FUZZY LANGUAGES AND THEIR RELATION TO

HUMAN AND MACHINE INTELLIGENCE

by

L. A. Zadeh

Memorandum No. ERL-M302

6 August 1971

FUZZY LANGUAGES AND THEIR RELATION TO HUMAN AND MACHINE INTELLIGENCE

by

L. A. Zadeh

Memorandum No. ERL-M302

6 August 1971

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

This is the text of a paper presented at the Conference on Man and Computer, Bordeaux, France, June 22-26, 1970. (Organized by Institut de la Vie, Paris.) The work described was supported in part by the National Science Foundation, Grant GK-10656X and the Joint Services Electronics Program, Contract F44620-71-C-0087.

Abstract

A fuzzy language as defined in this paper is a quadruple L = (U,T,E,N) in which U is a non-fuzzy universe of discourse; T (called term set) is a fuzzy set of terms which serve as names of fuzzy subsets of U; E (called an embedding set for T) is a collection of symbols and their combinations from which the terms are drawn, i.e., T is a fuzzy subset of E; and N is a fuzzy relation from E (or the support of T) to U called a naming relation.

As a fuzzy subset of E, T is characterized by a membership function $\mu_T : E \rightarrow [0,1]$, with $\mu_T(x)$ representing the grade of membership of a term x in T. Similarly, the naming relation N is characterized by a bivariate membership function $\mu_N : E \times U \rightarrow [0,1]$ in which $\mu_N(x,y)$ represents the strength of the relation between a term x and an object y in U.

The syntax and semantics of L are viewed as collections of rules for the computation of μ_T and μ_N, respectively. The meaning of a term x is defined to be a fuzzy subset, M(x), of U, whose membership function is given by $\mu_{M(x)}(y) = \mu_N(x,y)$.

Various concepts relating to fuzzy languages are introduced and their relevance to natural languages and human intelligence is pointed out. In particular, it is suggested that the theory of fuzzy languages may have the potential of providing better models for natural languages than is possible within the framework of the classical theory of formal languages.

1. Introduction

The question of whether or not machines can think has been the subject of many discussions and debates during the past two decades [1]-[10]. As computers become more powerful and thus more influential in human affairs, the philosophical aspects of this question become increasingly overshadowed by the practical need to develop an operational understanding of the limitations of machine judgment and decision-making ability. Can computers be relied upon to match people, decide on promotions and dismissals, make medical diagnoses, prescribe treatments, act as teachers, formulate business, political and military strategies, and, more generally, perform intellectual tasks of high complexity which in the past required expert human judgment? Clearly, this is already a pressing issue which is certain to grow in importance in the years ahead.

A thesis advanced in this paper is that there is indeed a very basic difference between human and machine intelligence which may well prove to be a very difficult obstacle in the path of designing machines that can outperform humans in the realm of cognitive processes involving concept formation, abstraction, pattern recognition, and decision-making under uncertainty. The difference in question lies in the ability of the human brain - an ability which present day digital computers do not possess - to think and reason in imprecise, non-quantitative, terms. Thus, a human being can understand and execute imprecise instructions such as "Increase x a <u>little</u> if y is <u>much</u> <u>larger</u> <u>than</u> 5," "rise <u>slowly</u>," "reduce speed if the road is <u>slippery</u>," and so forth. He can maneuver his car through dense traffic and park it in a tight spot. He can decipher sloppy hand-writing, understand distorted speech and untie a complicated knot. By contrast, the manipulative ability of digital computers is limited to precise instructions such as "add x to y," "if x = 5 then z = 3 else z = 7," "stop if x is non-negative," etc. In addition, a digital computer can accept digitized analog data and produce printed text, line drawings and the like under digital control. In all these cases, the input to the computer must be precisely defined.

The type of imprecision which is exemplified by the italicized words in the above instructions may be characterized as fuzziness, since it relates to the use of words such as little, slowly, slippery, etc. which in effect are labels for fuzzy sets,[1] that is, classes which admit of

---

[1] A fuzzy set is a class with fuzzy boundaries, that is, a class in which the transition from membership to non-membership is gradual rather than abrupt. More precisely, if $X = \{x\}$ is a collection of objects denoted generically by x, then a fuzzy subset of X, A, is a set of ordered pairs $\{(x,\mu_A(x))\}$, $x \in X$, where $\mu_A(x)$ is the grade of membership of x in A and μ_A is the membership function. Unless stated to the contrary, it will be assumed that $\mu_A(x)$ is a number in the interval [0,1], with 0 and 1 representing non-membership and full membership, respectively; more generally, $\mu_A(x)$ can be a point in a lattice. If A and B are fuzzy subsets of X, then A is a subset of B, written as $A \subset B$, iff $\mu_A(x) \leq \mu_B(x)$ for all x in X. The union of A and B is denoted by $A \cup B$ (or A + B when no confusion can arise) and is defined by $\mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x)$, $x \in X$, where $a \vee b$ denotes Max(a,b). The intersection of A and B is denoted by $A \cap B$ and is defined by $\mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x)$, $x \in X$, where $a \wedge b$ denotes Min(a,b). The complement, A', of A is defined by $\mu_{A'}(x) = 1 - \mu_A(x)$, $x \in X$. It should be noted that a membership function may be regarded as a predicate in a multivalued logic in which the truth values range over [0,1]. More detailed discussion of fuzzy sets and their properties may be found in [11]-[20].

grades of membership intermediate between full membership and non-membership. For example, the class of integers which are _much larger than_ 5 is a fuzzy set in which an integer such as 25 may be assigned a partial grade of membership, say 0.8, with 0 and 1 representing the extremes of non-membership and full membership, respectively. The same applies to classes characterized by words such as _green_, _tall_, _several_, _young_, _sparse_, _oval_, etc. Indeed, it may be argued that much, perhaps most, of human thinking and interaction with the outside world involves classes without sharply defined boundaries in which the transition from membership to non-membership is gradual rather than abrupt.

The ability of a human brain, weighing only a few hundred grams, to manipulate complicated fuzzy concepts and act on multidimensional fuzzy sensory inputs endows it with a capability to solve rather easily a wide variety of problems which, if formulated in precise quantitative terms, would exceed the computing power of the most powerful, the most sophisticated digital computer in existence. The explanation for this apparent paradox is that, in many instances, the solution to a problem need not be exact, so that a considerable measure of fuzziness in its formulation and results may be tolerable. The human brain is designed to take advantage of this tolerance for imprecision whereas a digital computer, with its need for precise data and instructions, is not. It is primarily for this reason that a problem which would be regarded as simple by a mentally retarded adult, might well be computationally infeasible for a machine equipped with a very large memory and operating at very high speed. A commonplace example of such a problem is that of parking a car.

Humans can park a car very easily and without making any use of quantitative measurements so long as the terminal position of the car is specified fuzzily rather than precisely. On the other hand, to program a computer to park a car in a specified location would be a very difficult problem involving precise quantitative data on the position of the car, its dimensions, dynamics and the parking space.

In general, complexity and precision bear an inverse relation to one another in the sense that, as the complexity of a problem increases, the possibility of analyzing it in precise terms diminishes. Thus it is a truism that the class of problems which are susceptible of exact solution is much smaller than that which can be solved approximately. From this point of view, the capacity of a human brain to manipulate fuzzy concepts and non-quantitative sensory inputs may well be one of its most important assets. Thus, "fuzzy thinking" may not be deplorable, after all, if it makes possible the solution of problems which are much too complex for precise analysis. For example, in the case of chess the choice of moves at an intermediate stage of the game is determined by subgoals, such as winning a piece or strengthening the center, which are fuzzily related to the ultimate goal - to win the game. Consequently, even though there is no imprecision or randomness in the rules of chess, the ability to play chess well depends in an essential way on the facility of the player in manipulating fuzzy concepts and relationships. The impressive performance of some chess-playing computer programs is not inconsistent with this assertion because the programs in question incorporate strategies which are arrived at through the ability of the programmer to operate on fuzzy sets and relations between them.

Although present day computers are not designed to accept fuzzy data or execute fuzzy instructions, they can be programmed to do so indirectly by treating a fuzzy set as a data-type which can be encoded as an array [21]. Granted that this is not a fully satisfactory approach to the endowment of a computer with an ability to manipulate fuzzy concepts, it is at least a step in the direction of enhancing the ability of machines to emulate human thought processes. It is quite possible, however, that truly significant advances in artificial intelligence will have to await the development of machines that can reason in fuzzy and non-quantitative terms in much the same manner as a human being.

A good illustration[2] of a problem which is far beyond the power of any existing computer is that of preparing a summary of a given document or book. The reason for this, in the first place, is that the notion of a summary is a fuzzy concept which cannot be defined in conventional terms for machine use. Second, and more important, the words in a natural language usually have fuzzy meaning, with the result that it is very difficult to devise an algorithm for constructing the meaning of a sentence, much less that of a concatenation of sentences, from the specification of the fuzzy meaning of individual words and the context in which they occur. Thus, to solve the problem of summarization, it would be necessary to develop a far better understanding of how to manipulate fuzzy concepts and relations than we possess at present.

---

[2] This example was suggested by Dr. M. Senko, Information Sciences Department, IBM, San Jose, California.

An essential step in this direction requires the construction of a conceptual framework for languages in which the syntax or semantics or both are fuzzy in nature. Such languages, which may appropriately be called _fuzzy languages_, could provide a significantly better approximation to natural languages than is possible within the framework of the classical theory of formal languages in which no provision is made for fuzziness in either syntax or semantics.

In what follows, we shall outline some of the basic aspects of the syntax and semantics of fuzzy languages, with the understanding that the theory of such languages is still in an embryonic stage at this juncture and our discussion of it will touch upon only a few of its many facets.

2. Fuzzy Languages

In the theory of formal languages [22]-[30], a language is defined as a set of strings over a finite alphabet. Such a definition is too narrow for many purposes because it fails to reflect the primary function of a language as a system of correspondences between strings of words and sets of objects or constructs which are described by these strings.

By contrast, in the definition of fuzzy languages given below, the correspondence between strings of words and sets of objects enters in an explicit fashion. Furthermore, the correspondence between words and objects is allowed to be fuzzy, as it is in the case of natural languages. In this way, the concept of a fuzzy language becomes much broader and more general than that of a formal language in its conventional sense.

Definition 1. A fuzzy language L is a quadruple

$$L = (U,T,E,N) \tag{1}$$

in which U is a non-fuzzy universe of discourse; T (called the term set) is a fuzzy set of terms which serve as names of fuzzy subsets of U; E (called an embedding set for T) is a collection of symbols and their combinations from which the terms are drawn, i.e., T is a fuzzy subset of E; and N is a fuzzy relation[3] from E (or, more specifically, the support of T) to U which will be referred to as a naming relation.

The first component of L is a universe of discourse, U, which may be any set of objects, actions, relations, concepts, etc. For example, U may be the set of integers; or the set of objects in a room; or the set of objects in a room together with the set of relations between

---

[3] A fuzzy relation R from $X = \{x\}$ to $Y = \{y\}$ is a fuzzy subset of the cartesian product $X \times Y = \{(x,y)\}$. E.g., if $x = y = R$ = real line, then >> (much larger than) is a fuzzy relation from R to R (or, more simply, a fuzzy relation in R). For a given ordered pair (x,y), the grade of membership $\mu_R(x,y)$ of (x,y) in R will be referred to as the strength of the relation between x and y. The domain of R is a fuzzy set in X denoted by dom(R) and defined by $\mu_{dom(R)}(x) = \bigvee_y \mu_R(x,y)$, where $\bigvee_y$ denotes the supremum over $Y = \{y\}$. Similarly, the range of R is a fuzzy set in Y denoted by ran(R) and defined by $\mu_{ran(R)}(y) = \bigvee_x \mu_R(x,y)$. (See [17] and [20] for additional details.)

The support of a fuzzy subset A of X is a non-fuzzy subset Supp(A) defined by $Supp(A) = \{x \mid \mu_A(x) > 0\}$. The cardinality of a fuzzy subset A with a finite support is denoted by $|A|$ and is defined by $|A| = \sum_i \mu_A(x_i)$, $x_i \in Supp(A)$. Essentially, the cardinality of a fuzzy set is a generalization of the notion of the number of members of a non-fuzzy set.

them; or the set of colors; or the union of the set of integers and the set of functions from integers to integers; etc. In essence, U, as its name implies, is the collection of objects or constructs which form the subject of discourse in L.

The second component of L, T, is a set of terms which serve as names of fuzzy subsets of U. The elements of T may have a variety of forms, e.g., they can be sounds, pictures, strings of letters, etc. In what follows, the terms will usually have the form of strings of letters or words drawn from a finite alphabet, with each word having a blank symbol (space) at its right end. For example, in the case of English, T would be the set of all English words and their well-formed concatenations.

The term set, T, is assumed to be a fuzzy subset of E, the embedding set for T, which in most cases is a collection of combinations of symbols drawn from an alphabet A. For example, in the case of English, A is the set of alphanumeric characters and E might be taken to be the collection of all finite strings of these characters. In the case of a formal language, A is usually denoted by V_T (set of terminals) and E is identified with V_T^* (the Kleene closure of V_T), which is the set of all finite strings over V_T.

A term may be <u>atomic</u> or <u>composite</u>. An atomic term is defined as a string which has no term as a substring. A <u>composite</u> term is a concatenation of atomic terms. E.g., words such as <u>red</u>, and <u>barn</u> are atomic terms, while their concatenation <u>red barn</u> is a composite term.

Since the term set, T, is assumed to be a fuzzy subset of E, it is characterized by a membership function $\mu_T : E \rightarrow [0,1]$ which associates with each[4] term $x \in E$ its grade of membership, $\mu_T(x)$, in T. For example, if E is the set of all finite strings over the alphabet $A = \{a,b,+\}$, then the grades of membership of some of the representative strings in T might be:

$$\mu_T(a+b) = 1.0 \qquad \mu_T(a+b+b) = 1.0$$

$$\mu_T(+a) = 0.8 \qquad \mu_T(+a+b) = 0.8$$

$$\mu_T(++a) = 0.1 \qquad \mu_T(a++b) = 0.1$$

The grade of membership, $\mu_T(x)$, may be used to represent the degree of <u>well-formedness</u> or <u>grammaticality</u> of x. For example, if T is the fuzzy set of words and phrases in English, then μ_T(John went home yesterday) = 1.0; μ_T(John yesterday went home) = 0.8; and μ_T(John home went yesterday) = 0.2. The important point to note is that in the model under discussion, the set of terms need not have a sharply defined boundary which separates these terms which belong to T from those that do not. Thus, the model allows a term to have a grade of membership in T which may lie somewhere between full membership on one end, and non-membership, on the other.

The fourth component of L is the fuzzy naming relation, N, from E to U. This relation is characterized by a bivariate membership function $\mu_N : \mathrm{Supp}(T) \times U \to [0,1]$, which associates with each[5] ordered pair (x,y),

---

[4] More generally, μ_T may be a partial function, i.e., $\mu_T(x)$ may be undefined for some x in E.

[5] As in the case of μ_T, μ_N may be a partial function over $\mathrm{Supp}(T) \times U$.

$x \in T$,[6] $y \in U$, the grade of membership $\mu_N(x,y)$, of (x,y) in N. In effect, $\mu_N(x,y)$ may be interpreted as the degree to which a term x fits an element y of U, and vice-versa. For example, if U is the set of ages from 1 to 100, x is the term young and y = 35 years, we may have μ_N(young, 35) = 0.2 while μ_N(old,35) = 0 and μ_N(middle-aged,35) = 0.02. Similarly, if y denotes the height, we may have

μ_N(tall, 5'8") = 0.6

μ_N(tall, 5'10") = 0.8

μ_N(tall, 6') = 1.0

μ_N(tall, 6'2") = 1.0

and likewise for other values of y.

The relationship between U, T, E and N is illustrated in Fig. 1.

Comment 2. In the above examples, the values of $\mu_N(x,y)$ are given for only a few representative values of x and y. To define a language completely, μ_N must be tabulated for all x in T and all y in U. In many practical situations, however, both μ_T and μ_N have to be estimated from partial information about them, such as the values which μ_T and μ_N take at a finite number of sample points in their respective domains of definition. When a fuzzy set is defined incompletely - and hence only

---

6 Here and elsewhere in this paper, $x \in T$ should be interpreted as $x \in \mathrm{Supp}(T)$.

approximately - in this fashion, it is said to be defined by exemplification.[7] The problem of estimating the membership function of a fuzzy set in X from the knowledge of its values over a finite set of points in X is the problem of abstraction, which plays a central role in pattern recognition [31],[32]. We shall not concern ourselves with this problem in the sequel and will assume throughout that μ_T and μ_N are either given or can be computed. It should be noted that the values assigned to $\mu_N(x,y)$ need not have an objective basis since they represent a subjective and, generally, context-dependent definition of a correspondence between the terms in T and elements of the universe of discourse.

When T and U are sets with a small number of elements, it may be practicable to define the naming relation N by a tabulation of $\mu_N(x,y)$. In general, however, both T and U are infinite sets, with the consequence that the characterization of T and N requires that they be endowed with a structure allowing the computation of $\mu_T(x)$ and $\mu_N(x,y)$ rather than a table look-up of their values. This is the rationale for the following definition of a structured fuzzy language.

Definition 3. A structured fuzzy language L is a quadruple

$$L = (U, S_T, E, S_N) \tag{2}$$

in which U is a universe of discourse; E is an embedding set for the term set T; S_T is a set of rules, called the syntactic rules of L,

[7] The definition of a fuzzy set by exemplification is an extension of the familiar linguistic notion of ostensive definition.

which collectively provide an algorithm for computing the membership function, μ_T, of the term set T; and S_N is a set of rules, called the <u>semantic</u> rules of L, which collectively provide an algorithm for computing the membership function, μ_N, of the fuzzy naming relation N.[8] The collection of syntactic and semantic rules of L constitute, respectively, the <u>syntax</u> and <u>semantics</u> of L.

<u>Comment 4</u>. Note that the only basic difference between Definition 1 and Definition 3 is that, in the case of an unstructured language, the set of terms T and the relation N are assumed to be defined explicitly by a tabulation of their respective membership functions or some equivalent means, whereas in the case of a structured language T and N are assumed to have an underlying structure which makes it possible to compute μ_T and μ_N through the use of syntactic and semantic rules, respectively.

It should be noted that when T is non-fuzzy, a procedure for computing μ_T reduces to a procedure for determining whether or not a given string x is an element of T, which in turn is equivalent to a procedure for generating elements of T.[9] Similarly, when N is non-fuzzy, a procedure for computing μ_N reduces to a procedure for determining whether or not a given ordered pair (x,y) belongs to N, which in turn is equivalent to a procedure for generating the ordered pairs (x,y) which are in N.

---

[8] As will be seen in Section 5, the semantic rules are used in the main to compute $\mu_N(x,y)$ when x is a composite term. For atomic terms, $\mu_N(x,y)$ will be assumed to be given as a function on U.

[9] We are tacitly assuming that T and N are recursively enumerable. See [30], pp. 5-7.

A language, whether structured or unstructured, will be said to be <u>fuzzy</u> if T or N or both are fuzzy. Consequently, a non-fuzzy language is one in which both T and N are non-fuzzy. In particular, a non-fuzzy structured language is a language with both non-fuzzy syntax and non-fuzzy semantics.

From this point of view, programming languages are non-fuzzy structured languages in which the <u>compiler</u> embodies the rules for computing the two-valued membership functions for the term set T and the naming relation N. Thus, by the use of syntactic rules, the compiler can determine whether or not a given string x is a term in T. If x is in T, then by the use of semantic rules the compiler can compute $\mu_N(x,y)$, $y \in U$ = set of machine language terms, and thus can determine a machine language instruction which corresponds to x.

In contrast to programming languages, natural languages have both fuzzy syntax and fuzzy semantics. The fuzziness of syntax manifests itself in the possibility that a sentence in, say, English, may have a degree of grammaticality[10] intermediate between complete correctness and incorrectness, e.g., μ_T(John yesterday went home) = 0.8. In most cases however, the degree of grammaticality of a sentence is either zero or one, so that the set of terms in a natural language has a fairly sharply defined boundary between grammatical and ungrammatical sentences.

The fuzziness of semantics, on the other hand, is a far more pronounced and pervasive characteristic of natural languages. For example, as was pointed out earlier, if the universe of discourse is identified

[10] For a discussion of grammaticality see [25] and [28].

with the set of ages from 1 to 100, then the atomic terms young and old do not correspond to sharply defined subsets of U. The same applies to composite terms such as not very young, not very young and not very old, etc. In effect, most of the terms in a natural language correspond to fuzzy rather than non-fuzzy subsets of the universe of discourse.

Our observation that natural languages are generally characterized by slightly fuzzy syntax and rather fuzzy semantics does not necessarily hold true when T is associated with an infinite rather than finite alphabet. Thus, when the terms of a language have the form of sounds, pictures, handwritten characters, etc., the fuzziness of its syntax may be quite pronounced. For example, the class of handwritten characters (or sounds) which correspond to a single letter, say R, is rather fuzzy, and this is even more true of concatenations of handwritten characters (or sounds).

3. The Meaning of Meaning

With the notion of a fuzzy language L = (U,T,E,N) as a point of departure, it becomes possible to give a concrete definition for the otherwise elusive concept of meaning. Specifically, let $\mu_N : \mathrm{Supp}(T) \times U \to [0,1]$ be the membership function characterizing N, with $\mu_N(x,y)$ representing the strength of the relation between a term x in T and an object y in U. Then, the definition of the meaning of x can be stated as follows [34]:

Definition 5. The meaning of a term x in T is a fuzzy subset M(x) of U in which the grade of membership of an element y of U is given by

$$\mu_{M(x)}(y) = \mu_N(x,y) \tag{3}$$

Thus, M(x) is a fuzzy subset of U which is conditioned on x as a parameter and which is a <u>section</u> of N in the sense that its membership function, $\mu_{M(x)}: U \to [0,1]$, is obtained by assigning a particular value, x, to the first argument in the membership function of N.

<u>Example 6</u>. As a very simple illustration of this definition, consider an unstructured language L = (U,T,E,N) in which among the elements of T are the terms <u>young</u>, <u>old</u> and <u>middle-aged</u>; U is the set of ages from 1 to 100; and N is a fuzzy naming relation from E to U defined by

$$\begin{aligned} \mu_N(\underline{\text{young}},y) &= 1 \quad , \quad \text{for } y < 25 \\ &= \left(1 + \left(\frac{y-25}{5}\right)^2\right)^{-1} , \quad \text{for } y \geq 25 \end{aligned} \tag{4}$$

$$\begin{aligned} \mu_N(\underline{\text{old}},y) &= 0 \quad , \quad \text{for } y < 50 \\ &= \left(1 + \left(\frac{y-50}{5}\right)^{-2}\right)^{-1} , \quad \text{for } y \geq 50 \end{aligned} \tag{5}$$

$$\begin{aligned} \mu_N(\underline{\text{middle}}\text{-}\underline{\text{aged}},y) &= 0 \quad , \quad \text{for } 1 \leq y < 35 \\ &= \left(1 + \left(\frac{y-45}{4}\right)^4\right)^{-1} , \quad \text{for } 35 \leq y < 45 \\ &= \left(1 + \left(\frac{y-45}{5}\right)^2\right)^{-1} , \quad \text{for } y \geq 45 \end{aligned} \tag{6}$$

for $1 \leq y \leq 100$.

Then the meaning of the term young is the fuzzy subset, M(young), of U = [1,100] whose membership function is given by

$$\mu_{M(\underline{young})}(y) = 1 \quad , \quad \text{for } y < 25 \tag{7}$$

$$= \left(1 + \left(\frac{y-25}{5}\right)^2\right)^{-1} , \quad \text{for } y \geq 25$$

and similarly for the meanings of old and middle-aged. (See Fig. 2.)

As another simple example, consider a fuzzy term such as several. If the universe of discourse is taken to be the set of non-negative integers, then several can be viewed as a name for a fuzzy subset M(several) of U given by the collection of ordered pairs

$$M(\underline{several}) = \{(3,0.4), (4,0.8), (5,1.0), (6,1.0), (7,1.0), (8,0.4)\} \tag{8}$$

in which we list only those pairs in which the grade of membership is positive.

In short, a term x, whether atomic or composite, is a name of a fuzzy subset of U. This subset, M(x), constitutes the meaning of x. The membership function of M(x) is given by (3), where μ_N is the naming relation in the language in which x is a term.

If N is a fuzzy naming relation from E to U, then its domain D(N) is a fuzzy set in T which is the shadow[11] of N on E. The membership

---

[11] If A is a fuzzy set in $X = X_1 \times X_2 \times \ldots \times X_n$, $X_i = \{x_i\}$, $i = 1,\ldots n$, with membership function $\mu(x_1,\ldots,x_n)$, then the shadow of A on $X_2 \times \ldots X_n$ is a fuzzy set in $X_2 \times \ldots \times X_n$ whose membership function, μ_1, is given by $\mu_1(x_2,\ldots,x_n) = \bigvee_{x_1} \mu(x_1,\ldots,x_n)$. (Additional details may be found in [11] and [12].

function of D(N) is given by

$$\mu_{D(N)}(x) = \bigvee_{y} \mu_N(x,y) \tag{9}$$

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

The grade of membership of x in D(N) may be interpreted as the <u>degree of meaningfulness</u> of x. Thus, x is <u>fully meaningful</u> if M(x) is a <u>normal</u>[12] fuzzy set, that is

$$\mu_{D(N)}(x) = \bigvee_{y} \mu_N(x,y) = 1 \tag{10}$$

Essentially, this implies that x is fully meaningful if there exists a y such that $\mu_N(x,y) = 1$. Conversely, x is <u>meaningless</u> if M(x) is an empty set, that is, $\mu_N(x,y) = 0$ for all y in U.

<u>Example 7</u>. Suppose that U is the set of integers U = {1, 2, ..., 10} and the terms <u>small</u>, <u>large</u>, <u>not small and not large</u>, <u>large and small</u>, are defined as the following fuzzy subsets of U.

M(<u>small</u>) = {(1,1.0), (2,1.0), (3,0.8), (4,0.2)}

M(<u>large</u>) = {(7,0.2), (8,0.8), (9,1.0), (10,1.0)}

M(<u>not small and not large</u>) = M'(<u>small</u>) ∩ M'(<u>large</u>)

M(<u>large and small</u>) = M(<u>large</u>) ∩ M(<u>small</u>)

---

[12] A fuzzy set A in X is <u>normal</u> iff $\bigvee_{x} \mu_A(x) = 1$; otherwise A is subnormal. If A has a finite support, this implies that A is normal iff there exists an x whose grade of membership in A is unity.

where M' denotes the complement of M and $\cap$ stands for the intersection (see Footnote 1). From the definition of complement and intersection, it follows that

$$M(\underline{not}\ \underline{small}\ \underline{and}\ \underline{not}\ \underline{large}) = \{(3,0.2),\ (4,0.8),\ (5,1.0),\ (6,1.0),\ (7,0.8),\ (8,0.2)\}$$

and

$$M(\underline{large}\ \underline{and}\ \underline{small}) = \text{empty set.}$$

Thus, <u>not</u> <u>small</u> <u>and</u> <u>not</u> <u>large</u> is fully meaningful, while <u>large and small</u> is meaningless.

An important aspect of meaning is its context - dependence. Thus, in general, the meaning of a term x when it is a component of a composite term depends on the context in which x occurs. To illustrate, in Example 6, the terms <u>young</u>, <u>old</u> and <u>middle-aged</u> were defined with a tacit understanding that they are adjectives applying to <u>man</u>. Clearly, the same adjectives when applied to, say, <u>dog</u>, would correspond to fuzzy sets in U quite different from those defined by (4), (5) and (6).

Can the terms like <u>young</u>, <u>old</u>, <u>tall</u>, etc. be defined in such a way as to make them relatively insensitive to the context in which they occur? One possibility lies in defining such terms on the basis of percentiles. Specifically, consider the term <u>tall</u> and assume that the tallness of an object y in a subset of U named z is measured in terms of its height, h(y). Further, let h_{50} denote the median of h(y) over z and h_r denote the r-percentile of h(y) over z, that is, a value of h such that 100-r percent of the number of members of z have height greater than or equal to h_r. Then, we would assign the grade of membership 0.5 in the fuzzy set labeled <u>tall</u> z to an object whose height is h_{50}, and the grade of membership $\mu_{\underline{tall}\ z}(y) = \frac{r}{100}$ to an object y whose height is h_r. More generally, the grade of membership of an object whose height is h_r might be related to r not linearly, as in $\frac{r}{100}$, but through an S-shaped function which takes the value 0.5 at r = 50 and tends to 0 and 1, respectively, as r approaches 0 and 100.

As a simple illustration, assume that U is the set of buildings in a city and z is the subset of hotels in that city. Suppose that the height of a particular hotel y is 150 feet and that this height represents the 75 percentile of the heights of hotels in the city. Then, the grade of membership of the hotel in question in the class named <u>tall hotel</u> in that city would be 0.75.

It should be noted that, in the case of natural languages, the context-dependence of meaning plays an important role in the resolution of ambiguities. Thus, if $x_1\ x_2\ x_3$ is a composite term and x_2 has two possible meanings, say $M^1(x_2)$ and $M^2(x_2)$, then x_2 would be assigned that meaning which maximizes the degree of meaningfulness of $x_1\ x_2\ x_3$. More generally, the rule governing the resolution of ambiguity may be stated informally as follows: If a component of a composite term has more than one meaning, assign that meaning to the component which maximizes the meaningfulness of the composite term in the context in which the latter occurs.

One of the most important aspects of the concept of meaning has to do with the semantic rules which make it possible to determine the meaning of a composite term from the knowledge of the meanings of its atomic components. This question will be considered in Section 5. As a preliminary, we shall turn our attention to some of the basic concepts underlying the syntax of fuzzy languages.

4. Syntax of Fuzzy Languages

As pointed out in [33], it is quite easy to generalize much of the theory of formal languages to the case where T is a fuzzy, rather than non-fuzzy, subset of strings over a finite alphabet. However, the resulting theory still falls far short of providing an adequate model for the syntax of natural languages for the case where the grade of membership of a composite term in T is equated with the degree of its grammaticality.

In what follows, we shall summarize and extend some of the main results of [33] and [34], and point to a connection between fuzzy term sets and non-fuzzy languages.

Following the standard notation of the theory of formal languages [30], let V_T be a finite alphabet of terminal symbols (e.g., alphanumeric characters in English) and let V_T^* denote the set of all finite strings composed of elements of V_T. Furthermore, let V_N denote a set of non-terminals, that is, a set of labels for the elements of a finite collection of fuzzy subsets of V_T^* called <u>syntactic categories</u>. For example, in the case of English, the elements of V_N would include N, standing for the syntactic category Noun; V, standing for Verb; NP, standing for Noun Phrase, etc. It is assumed that V_T and V_N are disjoint.

In the case of a fuzzy structured language $L = (U, S_T, V_T^*, S_N)$, the term set, T, is assumed to be a fuzzy subset of V_T^* characterized by a membership function $\mu_T : V_T^* \to [0,1]$ which associates with each string x in V_T^* its grade of membership, $\mu_T(x)$, in T, $0 \leq \mu_T(x) \leq 1$. The support of T is the set of all finite strings in V_T^* which have positive grades of membership in T.

It is convenient to represent T in the form of a power series (in the sense of [29])

$$T = \mu_1 x_1 + \mu_2 x_2 + \dots \tag{11}$$

where the x_i denote elements of the support of T and the μ_i are their respective grades of membership in T. Then, if the concatenation of two strings x and x' is denoted by xx', the concatenation of T with another fuzzy set of strings T'

$$T' = \mu_1' x_1' + \mu_2' x_2' + \dots \tag{12}$$

is denoted by TT' and is defined by

$$TT' = \sum_i \sum_j (\mu_i \wedge \mu_j') x_i x_j' \tag{13}$$

where

$$\mu_i \mu_j' = \mu_i \wedge \mu_j' = \mathrm{Min}(\mu_i, \mu_j') \tag{14}$$

and

$$\mu_i + \mu_j' = \mu_i \vee \mu_j = \mathrm{Max}(\mu_i, \mu_j') \tag{15}$$

Thus, (13), (14) and (15) imply that the grade of membership of the string $v = x_i x_j'$ in the fuzzy set TT' is given by

$$\mu_{TT'}(v) = \bigvee_{x_i, x_j'} (\mu_T(x_i) \wedge \mu_{T'}(x_j'))$$

where the supremum is taken over all x_i, x_j' such that $x_i x_j' = v$.

Example 8. As a very simple illustration, suppose that $V_T = \{a,b\}$ and

$$T = 0.2\ a + 0.3\ ab + 1.0\ aba$$

and

$$T' = 0.3\ a + 0.8\ aba + 1.0\ \varepsilon$$

where ε is the nullstring. Then

$$\begin{aligned} TT' = {} & (0.2 \wedge 0.3)aa + (0.3 \wedge 0.3)\ aba + (0.3 \wedge 1.0)\ abaa \\ & + (0.2 \wedge 0.8)\ aaba + (0.3 \wedge 0.8)\ ababa + (1.0 \wedge 0.8)\ abaaba \\ & + (0.2 \wedge 1)\ a + (0.3 \wedge 1.0)\ ab + (1.0 \wedge 1.0)\ aba \end{aligned}$$

which upon simplification becomes

$$\begin{aligned} TT' = {} & 0.2\ aa + 1.0\ aba + 0.3\ abaa + 0.2\ aaba + 0.3\ ababa \\ & + 0.8\ abaaba + 0.2\ a + 0.3\ ab \end{aligned}$$

The associativity of the concatenation of fuzzy sets of strings makes it possible to define the Kleene closure of T by the expression

$$T^* = \varepsilon + T + T^2 + T^3 + \dots \tag{16}$$

where + stands for union and T^n, n = 2, 3, ..., denotes an n-fold concatenation of T with itself. As will be seen presently, the notions of concatenation and Kleene closure of fuzzy sets of strings play significant roles in the definition of the syntax of T.

The function of the syntax, S_T, of L is to provide a set of rules for generating strings in the support of T together with their grades of membership in T. Such a set of syntactic rules constitutes a <u>fuzzy grammar</u> for L.

A particular form of fuzzy grammar can be obtained by generalizing the notion of a phrase-structure grammar [24]. Specifically, a fuzzy phrase-structure grammar or, simply, <u>fuzzy grammar</u> [33], is a quadruple

$$G = (V_T, V_N, S, P)$$

in which $S \in V_N$ is a starting symbol standing for the syntactic category "sentence" and P is a finite set of fuzzy productions of the form

$$\alpha \xrightarrow{\rho} \beta \tag{17}$$

where α and β are strings composed of elements of $V_T + V_N$ (except that $\alpha \neq \varepsilon$), and $0 < \rho \leq 1$. Thus, if $\alpha \xrightarrow{\rho} \beta$ and γ and δ are arbitrary strings in $(V_T + V_N)^*$, then

$$\gamma\alpha\delta \xrightarrow{\rho} \gamma\beta\delta \tag{18}$$

and $\gamma\beta\delta$ is said to be <u>directly derivable</u> from $\gamma\alpha\delta$. Note that α, β, γ and δ are, in effect, labels for fuzzy subsets of strings in V_T^*, and $\gamma\alpha\delta$ and $\gamma\beta\delta$ represent concatenations of these subsets in the sense of (13).

If u and v are two strings in $(V_T + V_N)^*$ and there exist strings $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ in $(V_T + V_N)^*$ such that

$$u \overset{\rho_1}{\rightarrow} \alpha_1 \overset{\rho_2}{\rightarrow} \alpha_2 \ldots \overset{\rho_{n-1}}{\rightarrow} \alpha_{n-1} \overset{\rho_n}{\rightarrow} v \tag{19}$$

then v is said to be <u>derivable</u> <u>from</u> u <u>via</u> <u>the</u> <u>derivation</u> <u>chain</u> $(u, \alpha_1, \alpha_2, \ldots, \alpha_{n-1}, v)$. The <u>strength</u> of this chain is defined to be the strength of its weakest link, that is,

$$\begin{aligned} \text{strength of } (u, \alpha_1, \alpha_2, \ldots, \alpha_{n-1}, v) &= \mathrm{Min}(\rho_1, \ldots \rho_n) \\ &= \rho_1 \wedge \rho_2 \wedge \ldots \wedge \rho_n \end{aligned} \tag{20}$$

The <u>strength</u>, ρ, <u>of</u> <u>the</u> <u>relation</u> <u>between</u> u and v is defined to be the strength of the strongest chain between u and v. Thus,

$$\rho = \mathrm{Sup}\ \mathrm{Min}(\rho_1, \ldots, \rho_n) \tag{21}$$

where Sup is taken over all derivation chains for u to v.
If v is derivable from u and the strength of the relation between u and v is ρ, then we write

$$u \overset{\rho}{\Rightarrow} v \tag{22}$$

The generation of T by G is governed by the definition:

<u>Definition 9</u>. A fuzzy grammar G generates a fuzzy term set T, or more

explicitly, T(G), in the following way. A terminal string[13] x is in T(G) (that is, in the support of T(G)) if x is derivable from S. The grade of membership of x in T(G), $\mu_T(x)$, is given by the strength of the relation between S and x.

Example 10. Suppose that $V_T = \{0,1\}$, $V_N = \{A,B,S\}$, and P is given by

$$
P: \quad \begin{array}{ll}
S \xrightarrow{0.5} AB & \qquad A \xrightarrow{0.5} a \\[1em]
S \xrightarrow{0.8} A & \qquad A \xrightarrow{0.6} b \\[1em]
S \xrightarrow{0.8} B & \qquad B \xrightarrow{0.4} A \\[1em]
AB \xrightarrow{0.4} BA & \qquad B \xrightarrow{0.2} a
\end{array}
$$

Consider the terminal string x = a. The possible derivation chains for this string are $S \xrightarrow{0.8} A \xrightarrow{0.5} a$, $S \xrightarrow{0.8} B \xrightarrow{0.2} a$, and $S \xrightarrow{0.8} B \xrightarrow{0.4} A \xrightarrow{0.5} a$. Hence

$$\mu_T(a) = (0.8 \wedge 0.5) \vee (0.8 \wedge 0.2) \vee (0.8 \wedge 0.4 \wedge 0.5) = 0.5$$

Similarly, the possible derivation chains for the terminal string x = ab are $S \xrightarrow{0.5} AB \xrightarrow{0.5} aB \xrightarrow{0.4} aA \xrightarrow{0.6} ab$, $S \xrightarrow{0.5} AB \xrightarrow{0.4} AA \xrightarrow{0.5} aA \xrightarrow{0.6} ab$, $S \xrightarrow{0.5} AB \xrightarrow{0.4} BA \xrightarrow{0.2} aA \xrightarrow{0.6} ab$, and $S \xrightarrow{0.5} AB \xrightarrow{0.4} BA \xrightarrow{0.4} AA \xrightarrow{0.5} aA \xrightarrow{0.6} ab$.

[13] A <u>terminal</u> string is a concatenation of terminals. A <u>sentential form</u> is a concatenation of terminals and non-terminals which is derivable from S.

Hence,

$$\mu_T(ab) = 0.4 \vee 0.4 \vee 0.2 \vee 0.4 = 0.4$$

Two fuzzy grammars G_1 and G_2 are <u>equivalent</u> if they generate the same fuzzy set of strings, that is,

$$T(G_1) = T(G_2).$$

For example, it is easy to verify that with G defined as in Example 10, the grammars G = ({0,1}, {A,B,S}, S, P) and G' = ({0,1}, {A,B,C,S}, S, P'), in which

$$\begin{aligned} P':\quad & S \xrightarrow{0.5} AB & & A \xrightarrow{0.5} a \\ & S \xrightarrow{0.8} A & & A \xrightarrow{0.6} b \\ & S \xrightarrow{0.8} B & & B \xrightarrow{0.4} A \\ & AB \xrightarrow{0.4} AC & & B \xrightarrow{0.2} a \\ & AC \xrightarrow{1.0} BC & & \\ & BC \xrightarrow{1.0} BA & & \end{aligned} \tag{23}$$

are equivalent.

For many purposes, it is convenient to express the productions in P in an algebraic notation which is similar in appearance to - but more general than - that used in connection with non-fuzzy languages [35]. The basic ingredients of this notation are: (a) the representation of a fuzzy set of strings in the power series form (11)

$$T = \mu_1 x_1 + \mu_2 x_2 + \dots . \tag{24}$$

where μ_i, $i = 1, 2, \dots$, is the grade of membership of the string x_i in T; (b) the definition of concatenation of fuzzy sets of strings (13); and (c) the definition of the expression γT[14] (in which $0 < \gamma \leq 1$) as a fuzzy set in which a generic string x has the grade of membership given by

$$\mu_{\gamma T}(x) = \gamma \wedge \mu_T(x) \tag{25}$$

With this understanding, a production of the form

$$\alpha \xrightarrow{\rho} \beta$$

in which α and β are labels for fuzzy subsets of V_T^*, may be replaced by the equation

$$\alpha = \rho\beta \tag{26}$$

in which $\rho\beta$ is a fuzzy set of strings defined by (25), i.e.,

$$\mu_{\rho\beta}(x) = \rho \wedge \mu_\beta(x) , \qquad x \in V_T^* \tag{27}$$

Furthermore, if P contains the productions

$$\alpha \xrightarrow{\rho_1} \beta_1 \tag{28}$$

[14] Note that an expression of the form γT in which $0 < \gamma \leq 1$ and T is a fuzzy set of strings may be regarded as a degenerate form of the concatenation T'T in which $T' = \gamma\varepsilon$, ε = nullstring. Then (25) follows from (13).

and

$$\alpha \xrightarrow{\rho_2} \beta_2 \tag{29}$$

then (28) and (29) give rise to the equation

$$\alpha = \rho_1 \beta_1 + \rho_2 \beta_2 \tag{30}$$

Example 11. Written in algebraic form, the production system of Example 10 reads

$$S = 0.5\ AB + 0.8\ A + 0.8\ B \tag{31}$$

$$AB = 0.4\ BA \tag{32}$$

$$A = 0.5\ a + 0.6\ b \tag{33}$$

$$B = 0.4\ A + 0.2\ a \tag{34}$$

The fuzzy set of strings generated by this grammar can be obtained by solving the system of equations (31)-(34) for S. Thus, on substituting (33) in (34) and using (13) and (25) we find

$$B = 0.4\ a + 0.4\ b$$

and hence

$$AB = (0.5\ a + 0.6\ b)(0.4\ a + 0.4\ b)$$

$$= 0.4\ aa + 0.4\ ba + 0.4\ ab + 0.4\ bb$$

Similar substitutions finally yield

$$T(G) = S = 0.5\ a + 0.6\ b + 0.4\ (aa + ba + ab + bb) \tag{35}$$

In solving a system of algebraic equations representing the production system of a fuzzy grammar, one frequently encounters linear equations of the form

$$u = \alpha u + \beta \tag{36}$$

in which u, α and β are fuzzy sets of strings over a finite alphabet, and + and the product denote the union and concatenation, respectively. A straightforward extension of Arden's theorem [37] to (36) yields the following proposition.

<u>Proposition 12</u>. If α does not contain the nullstring, then (36) has a unique solution for u which is given by

$$u = \alpha^* \beta \tag{37}$$

where α^* is the Kleene closure of α (in the sense of (16)).

<u>Example 13</u>. The solution of

$$u = (0.3\ a + 0.5b)u + 0.4\ a$$

is given by

$$u = (0.3\ a + 0.5\ b)^*\ 0.4\ a$$

which in expanded form reads

$$u = 0.4\ a + 0.3\ aa + 0.4\ ba + 0.3\ aaa + 0.3\ aba$$
$$+\ 0.3\ baa + 0.4\ bba + \ldots$$

A basic question in the theory of formal languages is whether or not there exists an algorithm for determining if a given terminal string x is in the language L(G) generated by a given G. The counterpart of this question in the case of fuzzy languages is the existence of an algorithm for computing the membership function μ_T for the fuzzy term set T(G) generated by a given fuzzy grammar G. If such an algorithm exists, then G is said to be recursive. In this sense, the grammar of Example 10 is recursive.

As in the case of non-fuzzy languages, it is convenient to classify the grammars of fuzzy languages into four principal categories, which in order of decreasing generality are:

Type 0 grammars

In this case, productions are of the general form

$$\alpha \overset{\rho}{\rightarrow} \beta \tag{38}$$

where α and β are strings in $(V_T + V_N)^*$, with $\alpha \neq \varepsilon$.

Example 14. Assume that V_T and V_N are as in Example 10. Then (38) is exemplified by $AB \overset{0.3}{\rightarrow} BA$, $ABa \overset{0.6}{\rightarrow} Bb$, $A \overset{0.8}{\rightarrow} b$.

Type 1 grammars (context-sensitive)

Here the productions are of the form

$$\alpha \overset{\rho}{\rightarrow} \beta \tag{39}$$

where α and β are strings in $(V_T + V_N)^*$, with $\alpha \neq \varepsilon$ and $|\beta| \geq |\alpha|$, that is, the length of the right-hand side (the consequent) must be at least as great as that of the left-hand side (the antecedent).

Example 15. (39) is exemplified by $AB \xrightarrow{0.5} BA$, $A \xrightarrow{0.8} bb$, but not by $BA \xrightarrow{0.9} B$.

If G is a context-sensitive grammar in the sense defined above, then it can readily be shown that there exists an equivalent grammar G' in which the productions are of the form

$$\beta A\gamma \xrightarrow{\rho} \beta\alpha\gamma \tag{40}$$

where $A \in V_N$, and $\alpha, \beta, \gamma \in (V_T + V_N)^*$, $\alpha \neq \varepsilon$. However, $S \xrightarrow{\rho} \varepsilon$ is allowed if S does not appear in any consequent. (40) implies that the nonterminal A can be replaced by α provided it occurs in the context (β,γ), that is, is preceded on the left by β and on the right by γ.

Example 16. (40) is exemplified by $aAb \xrightarrow{0.8} abb$, $Ab \xrightarrow{0.3} bbb$, but not by $AB \xrightarrow{0.3} BA$. However, by introducing a new nonterminal C, the latter production can be replaced by the following three productions of the form (40), with the resulting grammar being equivalent to the original one.

$$AB \xrightarrow{0.3} AC$$

$$AC \xrightarrow{1.0} BC$$

$$BC \xrightarrow{1.0} BA$$

An important property of context-sensitive grammars which is established in [33] is their recursiveness. This implies that, if the productions in a grammar G are of the form (40), there exists an algorithm for computing the grade of membership in T(G) of any terminal string x.

Type 2 grammars (context-free)

Here the allowable productions are of the form

$$A \overset{\rho}{\rightarrow} \alpha \tag{41}$$

where $A \in V_N$, $\alpha \in (V_T + V_N)^*$, and $S \overset{\rho}{\rightarrow} \varepsilon$ is allowed. Thus, in the case of a context-free grammar, A can be replaced by α regardless of the context in which A occurs.

In the case of non-fuzzy languages, context-free grammars are important because they can be used to generate, with some exceptions, well-formed strings in programming languages. Their relevance to natural languages, however, is not as great because context-sensitivity is a pervasive characteristic of such languages.

Type 3 grammars (regular)

In this case the allowable productions are of the form

$$A \overset{\rho}{\rightarrow} aB$$

$$A \overset{\rho}{\rightarrow} Ba$$

$$A \overset{\rho}{\rightarrow} a$$

$$S \overset{\rho}{\rightarrow} \varepsilon$$

where A, B $\in V_N$ and a $\in V_T$.

Comment 17. The algebraic notation which was described earlier is particularly useful in the case of context-free grammars. Thus if the nonterminals in V_N are denoted by $X_1, \ldots, X_n$, and $X = (X_1, \ldots, X_n)$, with $X_1 = S$, then the production system P can be put into the form

$$X = f(X) \tag{42}$$

where f is an n-vector whose components are multinomials in the X_i, $i = 1, \ldots, n$. In this way, the determination of the fuzzy set of strings generated by the grammar reduces to finding a fixed point of the function f. In this connection, it can really be shown that if we set $X' = \theta$ = empty set and form the iterates

$$X^{k+1} = f(X^k), \quad X^0 = \theta, \quad k = 1, 2, 3, \ldots \tag{43}$$

then, for each k, X^k is a fuzzy subset of the solution of (42).

Decomposition of a fuzzy grammar into non-fuzzy grammars

An important connection between fuzzy and non-fuzzy grammars relates to the possibility of decomposing a fuzzy grammar - in the sense defined below - into non-fuzzy grammars of the same type.

This possibility stems from a basic property of fuzzy sets which is stated below.

Let A be a fuzzy set in a space X = {x}, where x denotes a generic element of X. For λ in (0,1], define a λ-level-set or simply, a level set [11] of A as a non-fuzzy set A_λ comprising all elements of X whose grade of membership in A is greater than or equal to λ, i.e.,

$$A_\lambda = \{x \mid \mu_A(x) \geq \lambda\} \tag{44}$$

Clearly, the A_λ form a nested collection of subsets of X, with

$$\lambda \geq \lambda' \Rightarrow A_\lambda \subset A_{\lambda'} \tag{45}$$

As shown in [20], A admits of the resolution expressed by

$$A = \sum_\lambda \lambda A_\lambda \quad , \qquad 0 < \lambda \leq 1 \tag{46}$$

where Σ stands for the union of fuzzy sets and λA_λ denotes a fuzzy set with a two-valued membership function defined by[15]

$$\begin{aligned} \mu_{\lambda A_\lambda}(x) &= \lambda \quad , \quad \text{for } x \in A_\lambda \\ &= 0 \quad , \quad \text{elsewhere} \end{aligned} \tag{47}$$

To illustrate, let $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and assume that A, expressed as a power series, is given by

$$A = 0.3\, x_1 + 0.5\, x_2 + 0.6\, x_3 + 0.8\, x_4 + 1.0\, x_5 + 1.0\, x_6 \tag{48}$$

Then

---

[15] More generally, if A is a fuzzy set in $X = \{x\}$, with μ_A denoting the grade of membership of x in A, then λA is a fuzzy set in X such that the grade of membership of x in λA is $\lambda \wedge \mu_A(x)$. Thus, if A is expressed in power series form as $A = \mu_1\, x_1 + \mu_2\, x_2 + \ldots$, then $\lambda A = (\lambda \wedge \mu_1)x_1 + (\lambda \wedge \mu_2)x_2 + \ldots$. Note that this is consistent with (25).

$A_{1.0} = \{x_5,x6\}$

$A_{0.8} = \{x_4,x_5,x_6\}$

$A_{0.6} = x_3,x_4,x_5,x_6$

$A_{0.5} = x_2,x_3,x_4,x_5,x_6$

$A_{0.3} = x_1,x_2,x_3,x_4,x_5,x_6$

and the resolution of A reads

$$A = 0.3\ A_{0.3} + 0.5\ A_{0.5} + 0.6\ A_{0.6} + 0.8\ A_{0.8} + 1.0\ A_{1.0} \tag{49}$$

where + denotes the union of fuzzy sets. A straightforward way of verifying the equivalence of (48) and (49) is to substitute the power series expressions for $A_{0.3}, \ldots, A_{1.0}$ into (49), yielding

$$A = 0.3\ (x_1+x_2+x_3+x_4+x_5+x_6) + 0.5\ (x_2+x_3+x_4+x_5+x_6) + 0.6\ (x_3+x_4+x_5+x_6)$$
$$+ 0.8\ (x_4+x_5+x6) + 1.0\ (x_5+x_6)$$

Then, on noting that, by the definition of + ,

$$\lambda x_i + \lambda' x_i = (\lambda \vee \lambda')x_i \quad , \quad i = 1, \ldots, 6$$

we obtain (48).

To illustrate the application of the resolution expressed by (46) to fuzzy grammars, it will be convenient to focus our attention on context-free grammars, with the understanding that the same conclusions apply as well to grammars of type 0, 1, and 3.

Specifically, consider a fuzzy context-free grammar $G = (V_T, V_N, S, P)$ and let P_λ be the subset of productions in P such that if $\alpha \overset{\rho}{\rightarrow} \beta$, with $\rho \geq \lambda$, is in P, then $\alpha \overset{1.0}{\rightarrow} \beta$ (or simply $\alpha \rightarrow \beta$) is in P_λ. Further, let

$$G_\lambda = (V_T, V_N, S, P_\lambda) \tag{50}$$

be a non-fuzzy grammar with the production system P_λ.

The non-fuzzy grammar G_λ generates a non-fuzzy context-free term set $T(G_\lambda)$. As we shall see presently, T(G) can be resolved into the $T(G_\lambda)$ just as a fuzzy set A can be resolved into its level sets A_λ. More specifically, we can assert the following proposition.

<u>Proposition 18</u>. If $G(V_T, V_N, S, P)$ is a fuzzy context-free grammar and the G_λ, $0 < \lambda \leq 1$, are non-fuzzy context-free grammars defined by (50), then

$$T(G) = \sum_\lambda \lambda T(G_\lambda), \quad \lambda = \text{values of } \rho \text{ in } P \tag{51}$$

where T(G) and $T(G_\lambda)$ are the fuzzy context-free and non-fuzzy context-free term sets generated by G and G_λ, respectively, and $\lambda T(G_\lambda)$ is a fuzzy set of terms such that [16]

$$\mu_{\lambda T(G_\lambda)}(x) = \lambda \wedge \mu_{T(G_\lambda)}(x) \tag{52}$$

[16] Note that (52) is consistent with (25) as well as with the definition given in Footnote 15.

To prove (51) it is sufficient to note that (51) would be a special case of the resolution of a fuzzy set (see (46)) if the $T(G_\lambda)$ were the level sets of $T(G)$. Thus, all that is necessary to show is that $T(G_\lambda)$, which is the fuzzy set of terms generated by G_λ, is a λ-level set of $T(G)$. To this end, let x be a terminal string in $T(G_\lambda)$. Since the ρ of all productions in G_λ is greater than or equal to λ, it follows from the definitions of $\mu_{T(G)}(x)$ and $\mu_{T(G_\lambda)}(x)$ (see Definition 9) that $\mu_{T(G)}(x) \geq \mu_{T(G_\lambda)}(x) \geq \lambda$ and hence that x belongs to the λ-level set of $T(G)$. Conversely, let x be a terminal string in the λ-level set of $T(G)$. Then $\mu_{T(G)}(x) \geq \lambda$ and by Definition 9 it follows that x is derivable from S via a derivation chain which uses only those productions in G in which $\rho \geq \lambda$. Consequently, x belongs to $T(G_\lambda)$. Thus, both

$$T(G_\lambda) \subset \lambda\text{-level set of } T(G)$$

and

$$\lambda\text{-level set of } T(G) \subset T(G_\lambda)$$

are true, and hence

$$T(G_\lambda) = \lambda\text{-level set of } T(G) \tag{53}$$

which is what we set out to establish.

<u>Example 19</u>. Consider the fuzzy grammar

$$G = (\{a,b\}, \{A,B,S\}, S, P)$$

in which

$$P: \quad S \xrightarrow{0.8} bA \qquad B \xrightarrow{1.0} b$$

$$S \xrightarrow{0.6} aB \qquad A \xrightarrow{0.3} bSA$$

$$A \xrightarrow{1.0} a \qquad B \xrightarrow{0.3} aSB$$

In this case, the non-fuzzy production systems P_λ are given by

$P_{1.0}$: $A \to a$ $\quad$ $B \to b$

$P_{0.8}$: $A \to a$ $\quad$ $B \to b$

$S \to bA$

$P_{0.6}$: $A \to a$ $\quad$ $B \to b$

$S \to bA$ $\quad$ $S \to aB$

$P_{0.3}$: $A \to a$ $\quad$ $B \to b$

$S \to bA$ $\quad$ $S \to aB$

$A \to bSA$ $\quad$ $B \to aSB$

and the non-fuzzy context-free term sets generated by the corresponding grammars $G_{1.0}$, $G_{0.8}$, $G_{0.6}$ and $G_{0.3}$ are $T(G_{1.0})$, $T(G_{0.8})$, $T(G_{0.6})$ and $T(G_{0.3})$. In terms of these, the fuzzy term set generated by G is given by the resolution

$$T(G) = 0.3\ T(G_{0.3}) + 0.6\ T(G_{0.6}) + 0.8\ T(G_{0.8}) + T(G_{1.0}) \qquad (54)$$

It is easy to show that the converse of Proposition 18 also holds true.

Thus, if the G_λ, $0 < \lambda \leq 1$, constitute a nested sequence of non-fuzzy context-free grammars such that

$$\lambda \geq \lambda' \Rightarrow P_\lambda \subset P_{\lambda'} \tag{55}$$

then the expression

$$\sum_\lambda \lambda T(G_\lambda) \tag{56}$$

will represent a fuzzy term set which can be generated by a fuzzy context-free grammar.

As pointed out in [33], many of the basic results in the theory of non-fuzzy formal languages can readily be extended to fuzzy term sets defined by fuzzy grammars. For example, it is easy to show, both directly [33] and by making use of the resolution of fuzzy term sets,[17] that a fuzzy context-free term set can be put into the Chomsky and Greibach normal forms. Similarly, it can readily be shown [36] that a fuzzy context-free term set is accepted by a fuzzy push-down automaton. We shall not discuss these and other extensions[18] in the present paper and instead will turn our attention to the semantics of fuzzy languages.

5. Semantics of Fuzzy Languages

Consider a structured fuzzy language $L = (U, S_T, E, S_N)$ in which S_T is a set of syntactic rules defining a term set $T \subset E$, U is a universe

---

[17] The possibility of establishing the validity of the Chomsky and Greibach normal forms for fuzzy context-free grammars by making use of the resolutions of fuzzy term sets was suggested to the author by Professor R. Karp.

[18] A number of interesting results may be found in [39].

of discourse, and S_N is a set of semantic rules defining a fuzzy naming relation N from E to U. To simplify our discussion, we shall assume that T is a non-fuzzy subset of E which can be generated by a context-free grammar.

As was stated previously, the central problem of semantics is that of specifying a set of semantic rules, S_N, which can serve as an algorithm for computing the meaning of a composite term in T from the knowledge of the meanings of its components. In the case of an artificial language, especially a programming language, the semantic rules can be set by the designer of the language. In the case of natural languages, on the other hand, the semantic rules must be deduced from a partial knowledge of the membership function, $\mu_N : \mathrm{Supp}(T) \times U \to [0,1]$, of the naming relation N. More specifically, S_N must be deduced from a finite set of ordered pairs $\{((x_i,y_j), \mu_N(x_i,y_j))\}$, $i = 1, \ldots, k$, $j = 1, \ldots, m$, in which the x_i and y_j are examples (i.e., sample points) in T and U, respectively, and $\mu_N(x_i,y_j)$ is the strength of the relation between x_i and y_j. From this point of view, the deduction of S_N constitutes a problem in abstraction - which, as was pointed out earlier - plays a central role in the field of pattern recognition [31],[32].[19]

At present, there are no systematic techniques for solving the problem of abstraction and thus the deduction of S_T and S_N for natural languages must be carried out in an ad hoc fashion. Indeed, the complexity of natural languages is so great that it is not even clear, at this juncture, what the form of the rules in S_N should be.

To make at least a modest beginning toward the development of a

[19] For a discussion of grammatical inference see [40] and [41].

quantitative theory of semantics within the conceptual framework we have constructed so far, it is expedient to start with a few relatively simple special cases involving fragments of natural or artificial languages. In such cases, we can give explicit quantitative rules for determining the meaning of a composite term from the knowledge of the meanings of its components. The following simple examples are intended to illustrate the manner in which this can be done.

Example 20. Suppose the terms young and old are defined as in Example 6, that is, as fuzzy subsets of the set of integers K = [0,100], characterized respectively by the membership functions

$$\mu_N(\underline{young},y) = 1, \quad \text{for } y < 25$$

$$= \left(1 + \left(\frac{y-25}{5}\right)^2\right)^{-1}, \quad \text{for } y \geq 25$$

and

$$\mu_N(\underline{old},y) = 0 \quad \text{for } y < 50$$

$$= \left(1 + \left(\frac{y-50}{5}\right)^{-2}\right)^{-1}, \quad \text{for } y \geq 50$$

We wish to define also the modifiers not and very and the connectives or and and. To this end, let $\mathcal{F}(K)$ denote the fuzzy power set of K, that is, the set of all fuzzy subsets of K. Then, the modifier not can be regarded as a function from $\mathcal{F}(K)$ to $\mathcal{F}(K)$ defined by

$$\mu_N(\underline{not}\ x,y) = 1 - \mu_N(x,y), \quad y \in K \tag{57}$$

where the term x is a label for a fuzzy subset of K and <u>not</u> x is a composite term consisting of a concatenation of <u>not</u> and x. Thus,

$$\mu_N(\underline{\text{not young}},y) = 1 - \mu_N(\underline{\text{young}},y) \quad , \qquad y \in K$$

and

$$\mu_N(\underline{\text{not old}},y) = 1 - \mu_N(\underline{\text{old}},y)$$

with <u>not</u> acting as a <u>complementer</u>.

Similarly, the term <u>very</u> can be regarded as a function from $\mathcal{F}(K)$ to $\mathcal{F}(K)$ defined by, say

$$\mu_N(\underline{\text{very}}\ x,y) = \mu_N^2(x,y) \quad , \qquad y \in K \tag{58}$$

which has the effect of concentrating the membership function of x around its maximum value. Thus,

$$\mu_N(\underline{\text{very young}},y) = \mu_N^2(\underline{\text{young}},y) \quad , \qquad y \in K$$

and

$$\mu_N(\underline{\text{very old}},y) = \mu_N^2(\underline{\text{old}},y)$$

The effect of concentration on the term <u>old</u> is illustrated in Fig. 3. To place in evidence this property of the term <u>very</u>, it will be referred to as a <u>concentrator</u>.

The connective <u>or</u> is a function from $\mathcal{F}(K) \times \mathcal{F}(K)$ to $\mathcal{F}(K)$ which serves to generate the union of its arguments. Thus, if x_1 and x_2 are terms, then the meaning of the composite term x_1 <u>or</u> x_2 is defined by

$$M(x_1 \underline{\text{ or }} x_2) = M(x_1) \cup M(x_2) \tag{59}$$

or, in terms of membership functions,

$$\mu_N(x_1 \underline{\text{ or }} x_2, y) = \mu_N(x_1, y) \vee \mu_N(x_2, y) \tag{60}$$

Similarly, the connective <u>and</u> is a function from $\mathcal{F}(K) \times \mathcal{F}(K)$ to $\mathcal{F}(K)$ which serves to generate the intersection of its arguments. Thus,

$$M(x_1 \underline{\text{ and }} x_2) = M(x_1) \cap M(x_2) \tag{61}$$

or, more explicitly,

$$\mu_N(x_1 \underline{\text{ and }} x_2, y) = \mu_N(x_1, y) \wedge \mu_N(x_2, y) \quad , \quad y \in K \tag{62}$$

It should be noted that, whereas the meaning of x_1 <u>and</u> x_2 is a fuzzy subset of K defined by (62), the meaning of <u>and</u> is a fuzzy subset of $\mathcal{F}(K) \times \mathcal{F}(K) \times \mathcal{F}(K)$ rather than K. Consequently, to define the terms <u>young</u>, <u>old</u>, <u>not</u>, <u>or</u> and <u>and</u> as fuzzy subsets of the universe of discourse, K must be augmented[20] with the collections $\mathcal{F}(K) \times \mathcal{F}(K)$ and $\mathcal{F}(K) \times \mathcal{F}(K) \times \mathcal{F}(K)$, resulting in the expression

$$U = K + \mathcal{F}(K) \times \mathcal{F}(K) + \mathcal{F}(K) \times \mathcal{F}(K) \times \mathcal{F}(K) \tag{63}$$

where + stands for union and × for the cartesian product.

Another point that should be noted is that, in English, the connective <u>and</u> may be used in a sense other than that defined above. For

[20] This point is discussed more fully in [34].

example, in the sentence "The box contains nuts <u>and</u> bolts," <u>and</u> serves to define a set of objects consisting of the union rather than the intersection of its arguments. As was pointed out earlier, this type of context-dependence is characteristic of natural languages.

To facilitate the determination of the meaning of a composite term, it is convenient to construct a covering of the term set T with a collection of <u>syntactic categories</u> which are non-fuzzy subsets of T. For example, in the case of English the syntactic category Noun would contain such terms as <u>dog</u>, <u>cat</u>, <u>door</u>, <u>car</u>, etc., while the syntactic category Adjective would contain <u>red</u>, <u>tall</u>, <u>young</u>, <u>old</u>, <u>narrow</u>, etc.

Now suppose that x_1 is an adjective, i.e., $x_1 \in$ Adjective, and x_2 is a Noun. Then, if $M(x_1)$ and $M(x_2)$ are the fuzzy subsets of U representing, respectively, the meanings of x_1 and x_2, the meaning of the composite term x_1x_2 is defined as the intersection of $M(x_1)$ and $M(x_2)$, i.e.,

$$M(x_1x_2) = M(x_1) \cap M(x_2) \tag{64}$$

For example, if U consists of the totality of objects in a room and x_1 = <u>red</u> and x_2 = <u>chair</u>, then $M(x_1)$ is the fuzzy set of red objects in the room, $M(x_2)$ is the set of chairs in the room, and $M(x_1x_2)$ is the fuzzy subset of <u>red chairs</u> in the room. According to (62), if the grade of membership of an object in the fuzzy set of <u>red</u> objects is 0.8, say, while its grade of membership in the set of <u>chairs</u> is 1.0, then its grade of membership in the fuzzy set of <u>red chairs</u> is $0.8 \wedge 1.0 = 0.8$.

In the above example, $M(x_1x_2)$, with $x_1 \in$ Adjective and $x_2 \in$ Noun, is a subset of both $M(x_1)$ and $M(x_2)$. This would not necessarily be the case if x_1 were a member of a syntactic category other than Adjective. For example, if x_1 were a verb, e.g., x_1 = ran, and x_2 = home, then $M(x_1)$ would be a fuzzy subset of a set of actions, say A, while $M(x_2)$ is a fuzzy subset of a set of objects, say Q. In this case, M(ran home) would be a fuzzy subset of the cartesian product $A \times Q$, rather than a subset of either A or Q. However, if $M(x_1)$ and $M(x_2)$ are interpreted as cylindrical[21] fuzzy subsets of $A \times Q$, then $M(x_1 x_2)$ may be taken to be the intersection of $M(x_1)$ and $M(x_2)$.

In the cases considered so far, the semantic rules governing the construction of the meaning of a composite term are quite simple, e.g.,

$$M(x_1 \underline{\text{ or }} x_2) = M(x_1) \cup M(x_2) \tag{65}$$

and

$$x_1 \in \text{Adjective and } x_2 \in \text{Noun} \Rightarrow M(x_1 x_2) = M(x_1) \cap M(x_2) \;{}^{22} \tag{66}$$

[21] A fuzzy subset of a product space $X_1 \times X_2 \times \ldots \times X_n$, $X_i = \{x_i\}$, $i = 1, \ldots, n$, is cylindrical if it is characterized by a membership function whose arguments form a proper subset of $\{x_1, \ldots, x_n\}$. E.g., for $i = 2$, a fuzzy set where membership function is a function of x_1 alone is a cylindrical fuzzy subset of $X_1 \times X_2$.

[22] A rule such as (66) is much too simple to hold for all adjectives and all nouns in a natural language. In general, the usual syntactic categories, e.g. Adjective, Noun, etc. are too broad for rules like (66), necessitating the use of a finer covering of T than is provided by the syntactic categories in question. Thus, in the case of English, it should be understood that the validity of (66) is restricted to certain subcategories of the syntactic categories Adjective and Noun.

In a more complex example which is described in [34], the atomic terms in T are: young, old, very, not, and, (,), and the composite terms in T are generated by a grammar G in which S,A,B,C,D and Y are nonterminals and the production system is given by

$$
\begin{array}{ll}
S \rightarrow A & C \rightarrow O \\
S \rightarrow S \ \underline{\text{or}} \ A & C \rightarrow Y \\
A \rightarrow B & O \rightarrow \underline{\text{very}} \ O \\
A \rightarrow A \ \underline{\text{and}} \ B & Y \rightarrow \underline{\text{very}} \ Y \\
B \rightarrow C & O \rightarrow \underline{\text{old}} \\
B \rightarrow \underline{\text{not}} \ C & Y \rightarrow \underline{\text{young}} \\
C \rightarrow (S) &
\end{array}
\tag{67}
$$

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 $\mu_N(x,y)$ when x is a composite term, one can use an approach similar to that described by Knuth in [38]. Specifically, suppose that we are given $\mu_N(\underline{\text{young}},y)$ and $\mu_N(\underline{\text{old}},y)$. The remaining atomic terms are regarded as functions on $\mathcal{F}(K)$ or $\mathcal{F}(K) \times \mathcal{F}(K)$ 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 $\mu(H)$ as an abbreviation for $\mu_N(H,y)$, where H is a terminal or non-terminal symbol, the rules in question can be expressed as

$$\begin{array}{ll}
S \rightarrow A & \Rightarrow \mu(S_L) = \mu(A_R) \\
A \rightarrow B & \Rightarrow \mu(A_L) = \mu(B_R) \\
B \rightarrow C & \Rightarrow \mu(B_L) = \mu(C_R) \\
S \rightarrow S \ \underline{\text{or}} \ A & \Rightarrow \mu(S_L) = \mu(S_R) \vee \mu(A_R) \\
A \rightarrow A \ \underline{\text{and}} \ B & \Rightarrow \mu(A_L) = \mu(A_R) \wedge \mu(B_R) \\
B \rightarrow \underline{\text{not}} \ C & \Rightarrow \mu(B_L) = 1 - \mu(C_R) \\
O \rightarrow \underline{\text{very}} \ O & \Rightarrow \mu(O_L) = (\mu(O_R))^2 \\
Y \rightarrow \underline{\text{very}} \ Y & \Rightarrow \mu(Y_L) = (\mu(Y_R))^2 \\
C \rightarrow O & \Rightarrow \mu(C_L) = \mu(O_R) \\
C \rightarrow Y & \Rightarrow \mu(C_L) = \mu(Y_R) \\
C \rightarrow (S) & \Rightarrow \mu(C_L) = \mu(S_R) \\
O \rightarrow \underline{\text{old}} & \Rightarrow \mu(O_L) = \mu(\underline{\text{old}}) \\
Y \rightarrow \underline{\text{young}} & \Rightarrow \mu(Y_L) = \mu(\underline{\text{young}})
\end{array} \tag{68}$$

Now consider a composite term such as

$$x = \underline{\text{not}} \ \underline{\text{very}} \ \underline{\text{young}} \ \underline{\text{and}} \ \underline{\text{not}} \ \underline{\text{very}} \ \underline{\text{very}} \ \underline{\text{old}}.$$

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

$$\mu_N(x,y) = (1 - \mu_N^2(\underline{\text{young}},y)) \wedge (1 - \mu_N^4(\underline{\text{old}},y)). \tag{69}$$

More generally, as a first step in the computation of $\mu_N(x,y)$ 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 Figure 4. (The subscripts in this figure serve the purpose of numbering the nodes.)

Proceeding from bottom to top and employing the relations (68) for the computation of the membership function at each node, we obtain the system of non-linear equations:

$$\begin{aligned}
\mu(Y_7) &= \mu_N(\underline{\text{young}},y) \\
\mu(Y_6) &= \mu^2(Y_7) \\
\mu(C_5) &= \mu(Y_6) \\
\mu(B_4) &= 1 - \mu(C_5) \\
\mu(A_3) &= \mu(B_4) \\
\mu(O_{12}) &= \mu_N(\underline{\text{old}},y) \\
\mu(O_{11}) &= \mu^2(O_{12}) \\
\mu(O_{10}) &= \mu^2(O_{11}) \\
\mu(C_9) &= \mu(O_{10}) \\
\mu(B_8) &= 1 - \mu(C_9) \\
\mu(A_2) &= \mu(A_3) \wedge \mu(B_8) \\
\mu_N(x,y) &= \mu(S_1) = \mu(A_2)
\end{aligned} \tag{70}$$

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 (69).

The basic idea underlying the approach sketched above is the following: The semantic rules governing the computation of the meaning of a composite term x are induced by the syntactic rules by which x is generated from S in the grammar G defining the term set T. In particular, each production in G induces a relation between the membership functions of the fuzzy sets whose labels appear in the production in question.

Approaches such as this can be of use in the construction of query languages for information retrieval systems. It appears that they also have the potential for providing reasonably good models for the semantics of subsets of natural languages. Such models will be described in a subsequent paper.

Concluding Remarks

The concept of a fuzzy language differs from that of a formal language in two important respects. First, it incorporates a naming relation, N, which serves to define a correspondence between a set of terms, T, and a universe of discourse, U; and second, it allows both the set of terms and the naming relation to be fuzzy.

With the concept of a fuzzy language as a point of departure, the notions of syntax and semantics can be given a precise meaning as algorithms serving to compute the membership functions of T and N, respectively. From this point of view, the central problem in semantics may be regarded as that of computing the meaning of a composite term $x_1\ x_2 \ldots x_n$ from the knowledge of the meanings of its components, $x_1, x_2, \ldots, x_n$.

At present, the theory of fuzzy languages is in an embryonic stage. Eventually, it may serve to provide considerably better models for natural languages than is possible within the restricted framework of the classical theory of formal languages.

REFERENCES

[1] N. Rashevsky, "The Logical Mechanisms of Logical Thinking," Bull. Math. Biophysics, vol. 8, 1946.

[2] A. M. Turing, "Computing Machinery and Intelligence," Mind, vol. 59, pp. 433-460, 1950; also in Computers and Thought, edited by E. Feigenbaum and J. Feldman, pp. 11-35, McGraw-Hill Book Co., 1963

[3] J. von Neumann, "The Computer and the Brain," Yale Univ. Press, 1958.

[4] W. R. Ashby, "Design for a Brain," J. Wiley, 1960.

[5] J. Pfeiffer, "Thinking Machine," Lippincott, 1962.

[6] M. Minsky, "Matter, Mind and Models," Proc. IFIP Congress, 1965, pp. 45-49.

[7] H. L. Dreyfus, "Alchemy and Artificial Intelligence," Rand Paper P-3244, Rand Corp., Santa Monica, Calif., 1965.

[8] S. Papert, "The Artificial Intelligence of Hubert L. Dreyfus: A Budget of Fallacies," Memorandum 154, Project MAC, MIT, January 1968.

[9] H. L. Dreyfus, "A Critique of Artificial Reasoning," Harper and Row, 1971.

[10] R. E. Bellman, "Man, Society and the Computer," to appear in the Proceedings of the Conference on Man and Computer (Institut de la vie), Bordeaux, France, 1970.

[11] L. A. Zadeh, "Fuzzy sets," Information and Control, vol. 8, pp. 338-353, June 1965.

[12] L. A. Zadeh, "Shadows of Fuzzy Sets," Problems in Transmission of Information (in Russian), vol. 2, pp. 37-44, March 1966.

[13] J. Goguen, "L-Fuzzy Sets," Jour. Math. Anal. and Appl., vol. 18, pp. 145-174, April 1967.

[14] C. L. Chang, "Fuzzy Topological Spaces," Jour. Math. Analysis and Applications, vol. 24, pp. 182-190, 1968.

[15] E. J. Santos, "Maximim Automata," Information and Control, vol. 13, pp. 363-377, October 1968.

[16] M. Mizumoto, J. Toyoda, and K. Tanaka, "Some Considerations on Fuzzy Automata," Jour. of Computer and System Sciences, vol. 3, pp. 409-422, November 1969.

[17] S. Tamura, S. Niguchi and K. Tanaka, "Pattern Classification Based on Fuzzy Relations," Systems, Man and Cybernetics, vol. SMC-1, pp. 61-66, January 1971.

[18] J. G. Brown, "A Note on Fuzzy Sets," Information and Control, vol. 18, pp. 32-39, February 1971.

[19] L. A. Zadeh, "Toward a Theory of Fuzzy Systems," in Aspects of Network and System Theory, edited by R. E. Kalman and N. Declaris, Holt, Rinehart and Winston, 1971.

[20] L. A. Zadeh, "Similarity Relations and Fuzzy Orderings," Information Sciences, vol. 3, pp. 177-200, April 1971.

[21] S. K. Chang, "Fuzzy Programs - Theory and Applications," Proc. Polytechnic Institute of Brooklyn Conference on Computers and Automata, Brooklyn, New York, 1971.

[22] R. Carnap, "The Logical Syntax of Language," Harcourt, Brace and Co., New York, 1937.

[23] N. Chomsky, Syntactic Structures, Mouton, The Hague, 1957.

[24] Y. Bar-Hillel, M. Perles, and E. Shamir, "On Formal Properties of Simple Phrase-Structure Grammars," Z. Phonetic, Sprachwiss. Kommunikationsforch, vol. 14, pp. 143-172, 1961.

[25] H. Hiz, "Congrammaticality, Batteries of Transformations, and Grammatical Categories," Proc. Symp. in Applied Math., pp. 43-50, American Math. Soc., 1961.

[26] N. Chomsky, Aspects of the Theory of Syntax, MIT Press, 1965.

[27] S. Ginsburg, The Mathematical Theory of Context-Free Languages, McGraw-Hill Book Co., 1966.

[28] Z. Harris, Mathematical Structures of Language, Interscience Publishers, 1968.

[29] E. Shamir, "Algebraic, Rational, and Context-Free Power Series in Noncommuting Variables," in Algebraic Theory of Machines, Languages and Semigroups, edited by M. Arbib, pp. 329-341, Academic Press, 1968.

[30] J. E. Hopcroft and J. D. Ullman, Formal Languages and Their Relation to Automata, Addison-Wesley, 1969.

[31] R. E. Bellman, R. Kalaba and L. A. Zadeh, "Abstraction and Pattern Classification," Jour. Math. Analysis and Applications, vol. 13, pp. 1-7, 1966.

[32] S. M. Watanabe, "Methodologies of Pattern Recognition," Proc. of the International Conference on Methodologies of Pattern Recognition, Univ. of Hawaii, 1968, Academic Press, 1969.

[33] E. T. Lee and L. A. Zadeh, "Note on Fuzzy Languages," Information Sciences, vol. 1, pp. 421-434, October 1969.

[34] L. A. Zadeh, "Quantitative Fuzzy Semantics," Information Sciences, vol. 3, pp. 159-176, April 1971.

[35] D. J. Rosenkrantz, "Matrix Equation and Normal Forms for Context-Free Grammars," Jour. Assoc. for Computing Machinery, vol. 14, pp. 501-507, 1967.

[36] E. T. Lee and L. A. Zadeh, "Fuzzy Languages and Their Acceptance by Automata," Proc. 4th Princeton Conference on Information Sciences and Systems, p. 399, Princeton University, 1970.

[37] J. A. Brzozowski, "Regular Expressions for Linear Sequential Circuits," IEEE Trans. on Electronic Computers, vol. EC-14, pp. 148-156, 1965.

[38] K. Knuth, "Semantics of Context-Free Languages," Math. Systems Theory, vol. 2, pp. 127-145, 1968.

[39] M. Mizumoto, "Fuzzy Automata and Fuzzy Grammars," Ph.D. Thesis, Faculty of Engineering Science, Osaka University, February 1971.

[40] J. Feldman, J. Horning, J. Grips, and S. Reder, "Grammatical Complexity and Inference," Artificial Intelligence Laboratory, Stanford University, AI 89, June 1969.

[41] A. Koutsoudas, Writing Transformational Grammars, McGraw-Hill Book Co., 1966.

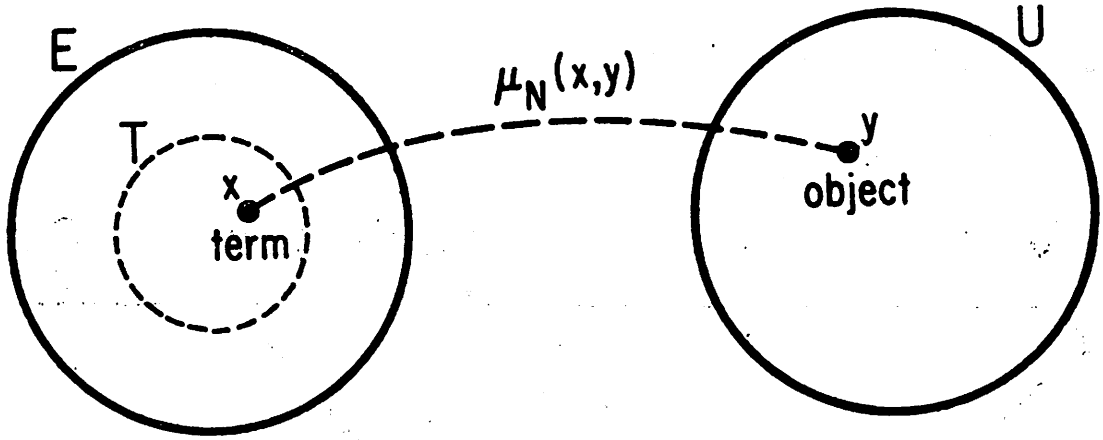

Fig. 1. The components of a fuzzy language: U = universe of discourse; T = term set; E = embedding set for T; N = naming relation from E to U; x = term; y = object in U; $\mu_N(x,y)$ = strength of the relation between x and y; $\mu_T(x)$ = grade of membership of x in T

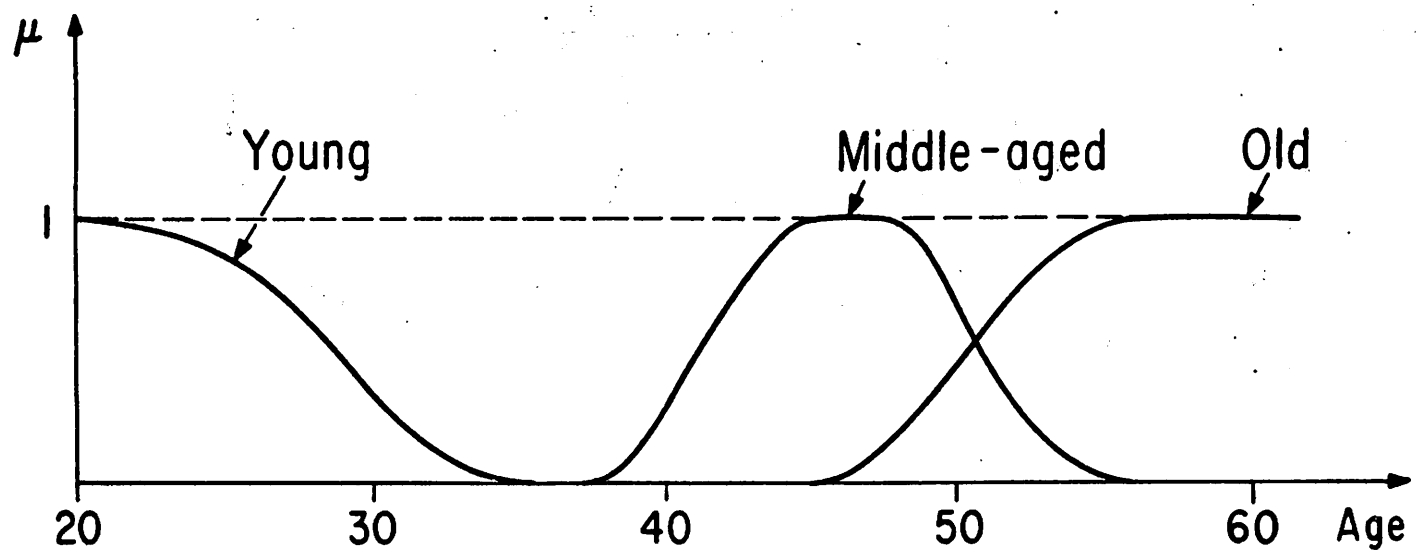

Fig. 2. Membership functions of the fuzzy sets M(young), M(middle-aged) and M(old).

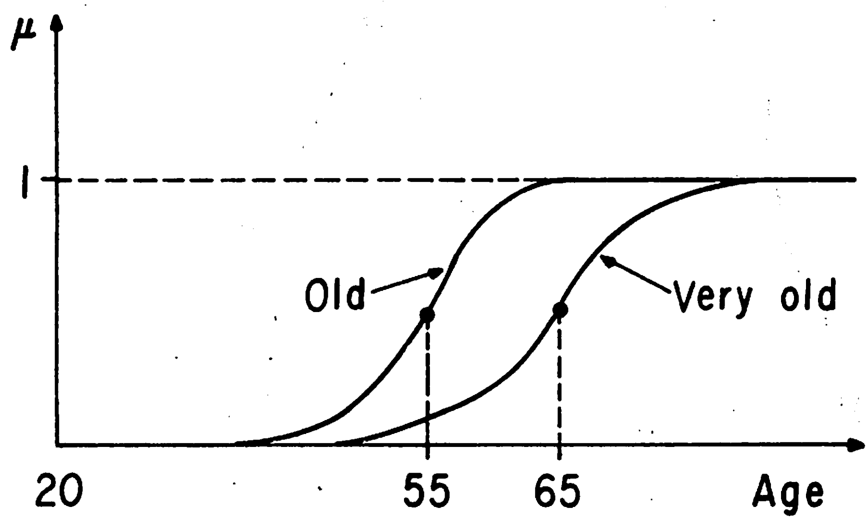

Fig. 3. The effect of the concentrator very on the fuzzy set M(old).

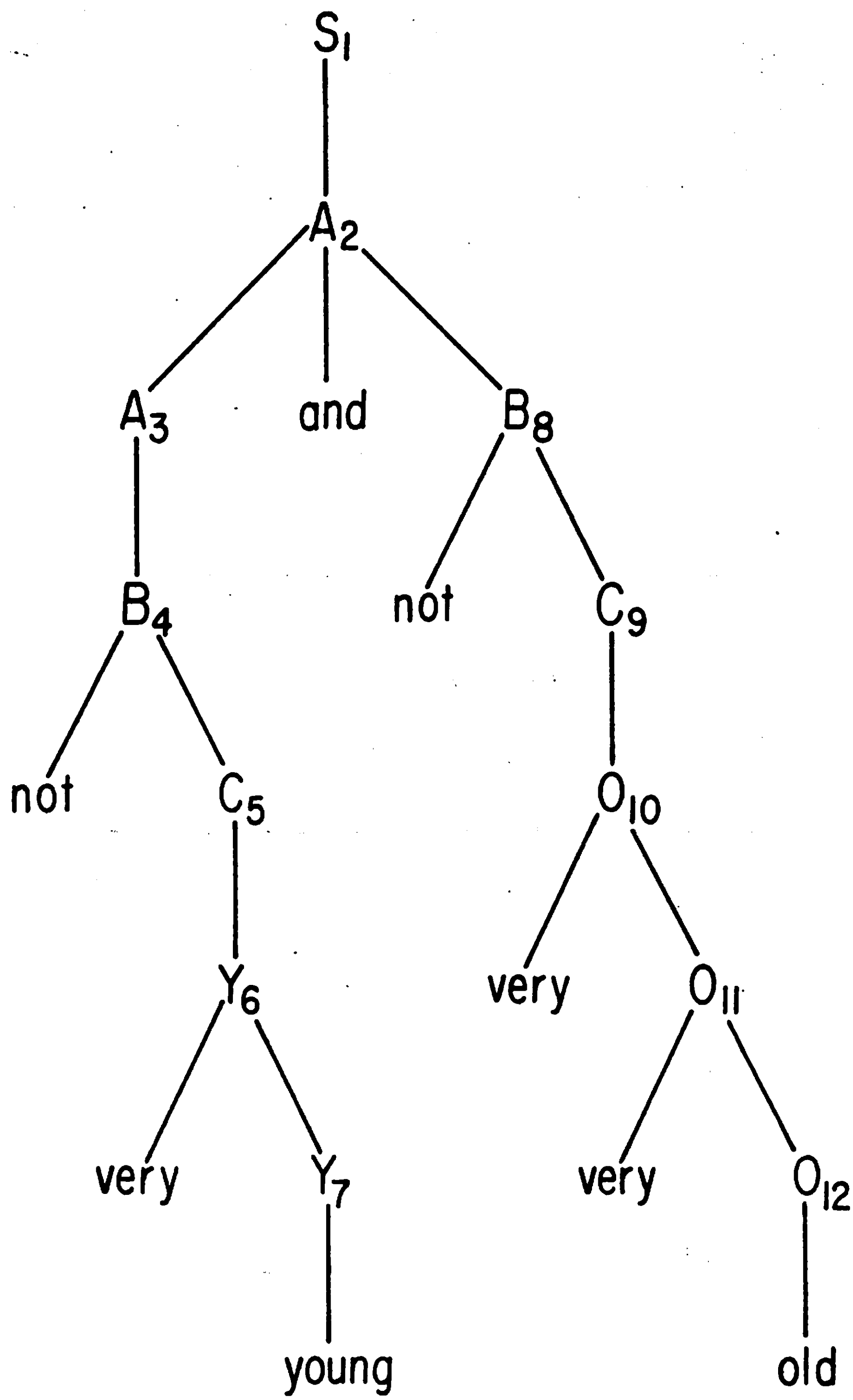

Fig. 4. Syntax tree for x = not very young and not very very old.