Language, Information and Computation(PACLIC 11), 1996, 337-346

Fuzzy statistics and computation on the lexical semantics : How much do you think? and how many?

Berlin Wu and Ching-min Sun
National Chengchi University and Jin-Wen College of Business and Technology, Taiwan
Berlin@math.nccu.edu.tw

Abstract

In this paper, we investigated the fuzzy statistics analysis in lexical semantics and apply the fuzzy logic to compute some uncertain and ambiguous problems. The fuzzy propositional computation for the cognitive semantics can account for the degree of typicality and similarity. Which provide a more precise expression in human thought and human cognition. Some essential definitions for fuzzy statistics are proposed to implement these procedures. The empirical results by a sampling survey and fuzzy statistical analysis suggests that the fuzzy statistics and computation are potentially powerful heuristics in analyzing lexical semantics.

1. Introduction

Procedures of semantic analysis may be one of the most complicated structures people have met with. Some conventional semantic theories presuppose the notion that natural language may be describable by a finite set of rules capable of generating an infinite set of sentences. The difficulty with this approach is that external features depends on being able to clearly determine for each relevant features whether or not an object processes it. Moreover, even those features which have been decided, such as 'heavy", 'short' or 'blue', might still be fuzzy since there are no clear cut boundaries distinguishing heavy from very heavy, short from a little short or blue from purple.

A fundamental problem of lexical semantics is the fact that what Ruhl (1989) calls the perceived meaning of a word can vary so greatly from one context to another. Some disadvantages about computational (numerical) semantics are: (i) the danger of overstraining the empirical data to meet the requirement of numerical precision; (ii) the danger of overinterpreting the numerical results of a term. One possible way to diminish the required amount of precision is to use fuzzy statistics. However, Zadeh (1972) and (1983) have proposed certain alternative approaches where the linguistic aspects are mostly emphasized. Since then, many papers have also been published on this topic, for examples, see Joyce (1976), Rieger (1976) and Morgan and Pelletier (1977) Sanchez et al (1982) etc. For an extensive treatment of the theory of fuzzy sets with applied linguistics the interested reader may refer to see Dubois and Prade (1980) or Manton, Woodbury and Tolley (1994).

In this paper, we will apply the fuzzy statistical analysis and computational lexical semantic method to investigate some uncertain and ambiguous problems. Especially we will discuss the degree of object's typicality and similarity, which

Especially we will discuss the degree of object's typicality and similarity, which provide a more precise expression in human cognition. . It should be pointed out that the concept of fuzzy statistical analysis applied in this research does not refer to the general notation of constructing certain theories, but to propose some alternative methods in computational lexical semantics. We hope that such analytic technique will be more reliable and significant for the future research.

2. Measuring the lexical semantics

In speaking of the semantic of a natural language, Langacker (1973, p.28) argued that we are referring not only to the fact that the words of the language have meanings but also to the way in which they divide the range of our conceptual experience into scaling. The arguments seem to demonstrate that such an analysis yields a notion of proposition which is insufficiently fine-grained to serve as the object of a human belief or a thought. Familiar considerations from lexical semantic theory cast doubt upon the conventional analysis of propositions as sets of possible words. Therefore, the fuzzy linguistic scaling used for measuring the meaning of terms measurement is necessary to be mathematically defined with membership function.

2.1 Fuzzy logic and lexical semantics

Let K be a class of generic elements called the kernel set. For example, K={the cats we have seen last week}, K={the gifts Peter received last year} or K={sages in the world history} etc.. In short K should contain all the specific objects we have met, thought or imagined. Let σ_k = the σ-filed generated by K. We called σ_k the semantic σ-field (cf. Kittay, 1992, p.237-240). For example the following lexical terms explained in the dictionary may be:

chair := a *usually* movable seat that is *essentially* designed to accommodate one person and usually has four legs and back and often has arms.

living cost : = the cost of buying the goods and services thought necessary to provide a person with the *average* accepted *standard* of living. (c.f. *Example 3.2*)

Basically, both the definitions does not seem sufficient or satisfied to the human cognition or thought. But if we make use of the semantic -field, we may reach a more concrete explanation. That is the σ_{chair} = the semantic σ-field generated by the kernel set chair, and the $\sigma_{\text{living cost}}$ = the σ-field generated by the kernel set living cost. The membership function corresponding to the object nouns could be constructed on the basis of the outer characteristics of factors, such as geometric patterns, topological properties or physical feature constrained in objects.

Because different kinds of morphemes have been found to be associated with different degrees of internal semantic sensitivity depending on whether the meaning of a morpheme is completely or partially rendered by the form of the morpheme, the component features have been arranged in the order of their fuzzy semantic value. Typically speaking, exactly one meaning to a sentence, its fuzzy semantic value will equal to 1.

Such constructions require the intervention of human thought to provide the logics and Bayesian probability, hence we can hardly assume those complicated phenomenon as measurable, not even approximately reasoning. Since that fuzzy methods are rather robust, the exact determination of the membership function is not as important as it might seem at first glance. A satisfactory definition about fuzzy measure can be found in Zimmermann (1991, p.45).

Example 2.1 The fuzzy set 'young' might be defined as :

$$\mu_{\text{young}}(x) = 1.0I_{20}(x) + .9I_{30}(x) + .8I_{40}(x) + .6I_{50}(x) + .4I_{60}(x) + .2I_{70}(x) + .1I_{80}(x);$$

where $I(x)$ is an indicator function; i.e. $I_c(x)=1$ if $x=c$, $I_c(x)=0$ if $x \neq c$. Which denotes that we adhere to the numerical age of 20 a grade of membership of the fuzzy set young of 1.0, that means 20 completely belongs to young. The age of 25 belongs with a grade of 0.9 to young, and so on.

On the other hand, the continuous membership function for the term 'young' might be defined as :

$$\mu_{\text{young}}(x)=\begin{cases} 1; & 0 < x < 25 \\ \exp\{-(x-220/20\}; & 25 \leq x \end{cases}$$

A fuzzy measurement makes use of the rating scale which contains pairs of adjectives from positive to negative (bipolar adjective) meanings. Since the statistical data provide some source of fuzzy semantic problems; in fact relevant concepts and relations can be ill-measured and vague.

2.2. Computation of semantic membership

A fuzzy quantity Q is a fuzzy set on the real numbers, i.e. a mapping $g_Q : [0, 1] \rightarrow [0,1]$. Here g_Q will naturally be viewed as a possibility distribution on the values that a variable can assume.

Thus if L is a linguistic quantifiers, such as most, then L can be represented as a fuzzy subset of L where for each t belongs to [0, 1], g_L (t) indicates the degree to which the proportion t satisfies the concept denoted by g_L. For example, let the linguistic quantifier L = some, then if $g_L(0.4)=1$ we would say that 40% is completely compatible with the idea conveyed by the linguistic quantifier some; while if g_L (0.2) = 0.8 it is indicating that the proportion of 20% is 0.8 compatible with the concept of some.

Moreover, the adverbs, e.g. very, extremely, highly, absolutely, slightly, hard, quit..., is usually called the linguistic modifier in the fuzzy set. One of the basic problems in psycholinguistic is to evaluate the meaning of a composite term from knowledge meaning of its atomic subterms. Considering here the meaning of composite terms of the form $x = h \circ n$, where n is a primary term and h is a linguistic modifier such as sort of, very, slightly etc.. The modifier h is viewed as a modifier of the meaning of n.

If f is a fuzzy set for the term n then the hedge h (modifier) generates a fuzzy set e (the term e) such that $e = m \circ t$. we define some of operator that may serve as a basis for modeling hedges:

$$\begin{cases} Normalization: & \mu_{norm(t)}(n) = \mu_{t(n)} / \sup \mu_t \\ Concentration: & \mu_{cont(t)}(n) = \mu_{t(n)}^{c}, \quad 1 < c \\ Dilation: & \mu_{dal(t)}(n) = \mu_{t(n)}^{c}, \quad 0 < c \le 1 \end{cases} \tag{2.1}$$

Foe instance, *very(n)=cont(n)=* $\mu_{t(n)}^{2}$, *more or less(n)=dia(n)=* $\mu_{t(n)}^{0.5}$, *Highly(n)=*n^3.

Thus, with the aid of linguistic hedges, a small number of basic functions can produce a wide range of models hedges. As in the case of linguistic variables, the set of possible or admissible values has thus been defined in a structural way and not by simple enumeration.

Example 2.2 Following *Example 2.1*, let us consider the term 'very young'. We might take the concentration c = 2. Then, the membership function becomes

$$\mu_{very\ joung}(n)=1.0I_{20}(x)+.81I_{30}(x)+.64I_{40}(x)+.36I_{50}(x)+.16I_{60}(x)+.04I_{70}(x)+.01I_{80}(x)$$

3 Computation of the terms relation and association

Semantic relations of a term have many characteristics in common with other concepts. Rosch's studies (1975) reported intersubjective agreements on typicality that were surprisingly high, in most cases a correlation greater than .9 A subsequent reconsideration of her statistical methods revealed that her measure of agreement was biased in that the larger populations automatically tended to produce a higher degree of agreement. The ability to perceive relations between ideas has long been taken to reflect human cognition. The most typical items in a category will be those that rank high in typicality on each features, where as the least typical will be those that rank low in typicality in each individual feature. Most of the known term relations are based on the three types: (i) semantic associations exhibit a degree of typicality. (ii) relations compared with one another. (iii) like other general terms (e.g. "cable"), association terms can be used to refer to a variety of different kind of situations and are instantiated or elaborated by their context.

On the other hand, semantic associations of an object have also long played an important role as explanatory constructs in psychological and computational linguistics. The use of association as theoretical primitives has obscured the fact that semantic associations are themselves concepts with interesting properties that are in need of explanation. The representation of objective association in cognition must be explained in terms of more basic meaning elements that are common to a variety of different concepts, c.f. Lyons (1977, p.317).

3.1 The internal structure of fuzzy subjective categories

There are many methods proposed in the literature of mathematical psychology or lexicon for scaling a subject's perception of an attribute, e.g. Nunnally (1978), Cruse (1986), Salton (1986), Dubois and Prade (1988), Hamers *et al* (1989), Ruge and Schwarz (1990) and others. Most of their research are based on co-ocurrence statistics of the terms in the text databases for which the associations are used.

For determining the semantic similarity of two objects or concepts, the set of their features must be compared. The following three schemes are proposed to measure the general term relations and associations.

(a) Typicality

A category vary in the degree to which they are typical of the concept. For example, a trout is a very typical fish, a skipper is slightly typical, a whale is less typical, and a frog is not very fish-like at all. As Medin (1989) points out, this kind of grade structure has been found for every kind of category that has been studied: taxonomic categories, formal categories, goal-derived categories, ad hoc categories, and linguistic categories. If a typical pet is a dog, then the subject must have a representation of a dog in his/her mental warehouse. Typicality is measured by asking subjects from random samples to rate how good an example a concept is of category. Typical members of a category are those that are most similar to prototype of the category. For instance, the typical dog has four lags and tail, is about 1 foot long, and runs and barks around the house. Ruge and Schwarz (1990) have all these attributes and so are similar to the prototypical dog and are judged to high in typicality. Here, I define the typicality T for any object o as

$$T(o) = \frac{\text{membership value of o}}{\text{maximum membership value in the population}} \quad (3.1)$$

Decision rules of this kind were originally proposed to account for effects of typicality on the latency of category verification. Evidence for or against category membership was based on a comparison of attributes of the stimulus concept with those of the prototype for the category. Evidence is more consistently positive for high typicality than for low typicality category members and more consistently negative for dissimilar than for similar nonmembers.

(b) General Similarity

Because people can easily make similarity judgments about relations, their judgments can be used to identify the elements that are used in comparing semantic relations. As noted before, concepts have generally been viewed as composed of more basic components. For example, it is easy to decide that a gorilla and a chimpanzee are more similar than a gorilla and a panda. Comparison requires the identification of ways in which the things compared are similar and different, e.g. shape, lags, fingers, taste etc... A description of the hyperterm system REALIST (REtrieval Aids by LInguistics and STatistical) and in more detail a description of its semantic component is given by Ruge (1992). Various experiments with different similarity measures are also presented in his paper. the similarity measure $S(o_i, o_j)$ he used is

$$S(o_i, o_j) = \frac{|H_i \cap H_j| + |M_i \cap M_j|}{|H_i \cup H_j| + |M_i \cup M_j|}, \quad (3.2)$$

where H_i, M_i, H_j and M_j are the characteristic sets of the heads and modifiers of the term pairs (o_i, o_j) are taken into account with equal weights. Table 3.1 shows the results of an experimental version of Ruge and Schwarz's (1990) approach based on the heads and modifiers from 200,000 abstracts of the U.S. Patent and Trademark Office (PTO). Those pair terms are semantically similar in a general sense. For example, synonyms like 'cable vs. wire' or 'efficient vs. economical' or 'container vs. receptacle'; antonyms like 'acceleration vs. deceleration'; broader terms like 'acceleration vs. inclination'; narrower terms like 'container vs. tank' etc.

Table 3.1 Heads, modifier, and their frequency in 200,000 abstracts from the terms similarity.

Container		cable		acceleration		efficient	
Term	*similarity*	*Term*	*similarity*	*Term*	*similarity*	*Term*	*similarity*
container	1.000	cable	1.000	acceleration	1.000	efficient	1.000
enclosure	0.466	conductor	0.333	deceleration	0.416	economical	0.466
bottle	0.466	connector	0.283	speed	0.283	simple	0.466
receptacle	0.433	wire	0.283	velocity	0.250	effective	0.433
cavity	0.433	rope	0.266	inclination	0.200	easy	0.433
vessel	0.433	rod	0.250	movement	0.166	compact	0.433
tank	0.416	line	0.233	correction	0.150	simultanious	0.416
pouch	0.400	pipe	0.216	rotation	0.150	direct	0.400
housing	0.383	unit	0.216	engine	0.083	low	0.383
compartment	0.366	chain	0.200	exhaust	0.005	utilizable	0.366

(c) Partial Similarity

If we consider that the semantic similarity of terms only depends on the specific features, we may encounter the partial similarity measurement. For this purpose I present the definition about the similarity of partial determination $PS(o_i, o_j)$ between object (o_i, o_j). They may properly display the $o_i \cdot o_j$ relationship at fixed features f.

$$PS(o_i, o_j) = \frac{\sqrt{T(o_i|f)\cdot T(o_j|f)}}{\max\{T(o_i|f), T(o_j|f), 1-T(o_i|f), 1-T(o_j|f)} \tag{3.3}$$

Example 3.1. Let the typicality gradient of fish for typicality be T(trout | hape) = 0.9, T(skipper | hape) = 0.7, T(whale | hape)=0.5 and T(frog | hape) = 0.1. By equation (2.3), their partial similarity under the condition of shape of fish is exhibited in Table 3.2

Table 3.2 Partial similarity for the shape of fish

Term	*trout*	*skipper*	*whale*	*frog*
trout	1	0.88	0.75	0.33
skipper		1	0.82	0.29
whale			1	0.25
frog				1

The computational rules make use of the rating scale which contains pairs of adjectives from positive to negative (bipolar adjectives) meanings. Since the statistical noise provided some source of fuzzy semantic problems; in fact relevant concepts and relations can be ill-measured and vague. To this purpose, the fuzzy statistics seems the most appropriate tool for handling this type of uncertainty.

3.2 Fuzzy statistical analysis for human thought

In this section we propose definition of essential fuzzy statistics and its applications in semantic measurement.

Definition 3.1 Let $S_i = (a_i, b_i)$ be the survey of fuzzy intervals, i=1, 2, ..., n. If the frequency of the lower value for a_i is f_i and the frequency of the upper value for b_i is g_i, then the fuzzy mean μ_S of the samples $\{S_i\}$ is the average of weighted sum of f_i and g_i respectively, i.e. the average fuzzy interval $\mu_S = (a, b)$, where

$$a = \frac{\Sigma f_i a_i}{\Sigma f_i} \text{ (average minimum)}, \ b = \frac{\Sigma g_i b_i}{\Sigma b_i} \text{ (average maximum)}.$$

Definition 3.2 The fuzzy expect value of the sample $\{S_i\}$is E $\mu_S = \frac{a+b}{2}$.

Definition 3.3 The fuzzy median of the samples $\{S_i\}$ is defined as $Median_S$ = (m_l, m_u) where m_l = median of $\{a_i\}$ and m_u = median of $\{b_i\}$.

Definition 3.4 The fuzzy mode of the samples $\{S_i\}$ is defined as $\text{mod}_S = (m_l, m_u)$, where m_l is mode of $\{a_i\}$ and m_u is model of $\{b_i\}$.

Empirical study 3.1 The following studies are based on survey of the respondents of Taipei metropolitan area conducted during may 1992, see Wu and Yang (1993). A total of 100 respondents were contacted in the survey with 100 completed questionnaires. The response rate is 98%.
Question: How much do you expect for the cost of living, including rent, daily expenditure, food stuff and commute for a four persons family in Taipei area.

Note that the term 'living cost' has different implications in the minds of different individuals. For example:(i) in terms of the social class (high, middle, low-class people (ii) profession: doctor, professor, manager, etc. (iii) the number of persons one has supported in his family (iv) sex and age difference: male vs female, young vs aged person. (v) society: different society will have different standard on the cost of living, e.g.: Taipei downtown, Shin-Den, Wu-Lai etc.. Table 3.3 shows the survey data.

Table 3.3 People's preference for the living cost (thousands)

Low	10	15	20	25	30	35	40	45	50	55	60	70	150
frequency	2	4	12	9	25	7	19	1	14	1	2	1	1
High	20	25	30	35	40	45	50	55	60	70	80	100	200
frequency	6	2	17	3	25	4	17	2	12	4	2	3	1
Exact	15	20	25	30	35	40	45	50	60	70	80	100	150
frequency	3	4	10	27	7	10	6	23	3	2	1	1	1

From Table 3.3 we compute the frequency for each intervals, which is shown at Table 3.4.

Table 3.4 Frequency of each fuzzy intervals

Living Cost (NT$1000)	Frequency	Relative Frequency
10-15	2	0.02
15-20	5	0.05
20-25	13	0.13
25-30	20	0.20
30-35	28	0.29

35-40	32	0.33
40-45	25	0.26
45-50	23	0.24
50-55	20	0.20
55-60	19	0.19
60-65	8	0.08
65-70	8	0.08
70-75	5	0.05
75-80	5	0.05
80-85	3	0.03
85-90	3	0.03
90-95	3	0.03
95-100	3	0.03
100-150	1	0.01

Its membership function is:

$$\mu(x) = .02I_{[10,15]} + .05I_{[15,20]}(x) + .13I_{[20,25]}(x) + .20I_{[25,30]}(x) + .29I_{[30\text{-}35]}(x) + .33I_{[35,40]}(x) + .26I_{[40,45]}(x) + .24I_{[45\text{-}50]}(x) + .20I_{[50,55]}(x) + .19I_{[55,60}(x) + .08I_{[60,65]}(x) + .08I_{[65,70]}(x) + .05I_{[70,75]}(x) + .05I_{[75,80]}(x) + .03I_{[80,85]}(x) + .03I_{[85,90]}(x) + .03I_{[80,95]}(x) + .03I_{[95,100]}(x) + .01I_{[100,150]}(x)$$

Figure 3.1 plots the distributions of membership function for the living cost.

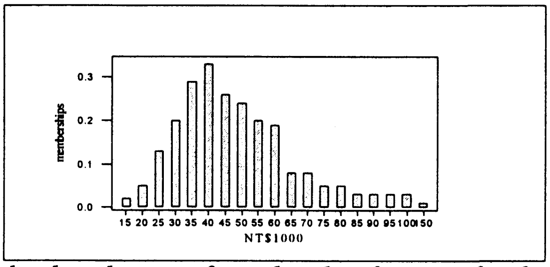

Figure 3.1 the distribution of membership function for the living cost.

Table 3.5 makes a comparison for belief measurement about the fuzzy survey and the conventional survey.

Table 3.5 Comparison results with traditional statistics

Exact	Mean = 39.9	σ = 18.2	Median = 35	Mode = 30
Fuzzy	E μ_S = 40.9	μ_S =(34.1, 46.7)	$Median_S$ =(30,40)	mod_S = (30,40)

The typicality gradients for the living cost are calculated according to equation (3.1) and are exhibited in Table 3.6.

Table 3.6 typicality of the living cost

Living Cost (NT$1000)	Typicality
10-15	0.06
15-20	0.15
20-25	0.39
25-30	0.61
30-35	0.88
35-40	1.00
40-45	0.76

45-50	0.70
50-55	0.62
55-60	0.58
60-65	0.24
65-70	0.24
70-75	0.15
75-80	0.15
80-85	0.09
85-90	0.09
90-95	0.09
95-100	0.09
100-150	0.03

Hence, from the membership function, we can get a more precise picture about those ambiguous terms in our ordinary life. In this study, we find the amount of thirty five to forty thousands (NT$-dollars) is the typicality of people's common agreement for living cost at Taipei area in 1992.

4. Conclusion

In the real world, the concepts involved in various domains of information or knowledge are much too complex and sophisticated to admit conventional logic as well as linguistic semantics. Using the fuzzy logic in analyzing the semantic system as well as measuring words sense have contributed not only to attain the identification of the situation stated above, but also exert a significant impact on the orientation of linguistic semantics. Although there are many different approaches given in the literature, each has its own advantages as well as its own drawbacks.

One of the problems in practical applications of fuzzy theory is how to obtain the membership functions and how to be sure that they do represent the meaning of the linguistic terms. In this paper, we described the fuzzy system analysis process in psycholinguistic cognition. The fuzzy propositional model for the semantic system can account for the degree of typicality and similarity. Which provide a more precise expression in human cognition. It is not difficult to imagine that there exists alternative models that do not directly involve either typicality, similarity and partial similarity membership information. The viability of such models will mostly depend on whether it is a satisfactory description of human perceptual primitives.

To this aim, some essential definitions for fuzzy statistics are proposed to implement these procedures. Empirical results of this research suggests that fuzzy modeling and statistics analysis are potentially powerful heuristics.

Finally, a neural network is a system of interconnected computational elements operated in parallel, arranged in patterns similar to biological neural nets and modeled after the human brain. Recently interest in this field has increased mainly because of the developments in many fields. We hope this direction of research would provide a useful tool in computing linguistics. In order to get an appropriate accuracy for human thought, we expect neural computing will be a worthwhile approach and may simulate more future empirical work in lexical semantics.

REFERENCE

Cruse, D. A. 1986. *Lexical Semantics*. Cambridge, England: Cambridge University Press.

Dubois, D. and Prade, H. 1980. *Fuzzy Sets and Systems: Theory and applications.* London: Academic press.

Dubois, D. and Prade, H. 1988. *Possibility Theory.* An Approach to Computerized Processing of Uncertainty. New York: Plenum.

Joyce, J. 1976. Fuzzy set and the study of linguistics. *Pac. Coast. Philol.* 11, 39-42.

Kittay, E. 992. *Semantic Fields and the individuation of Content. Frames, Fields, and Contrasts* ed. Lehrer, A and Kittay, E.. New Jersey; Lawrence Erlbaum Associates.

Kosko, B. 1992. *Neural Networks and Fuzzy System.* New Jersy: Prentice Hall.

Langacker, R. 1973. *Language and Its Structure*. 2nd. New York: Harcourt Brace Jovanovich.

Lee, E.T., and Zadeh, L. A. 1969. Note on fuzzy languages. *Inf. Sci.* 1, 421-434.

Manton,K., Woodbury, M, and Tolley, H. 1994. *Statistical Applications - Using Fuzzy Sets*. New York: John Wiley & Sons, Inc.

Medin, D. L. 1989. Concepts and conceptual structure. *American Psychologist*. 12, 1469 - 1481.

Mongan, C., and Pelletier, F. 1977. Some notes concerning fuzzy logic. *Linguist. Philos*. 1, No 1. 11-23.

Nunnally, J.M. 1978 *Psychometric Theory.* 2nd ed. New York: Mcgraw-Hill.

Rieger, B. 1976. Fuzzy structural semantics: on a generative model of vague natural language meaning. *Eur. Meet. Cybern. Syst. Res.*,3rd, Vienna.

Ruge, G. 1992. Experiments on linguistically-based term associations. *Information Procession & Management.* 283, 317-332.

Ruhl, C. 1989. *On Monosemy.* New York: SUNY Press.

Sanchez, E., Gouvernet, J., Bartolin, R. and Vovan, L. 1982. Linguistic approach in fuzzy logic of the who classifications of dyslipoproteinemias. in *Fuzzy Set and Possibility Theory* [Yager, R. ed.], Oxford:Pergamon Press.

Wu, B. and Yang, W. S. 1993. Measuring beliefs: an applications of fuzzy sets to social and economic analysis. *1993 Far-Eastern Meeting of the Econometric society,* Taipei.

Zadeh, L. 1972. A Fuzzy set-theoretic interpretation of linguistic hedges. *J. Cybern.* 2, No.3, 4-34.

Zadeh, L. 1983. *A computational approach to fuzzy quantifiers in natural languages. Computers and Mathematics*, 9, 149-184

Zimmermann, H.-J. 1991. *Fuzzy Set Theory - and Its Applications*. 2nd. Kluwer-Nijhoff, Boston.