ABOUT SEMANTICAL STRUCTURE OF LINGUISTIC HEDGES : AN EXPERIMENTAL HYPOTHESIS

V.B.KUZ'MIN (MOSCOW, USSR)

The hypothesis, infered from experimental data, about the ordinal structure of the verbal estimate scale of the type "modifier - term" for linguistic variables is proposed.

As it is represented there are two important problems. The first is the problem of the measurement of the membership degrees, the second one is treated as a problem of the description of the hedges (and connectives). Further we consider the problem of the description of the hedges.

For the calculation of the composite expression by means of the fuzzy sets theory we can use the descriptions of the hedges and the connectives similar to them which was proposed by L.Zadeh [1]. But as it may be seen from the papers [2 - 5] , expressions calculated in this manner are rather different from experimental results. Because of this one would like to have experimentally supported representation about both the general idea of linguistic estimate scale and of each specific hedges in distinct contextes.

It will be recalled some resultes related to the problem at hand. Mosier [6] suggested that the evaluative adjectives are arranged along an unidimensoinal continuum. He experimentally demonstrated that the intensificators like "very" induce a shift of the meaning of the modifaing word from the neutral point to the extremum. In his model Cliff [7] made use of the distance from the neutral point of the scale. Cliff's experiments showed that this distance is functionally dependet on the hedges.

L.Zadeh [1] described the types of effects (dil, int, con, neg) on the curve of the membership function of the term which are caused by the implementation of the hedges. Lakoff [8]developed the typology of the meaning components (def, prim, sec and char) defining the hedges. The deep analysis made by Lakoff confirmed Mosier's conclusion about the character of the operation "very" of the meaning of the term (see also [2]).

In the paper [9] " a sufficiently rich set F of verbal expressions" is divided on two parts: the first one is composed of couples $A^-, A^+ \in F$ of antonymous expressions such as "young, old" and the second one is the complement: $F \setminus \{(A^-, A^+)\}$. Three point $m \leq s \leq v$ are separated from the ordered support $\mathcal{U}$. The point s is called by author "the semantical center of $\mathcal{U}$ " and it plays a role of neutral point whi le the points m and v are extremal. The effect of the hedges is described by the model in which the term is shiffed from s to a proper extremal point. In conclusion author manifestes the confidence in that "there really exist some general laws in natural language semantics which can be quantified".

I hope that the formulated below hypothesis will be move us one little step further to the understanding of thies laws.

Let us first consider the experimental data. Three levels of intensity were assigned for each specific linguistic variable in the conducted experiment:

$\mathcal{T}$ - the water temperature for bathing in the Black Sea,

$\mathfrak{R}$ - rendering the material of one scientific edition in other,

$\mathcal{A}$ - actuality of the scientific edition.

Their term-set were respectively

$T_{\mathcal{T}} = \{t_1 = \text{cold}, t_2 = \text{fresh}, t_3 = \text{warm}\}$,

$T_{\mathfrak{R}} = \{t_1 = \text{partial}, t_2 = \text{essential}, t_3 = \text{complete}\}$,

$T_{\mathcal{A}} = \{ t_1$ = out of date, t_2 = up-to-date, t_3 = perspective$\}$.

The following hedges were used:

$M = \{$ very, quite, more or less, rather, absolutely, not, not very, very not, more or less not, not absolutely, rather not $\}$.

For each linguistic variables the membership functions of the terms and of the composite expressions "modificator-term" were constructed by the method discussed in [5] with the help of the panal of ten persons. It have to be recalled that according to the method the membership functions of the terms and their modifications are of triangular form or of S-shaped form. For each variable the experimental data were averaged by the rules described in [10] and constructed their unimodal membership functions. The crisp scales were constructed on the corresponding supports $U_{\mathcal{T}}$, $U_{\mathcal{R}}$ and $U_{\mathcal{A}}$ using the everaged data.

The marks of the terms $t_i, i = 1,2,3$ and their modifications $mt_i, m \in M$ were arranged on these scales in the points $u \in U : \mu_{t_i}(u) = 1$. Let us note that such scales may be constructed on averaged data by means of any suitable metric (for example $d(A,B)$, $\delta(A,B)$ or $\varepsilon(A,B)$, see [11]). The scale of verbal expressions for variable $\mathcal{A}$ is shown in Fig.1 where the point of support $u = 0$ stands for year of issue.

Further on the base of such linear scales three ordinal scales P_{t_i} ($i = 1, 2, 3$) were constructed for every variable $\mathcal{T}, \mathcal{R}, \mathcal{A}$. and for every term (see tables 1, 2 and 3). In the table 1 (an variable $\mathcal{T}$) the arrangement of the verbal marks for the terms <u>cold</u> and <u>fresh</u> (P_{t_1} and P_{t_2}) is changed by the inverse ordering ($P^{-1}_{t_1}$ and $P^{-1}_{t_2}$). In the table 2 ($\mathcal{R}$) this was done for the term <u>weak</u> ($P^{-1}_{t_1}$), in the table 3 ($\mathcal{A}$) for the terms <u>up-to-date</u> and <u>perspective</u>. The circuled marks were disregardet in further treatment because they occured only in combinations with one of three terms.

Table 1. Linguistic variable $\mathcal{T}$

ordinal scale of verbal expressions for the terms of $\mathcal{T}$			rankings			a priori (mean) ranking for $\mathcal{T}$
mt_1	mt_2	mt_3	$P^{-1}_{t_1}$	$P^{-1}_{t_2}$	P_{t_3}	
very	absol.	very	1	2	1	1
absol.	very	absol.	2	1	1	2
t_1	t_2	t_3	3	3	3	3
rather	mor/les	rather	4	-	4	4
mor/les	not absol.	mor/les	5	4	5	5
not absol.	mor/les not	(not very)	6	5	6	6
mor/les not		not absol.	7	6	7	7
not		mor/les not	8	-	8	8
(absol. not)		not				
		(very not)				

Table 2. Linguistic variable $\mathcal{R}$

ordinal scale of verbal expressions for the terms of $\mathcal{R}$			rankings			a priori (mean) rankings for $\mathcal{R}$		
mt_1	mt_2	mt_3	$P^{-1}_{t_1}$	P_{t_2}	P_{t_3}	r_1	r_2	r_3
very	very	very	1	1	1	1	1	1
mor/les	rather	t_3	3	2	3	2,5	3,5	2
rather	t_2	rather	4	3	2	2,5	2	3
t_1	mor/les	mor/les	2	4	4	4	3,5	4
not	mor/les not	not	-	5	7	5,5	5	5,5
not very	not absol.	not absol.	-	6	5,5	5,5	6,5	5,5
(rather not)	not very	mor/les not	6	7	8	8	8	-
	not	not very	5	8	5,5	7	6,5	7
	(very not)							
	(absol.not)							

With the first sight on these tables one may have an impression th t the ordinal scales of verbal expressions for each specific linguistic variable are consistent. Two methods of nonparametric statistics [12] for analysis of incomplete rankings with

Table 3. Linguistic variable $\mathcal{A}$

ordinal scale of verbal expressions for the terms of $\mathcal{A}$			rankings			a priori (mean) ranking for $\mathcal{A}$
mt_1	mt_2	mt_3	P_{t_1}	$P^{-1}_{t_2}$	$P^{-1}_{t_3}$	
very	very	very	1	1	1	1
quite	quite	t_3	2	2	3	2
t_1	t_2	quite	3	3	2	3
not very	rather	mor/les	4	5	6	4,5
mor/les	not very	rather	5	6	4	4,5
mor/les not	mor/les	not very	6	7	7	7
not	not	not	7	4	5	6
rather		rather not	8	-	8	8
rather not						

coinciding ranks were used to check the original hypothesis about consistency of initial rankings of verbal expressions. The first method is based on M.J.Prentice's statistics [13] (let us denote it C), the second - on statistic B proposed in [14] . Resulting from these two method data allows to make a conclusion about high degree of consistency of the order scale of verbal expressions for linguistic variables under consideration. Moreover a priori avarage rankings for each variable were compared with the whole totality of the initial rankings and their consistency was estimated by means of suitable for this case statistics L [12] . Results of calculations are also shown in table 4. They enable us to come to the conclusion about a very good agrement between the initial rankings and a priori one.

As it seems there is no need to comment on the results for checking the consistency of data, which allow us to make the following hypothesis.

Let the set $T = \{t\}$ of the terms - the verbal expressions of an linguistic variable $\mathcal{L}$ - is defined on the universe

Table 4

ling. variab.	statistics C			statistics B			statistics L			
	H	$\alpha\%$	Δ	H	$\alpha\%$	Δ	H		$\alpha\%$	Δ
$\mathcal{T}$	18.663	1	9.0	18.44	1	9.0		3.950	1	2.326
$\mathcal{R}$	9.997	1	9.0	16.75	1	9.0	r_1	3.830	1	2.326
							r_2	3.470	1	2.326
							r_3	1.756	5	1.645
$\mathcal{A}$	6.669	5	6.25	17.87	1	9.0		6.25	1	2.326

where H - calculeted value of statistics,

$\alpha\%$ - significance level,

Δ - table value.

$\mathcal{U} \subseteq$ Real numbers. In the common case the terms are ordered naturally respectively to the intensity levels of $\mathcal{L}$. Let $M = \{m\}$ be the set of modificators (linguistic hedges) defined on $\mathcal{L}$, $m : [0,1] \to [0,1]$. We will map the meaning of the composite expression mt , $t \in T$, $m \in M$ to the point u of support $\mathcal{U}$ such that the membership function $\mu_{mt}(u)=1$. In the most cases it is possible to agregate the group description for the verbal expressions meaning in such a way that the point $u \in \mathcal{U}$ will be unique. Let us consider the scale of commonly used verbal expressions mt for every $t \in T$ over all $m \in M$. The term t is a member of the set $\{mt\}$. The decribed observations lead to the following hypothesis.

Hypothesis. There is a point $u \in \mathcal{U}$ on a support $\mathcal{U}$ of a variable $\mathcal{L}$ (which according to above-mentioned authors we may call zero, neutral, semantical center for $T = \{t\}$) which divides the set T on left t^- and right t^+ terms and such that 1) the order P_{t^+} (P_{t^-}) of arrangement of combinations $m_i t^+$ ($m_i t^-$), $i \in I$ for all t^+ (t^-)

coincides with high significance level in a sense of suitable statistics almost for all m_i, 2) $P_{t^+} = P_{t^-}^{-1}$ almost for all m_i, where P^{-1} is the inverse order of P.

The hypothesis has some attractive aspects : it combines almost all related.factors earlier menthioned by other researchers and it is very simple. These merits make this hypothesis deserving further experimental verifications on more rich experimental data.

References

1. Zadeh L.A., The concept of a linguistic variable and its application to approximate reasoning. Information Sciences, I, v.8, pp 199-249; II, v.8,pp301-357; III,v.9,pp 43-80 (1975).
2. Hersh H.M., Garamazza A., A fuzzy set approach to modifiers and vagueness in natural language. J. of Experimental Psychology: General, 105, 1976, n.3, 254 - 276.
3. MacVicar - Whelan P.J., Fuzzy aets, the concept of height and the hedge "very". IEEE Transactions on SMC, 8, 1976, n.6, 507 - 511.
4. Badre A.N., On the precision of adjectives which denote fuzzy sets. J.of Cybernetics, 4, 1974, n.1, 49 - 59.
5. Kuz'min V.B., A parametric approach to description of linguistic values of variables and hedges. Fuzzy Sets and Systems, 6, 1981, n.1, 27 -41.
6. Mosier C.I., A psychometric study of meaning. J. of Social Psychology, 13, 1941, 123 - 140.
7. Cliff N., Adverbs as multipliers. Psychological Review, 66, 1959, 27 - 44.
8. Lakoff G., Hedges: a study in meaning criteria and the logic of fuzzy concepts. J.of Psychological Logic,2,1973,458 -508.

9. Novák V., A note on linguistic hedges and the hedge "very". BUSEFAL, n.20,1984, 5 - 9.

10. Dubois D., Prade H., Operations on fuzzy numbers. Int. J. of System Sci., 9, 1978, n.6, 613 - 626.

11. Kaufmann A., Introduction a la Théorie des Sous - Ensembles Flous,1 : Elements Theoretiques de Base. Paris, France.

12. Miskina N.P., Teiman A.I., To the construction of the system of the programs for the treatment of statistical data by the nonparametric methods. In: Nonnumerical data analysis in system studies. Collections of papers,v.10, M., All - union Institute of System Studies. 1982, 35 - 43.

13. Prentice M.J., On the problem of in incomplete rankings. Biometrica, 66, 1979, pt.1, 167 - 170.

14. Benard A., Van Elteren Ph.A., A generalization of the method of m ranking. Konink Ned. Akad. van. Wet, 1953,A,56,358 - 369.

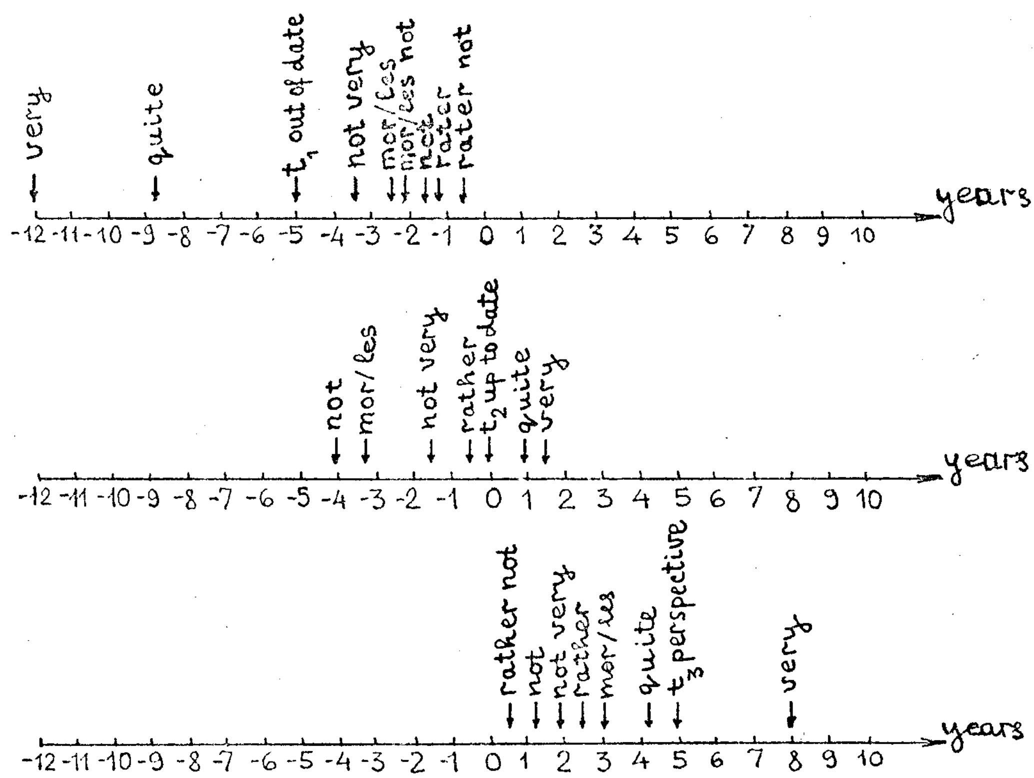

Fig. 1.