

The Fuzziness of Verbal Response Scales: The STAI-T Questionnaire

Franziska Bocklisch¹ Steffen F. Bocklisch² Josef F. Krems¹

¹Department of Psychology, Chemnitz University of Technology

²Department of Electrical Engineering, Chemnitz University of Technology

Abstract

In this paper we used a two-step procedure for the numerical translation of verbal frequency expressions of the response scale of a questionnaire (STAI-T). In an empirical study, 70 participants estimated numerical equivalents for verbal frequency expressions, data was modeled, and fuzzy membership functions were calculated. Results show that the scale's visual arrangement does not influence the interpretation of the words' meanings. We argue that traditional statistics are inappropriate for the analysis of verbal response data and demonstrate the alternative of fuzzy analysis, providing an example.

Keywords: verbal response scale, parametric fuzzy membership function, translation procedure, STAI-T questionnaire

1. Introduction

In the social sciences and humanities, empirical research questions are commonly addressed by directly interrogating participants' attitudes or behaviors. For this purpose, questionnaires are widely used as an appropriate measurement tool. Most often, participants are asked to respond to presented items by choosing a category of a related response scale. Afterwards, the research data are analyzed statistically. Although response scales determine primary data collection directly, only little systematic research has been done on these, for instance, compared to the construction of questionnaire items. One of the most popular scales is the Likert-type response scale. It commonly consists of five to seven response categories labeled with linguistic terms and integers (e.g., probability from 1 = *very improbable* to 5 = *very probable*). Linguistic expressions assigned to the scale's categories are highly comprehensible, even for subjects who are not accustomed to expressing their opinion on a response scale.

Scale categories are assumed to have a rank order but the distance between intervals cannot be presumed equal [1]. A counterargument might suggest that lack of equidistance in response categories is compensated for by the conventional visual arrangement of scales. However, the question of whether visual arrangement has an influence on the interpretation of the words' meanings has not yet been clarified empirically. Moreover, Likert scale responses comprise ordinal- but not interval-level

data and thus parametric statistical analysis cannot be employed. It is a common practice to disregard the fact that mathematical operations, such as multiplication or division, are not valid for ordinal data and: "Therein lies the sin: simply expressing ordinal data using integers does not justify the use of parametric statistics." [2] (S. 266).

Attributing interval-level qualities to Likert categories and employing inappropriate statistical techniques may lead to the misinterpretation of results and wrong conclusions [1]. The substitution or replacement of Likert scales with numerical scales (e.g., rating scales from 0 to 100) is a suitable alternative for subjects in academia and experts but not for field studies because numerical scales are too abstract and differentiated for less-educated subjects [3].

The fact that linguistic terms are fuzzy in nature [4] is widely neglected in discussions about Likert response scales. An important result of efforts to translate linguistic terms into numbers [5, 6, 7] is that verbal expressions may be conceptualised as fuzzy sets and can be mathematically described using fuzzy membership functions (MFs). We argue that Likert scales should be employed where appropriate and that respective linguistic categories should be treated as fuzzy classes instead of simply being assigned integers. Furthermore, in our opinion a fuzzy analysis of the response data is more suitable than a statistical one. If, however, parametric statistics must be employed for some reason, problems with scale levels should be minimized by carefully choosing response categories with a view to forming an equidistant distribution.

Therefore, we briefly present a two-step procedure for the numerical translation of linguistic terms [6, 7] that can be used to test and construct Likert scales with nearly equidistant response categories. This general procedure is composed of an (1) empirical study where participants estimate numerical equivalents for linguistic expressions and the (2) modeling of the data using parametric fuzzy MFs of the potential type. We used the procedure to explore the 4-point Likert scale of the State-Trait Anxiety Inventory, version T (STAI-T) [8] exemplarily and present data of 70 participants who took part in our empirical study. To rule out the argument that the visual arrangement of response scales might suggest equidistance between response categories, we manipulated the visual presentation format. We then outline the fuzzy analysis of response data and dis-

cuss this as a potential alternative to common statistical analyses.

2. Method

Here, we present details of the two-step translation procedure for the numerical translation of the four verbal frequency expressions used in the Likert response scale of the psychological questionnaire STAI [8]. The STAI is a widely used self-report measure of anxiety. It is composed of two subscales to measure anxiety as trait (STAI-T) or state (STAI-S). Answer scales of the STAI consist of 4-point Likert scales with the following verbal frequency labels (original in German): *almost never* (*fast nie*), *sometimes* (*manchmal*), *often* (*oft*) and *almost always* (*fast immer*). We chose STAI-T for our experimental study. In the following we outline the estimation technique and method applied in the empirical study. Thereafter, fuzzy analysis and MFs are specified.

2.1. Two-step Translation Procedure

2.1.1 Step One: Empirical Study

Seventy undergraduate students (18 males) of Chemnitz University of Technology with an average age of $M = 21.3$ years ($SD = 2.6$) took part in the study. The survey instrument was a paper questionnaire and consisted of two parts. In the first part participants answered the 20 items (e.g., “I am cheerful.”) of the STAI-T by marking the response category (e.g., *sometimes*) that best described their feelings. The visual arrangement of the response scale was manipulated between subjects. One group of participants ($N = 37$) received the questionnaire with the original presentation format and the other ($N = 33$) obtained an alternative format (see Figure 1). If the visual arrangement of the response scale has an influence on the interpretation of the linguistic terms, then the original presentation format should lead to more equidistance between the response categories compared to the alternative format.

Original presentation format

	Almost never	Sometimes	Often	Almost always
I am cheerful. (Ich bin vergnügt.)	1	2	3	4
I feel like crying. (Mir ist zum Weinen zumute.)	1	2	3	4

Alternative presentation format

I am cheerful. (Ich bin vergnügt.)	Almost never, Sometimes, Often, Almost never
I feel like crying. (Mir ist zum Weinen zumute.)	Almost never, Sometimes, Often, Almost never

Fig. 1: STAI response scale presentation formats

Each participant randomly received one of the presented formats and marked their answers with crosses. In the second part subjects were requested to estimate three numerical values corresponding to each of the four verbal frequency expressions. The three numerical values were: (1) the one that best represented the given frequency, (2) the minimal and (3) maximal values of correspondence. The first value identified the typical numerical equivalent to the word, whereas values (2) and (3) indicate lower and upper boundaries of the frequency expression. The instructions required that the estimates should be given in frequency format referring to 100 cases (e.g., “I am *sometimes* cheerful” means “in X out of 100 cases”). This format was used based on the rational that it is a natural mode, which directly represents information, compared to that of a percentage form [9]. Thereafter, empirical estimates for these three values were used to fit fuzzy MFs.

2.1.2 Step Two: Fuzzy Analysis

MFs are truth-value functions. The membership value (μ) represents the value of truth that an object belongs to a specific class. For the analysis of the empirical data provided by the 70 participants, a parametric fuzzy MF of the potential type [6, 7, 10, 11] was used (see Figure 2).

$$\mu(u) = \begin{cases} \frac{a}{1 + (\frac{1}{bl} - 1)(\frac{r-u}{cl})^{dl}} & u \leq r \\ \frac{a}{1 + (\frac{1}{br} - 1)(\frac{u-r}{cr})^{dr}} & u \geq r \end{cases}$$

Fig. 2: Parametric MF of potential type

This function is based on eight parameters:

- r marks the position of the arithmetic mean of the empirical estimates of the typical values.
- a represents the maximum value of the MF and regarding class structure, a expresses class “weight” in the given structure.
- c_l and c_r characterize left- and right-sided expansions of the class (equal to the statistical range)
- b_l and b_r assign left- and right-sided membership values at the boundaries of the function and represent border membership.
- d_l and d_r specify continuous decline of the MF starting from the class centre, and are denoted as representative of a class determining the shape and fuzziness.

In this study we used $a = 1$ such that all frequency terms were weighted equally. The parameters c , b and d were calculated from the distribution of the empirical data using Fuzzy Toolbox Software [11], which is specialized for fuzzy analyses and modelling of MFs. A continuous variation of MFs, ranging from highly fuzzy to crisp, is available through this parametric function type.

In contrast to the non-parametric individualized MF approach of Budescu and colleagues [5, 12] we fit group MFs to obtain a generalized model of a certain

population of subjects. Furthermore, our MFs are defined continuously such that in addition to the expansions of the class (*c* parameters) the MFs' shape (*d* parameters) carries information about the distribution of the empirical estimates. This is an advantage insofar as potential contradictions between participants' estimates are considered (e.g., in contrast to a triangular MF type that describes the graded interval between $\mu = 0$ and $\mu = 1$ with a rather arbitrary linear model and thus would not account for the empirical data). Our MF type also allows for asymmetry in fuzzy classes by providing individual parameters for the left- and right-hand branches of the function. As previous research shows [5, 6], many verbal expressions are best described by asymmetric MFs. Therefore, we expect this feature to be important. For a more detailed specification of the function type and its parameters see [6, 7].

3. Results

We first present the descriptive statistics of the data set. Thereafter, the fuzzy MF procedure is specified. In our opinion, it is valuable to present both results for purposes of comparison, even though we favor the second approach. It is important that the two approaches be understood independently. Moreover, fuzzy analysis and modeling of the MFs by definition do not refer to the background of probability theory and statistics. We then outline the fuzzy analysis of the STAI-T questionnaire response data.

3.1. Descriptive Statistics

Table 1 shows the typical values that corresponded to the frequency expressions presented. Both presentation formats (original vs. alternative) are depicted. Minimum and maximum estimates of the semantic meaning of linguistic terms were necessary for the modeling of the MFs (*c* parameters; see Section 3.2.).

STAI expressions (Original German)	Mean	SD	Skew	Kurtosis
Original format:				
Almost never (Fast nie)	7.3	3.9	0.3	-0.5
Sometimes (Manchmal)	33.5	17.6	1.2	2.0
Often (Oft)	68.5	14.3	-1.1	2.2
Almost always (Fast immer)	84.8	7.3	-0.4	-0.2
Alternative format:				
Almost never (Fast nie)	7.6	3.6	-0.4	-0.9
Sometimes (Manchmal)	30.9	15.2	0.3	0.9
Often (Oft)	67.9	9.8	-0.7	0.9
Almost always (Fast immer)	87.3	6.3	-0.8	0.5

Table 1: Descriptive statistics for the estimates

At first glance, results show that the mean values of the original and alternative presentation formats are almost equal. The largest difference is 2.6 for values corresponding to *sometimes*. The other statistical parameters also show the same pattern (e.g., smaller standard deviations for linguistic terms at the scales' borders such as *almost never* and *almost always* and larger SDs for terms mid-scale, such as *sometimes* and *often*). Therefore, we decided to aggregate the results (Table 2) for further analyses.

STAI expressions (Original German)	Mean	SD	Skew	Kurtosis
Almost never (Fast nie)	7.4	3.7	0.02	-0.8
Sometimes (Manchmal)	32.3	16.5	0.9	1.4
Often (Oft)	68.2	12.3	-1.0	2.3
Almost always (Fast immer)	86.0	6.9	-0.6	-0.1

Table 2: Descriptive statistics for aggregated estimates (N=70)

3.2. Fuzzy MF

Figure 3 shows the resulting MFs for the STAI response scale (i.e., aggregated estimates). The representative values (*r*) indicate the peak of the MFs (see vertical lines) and therefore mark the highest membership value. The *r* values are identical to the reported statistical mean values in Table 2. The MFs differ slightly in terms of shape and overlap (e.g., the MFs for *often* and *almost always* overlap more than *almost never* and *sometimes*).

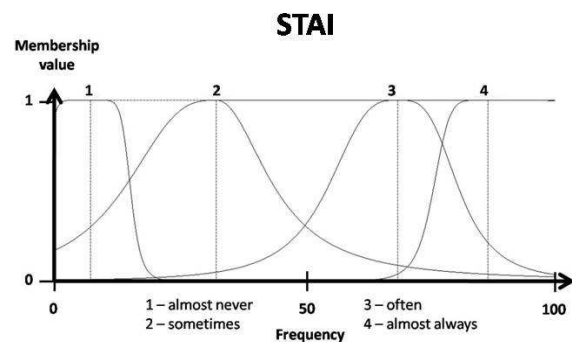


Fig. 3: MFs of the verbal frequency expressions (STAI)

The MFs' overlaps allow conclusions about the semantic meaning of the linguistic terms [6, 7] such that a large overlap reveals similarity in meaning compared to a smaller overlap. Therefore, for instance, *sometimes* and *often* are more distinct than *often* and *almost always* (Figure 3). The difference between the *r* values of the functions determines the possible equidistance of the scale (see Figure 4). Here, the distances differ considerably and, therefore, the linguistic terms are not distributed equidistantly.

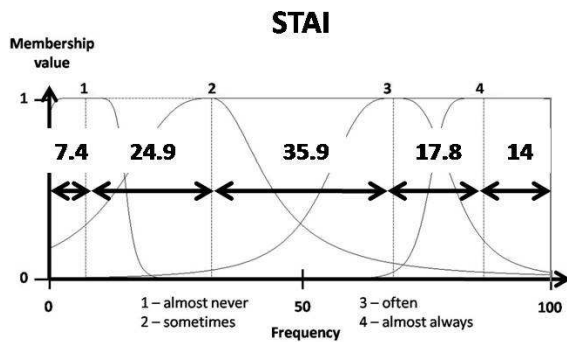


Fig. 4: Distances between the r values of the MFs

3.3. Fuzzy Analysis of Response Data

Beginning with the MFs (see Figure 3) we outline the fuzzy analysis of STAI-T response data exemplarily and compare it to the practice of common statistical analysis. The STAI-T consists of 20 items (see exemplar items in Figures 1 and 5) to measure anxiety as a trait.

The example in Figure 5 shows a fictive response pattern for the STAI-T. The following analyses were carried out for this example.

Im folgenden Fragebogen finden Sie eine Reihe von Feststellungen zur Selbstbeschreibung. Bitte wählen Sie aus den vier Antwortmöglichkeiten diejenige aus, die angibt, wie Sie sich im Allgemeinen fühlen.				
	fast nie	manchmal	oft	fast immer
Ich bin vergnügt.	1	✗	3	4
Ich werde schnell müde.	1	✗	3	4
Mir ist zum Weinen zumute.	✗	2	3	4
Ich glaube, mir geht es schlechter als anderen Leuten.	✗	2	3	4
Ich verpasse günstige Gelegenheiten, weil ich mich nicht schnell genug entscheiden kann.	✗	2	3	4
Ich fühle mich ausgeruht.	1	✗	3	4
Ich bin ruhig und gelassen.	1	2	✗	4
Ich glaube, dass mir meine Schwierigkeiten über den Kopf wachsen.	✗	2	3	4
Ich mache mir zu viele Gedanken über unwichtige Dinge.	1	✗	3	4
Ich bin glücklich.	1	2	3	✗
Ich neige dazu, alles schwer zu nehmen.	1	✗	3	4
Mir fehlt es an Selbstvertrauen.	✗	2	3	4
Ich fühle mich geborgen.	1	2	✗	4
Ich mache mir Sorgen über mögliches Missgeschick.	✗	2	3	4
Ich fühle mich niedergeschlagen.	✗	2	3	4
Ich bin zufrieden.	1	2	3	✗
Unwichtige Gedanken gehen mir durch den Kopf und bedrücken mich.	1	✗	3	4
Enttäuschungen nehme ich so schwer, dass ich sie nicht vergessen kann.	1	✗	3	4
Ich bin ausgeglichen.	1	2	3	✗
Ich werde nervös und unruhig, wenn ich an meine derzeitigen Angelegenheiten denke.	✗	2	3	4

Fig. 5: Fictive response pattern of the STAI-T (grey lines highlight positive response items)

Usually, the sum of reported answers (integers) given on the Likert response scale indicates how anxious a person is. This sum is then compared to averaged normative values of certain populations (e.g., either healthy individuals or those with anxiety disorder) for the purposes of drawing conclusions, for instance, about the need for therapeutic interventions. Comparison norms (i.e., standard values) are found by averaging the sum scores of many subjects' questionnaire responses ($N > 2300$ for the German version of STAI). As we argued in the Introduction, this procedure is questionable due to the ordinal scale level of the Likert response scale.

3.3.1 Item polarity

This study employed two types of items with regard to item polarity or phrasing: (1) positive items (e.g., "I am cheerful.") and (2) negative ones (e.g., "I feel like crying."). The STAI-T involves 7 positive and 13 negative items. Figure 5 shows the German items with positive responses highlighted in grey. The interpretation of a subject's response depends on item polarity such that high anxiety is indicated if a high frequency (e.g., 4 = *almost always*) is chosen for a negative item but low anxiety if chosen for a positive item. In order to obtain a sum that can be interpreted as an anxiety score, positive item responses have to first be inverted. This procedure was transferred for the fuzzy analysis of the response data by inverting the MFs shown in Figure 3. That is, the inverted r values were calculated by subtracting the r values (Figure 3) from the endpoint of the scale (e.g., $r_{\text{often inverted}} = 100 - r_{\text{often}}$). The other parameters were obtained by exchanging the left- and right-sided values of the original parameters (e.g., $b_{\text{I often inverted}} = b_{r \text{ often}}$). The resulting inverted fuzzy MFs' scale of the STAI is presented in Figure 6.

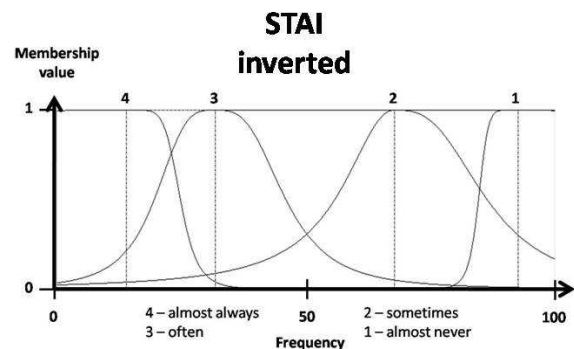


Fig. 6: MFs for the inverted STAI

3.3.2 Aggregation of MFs

To obtain a total score of anxiety for a certain participant, traditionally, integers of the marked answers on the Likert response scale are summated after inverting positive-item responses (see Figure 5, marked response integers were aggregated). In our example the sum is 31 and the possible sum score ranges from 20 to 80. We transferred the general idea of this procedure to fuzzy analysis by aggregating the MFs (see Figures 3 and 6) of the marked answers. For the example (see Figure 5) we aggregated the MFs that corresponded to the chosen response categories of the 20 items. Therefore, 20 MFs were aggregated: eight times *almost never*, five times *sometimes*, two times *sometimes-inverted*, two times *often-inverted* and three times *almost always-inverted*. The parameters of the MFs are calculated as follows: the r_{fus} (representative value of the merged MF) is the weighted arithmetic mean of the r values of the k MFs (the STAI-T has 20 items therefore $k = 20$ in this case) that are aggregated. The weight w is equal to the number of responses per category (e.g., in the example questionnaire shown in Figure 5: $w = 8$ for *almost never*). The equations for the computation of the r parameters are shown in Figure 7a.

$$r_{fus} = (\sum_{k=1}^K w_k \times r_k) / \sum_{k=1}^K w_k$$

Fig. 7a: Equation for the aggregation of r parameters

The other parameters (b_l , b_r , c_l , c_r , d_l , d_r) are aggregated by calculating the arithmetic mean of the original parameters. The equations in Figure 7b describe the b parameters.

$$b_{l, fus} = \sum_{k=1}^K b_{l, k} / K$$

$$b_{r, fus} = \sum_{k=1}^K b_{r, k} / K$$

Fig. 7b: Equation for the aggregation of b parameters (also valid for c and d parameters)

All 20 MFs (light grey lines) were aggregated and generated the resulting MF (black line) shown in Figure 8 ($r_{fus} = 23.1$).

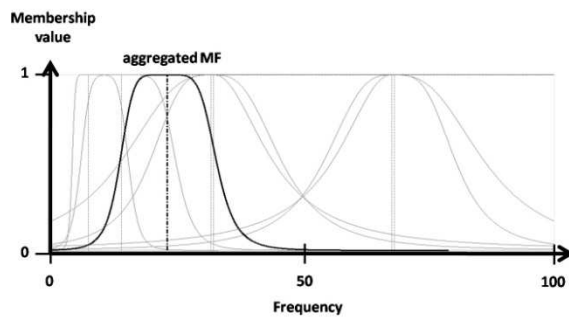


Fig. 8: Aggregated MF for the example presented (see Figure 5)

Following this aggregation procedure, comparison norms of populations can likewise be obtained by merging all “aggregated MFs” of the subjects of the population.

The MF is located in the lower half of the scale between 0 and 50 which indicates low anxiety. The sum score for the example is 31 on a scale ranging from 20 to 80. Transferred to a scale ranging from 0 to 100 the resulting sum score is 18.3. Therefore, the difference between the total score of the fuzzy analysis ($r_{fus} = 23.1$) and the sum score of the statistical analysis is, with 4.8, rather small.

4. Discussion

We used a general two-step translation procedure [6, 7] to numerically translate vague frequency expressions of the Likert scale used on the STAI-T questionnaire. Relative position and shape of the resulting MFs (see Figure 3) provide details about the semantic meaning of the

linguistic terms and can be used as basis for fuzzy analysis of the response data of questionnaires. In an empirical study we manipulated the visual arrangement of the response scale to interrogate if the original presentation format (see Figure 1) influences the interpretation of verbal response categories as having equal distances. However, the scale’s visual arrangement had no significant influence on participants’ estimates of the words meanings (see Table 1). Furthermore, results show that the frequency expressions attributed to the STAI-T response categories are not perceived as being equidistant (see Figure 4). This has implications for the analysis of response data because the STAI-T scale has only ordinal but not interval scale level. Therefore, certain statistical analyses, such as averaging sum score values for the purpose of creating comparison norms for populations, are not valid and should not be used.

We suggest analyzing verbal response data using a fuzzy method as an alternative to traditional statistics. Frequency expressions are vague in nature and in comparison to statistical methods that neglect vagueness the fuzzy approach provides a coherent and adequate framework. In the last part of the paper we showed how the fuzzy MFs of the verbal response labels can be aggregated. The resulting aggregated MF can be interpreted as a participant’s total anxiety score. The r_{fus} is comparable with the sum score obtained by the statistical analysis. Moreover, for the example response data, there is a high similarity between the values ($r_{fus} = 23.1$, sum score $\Sigma = 18.3$). This shows that in general our method leads to similar results (e.g., a low total anxiety value for the example response data).

Interestingly, the MFs provide additional information about the vagueness and coherence of the participant’s answering behavior. The meaning of the resulting aggregated MF is easy to understand because the underlying numerical scale from 0 to 100 can be interpreted as numerical frequency (X out of 100 “cases/events”). This is more comprehensible than the traditional sum score (on a scale between 20 and 80 without a reasonable single unit). The total score represented by $r_{fus} = 23.1$ means that the subject feels anxiety in about one quarter of 100 possible frightening events or situations. The MF parameters that determine the shape (d s) and spread (c s) illustrate the vagueness associated with the estimates. Additionally, the problem with the scale-level requirement for statistical analyses [1, 2] does not exist within the fuzzy framework. This also simplifies the construction of standard norms of populations.

In this paper we illustrated the fuzzy approach for the psychological STAI-T questionnaire. However, the scope of such an approach is broad and may be used for other questionnaires applied in a variety of research settings, as well. Thus, future research should test the suggested fuzzy approach in other contexts and domains. Moreover, the fuzzy analysis of response data has to be further validated to confirm its reliability and usefulness compared to the established alternatives. It would also be interesting to examine how questionnaires with different verbal (and numerical) response scales can be analyzed and interpreted using this fuzzy approach.

References

- [1] S. Jamieson, Likert scales: how to (ab)use them, *Medical Education*, 38: 1217—1218, Blackwell Publishing Ltd, 2004.
- [2] W.M. Kuzon, Jr, M.G. Urbanchek and S. McCabe, The Seven Deadly Sins of Statistical Analysis, *Annals of Plastic Surgery*, 37: 265—272, Little, Brown and Company, 1996.
- [3] B. Rohrmann, Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung, *Zeitschrift für Sozialpsychologie*, 9: 222—245, 1978.
- [4] L. A. Zadeh, Fuzzy sets. *Information and Control*, 8, 338—353, 1965.
- [5] D.V. Budescu, T.M. Karelitz and T.S. Wallsten, Predicting the Directionality of Probability Words from Their Membership Functions, *Journal of Behavioral Decision Making*, 16, 159—180, 2003.
- [6] F. Bocklisch, S.F. Bocklisch and J.F. Krems, How to Translate Words into Numbers? A Fuzzy Approach for the Numerical Translation of Verbal Probabilities. In E. Hüllermeier, R. Kruse and F. Hoffmann (Eds.) *IPMU 2010*, Lecture Notes in Artificial Intelligence 6178, 614—623, Springer, 2010.
- [7] F. Bocklisch, S.F. Bocklisch and J.F. Krems, Sometimes, Often, and Always – Exploring the Vague Meanings of Frequency Expressions, *Behavior Research Methods* (in press).
- [8] C.D. Spielberger, R.L. Gorsuch and R.E. Lushene, Manual for the State-Trait Anxiety Inventory. Palo Alto, California: Consulting Psychologists Press, 1970.
- [9] U. Hoffrage, S. Lindsey, R. Hertwig and G. Gigerenzer, Communicating statistical information, *Science*, 290, 2261—2262, 2000.
- [10] S.F. Bocklisch and N. Bitterlich, Fuzzy Pattern Classification – Methodology and Application. In R. Kruse, J. Gebhardt and R. Palm (Eds.) *Fuzzy Systems in Computer Science*. Vieweg, 1994.
- [11] S.F. Bocklisch, *Handbook Fuzzy Toolbox*, GWT-TUD, Germany, 2008.
- [12] T.S. Wallsten, D.V. Budescu, A. Rapoport, R. Zwick and B. Forsyth, Measuring the Vague Meaning of Probability Terms. *Journal of Experimental Psychology: General*, 115, 348—365, 1986.