ISSN 0254-9271

Semiotische Berichte

mit:
Linguistik Interdisziplinär

Heftherausgeber:
Jeff Bernard, Imre Gráfik, Vilmos Voigt, Gloria Withalm

Zeichen / Kultur

Akten des
3. Österreichisch-Ungarischen
Semiotik-Kolloquiums
Szombathely/Velem 1992

Jg. 17
3,4/1993

Rüdiger GROTJAHN, Anna TÓTHNÉ LITOVKINA, Peter GRZYBEK, Christoph CHLOSTA, Undine ROOS

Statistical Methods in the Study of Proverb Knowledge

An Analysis of the Knowledge of Proverbs in Contemporary Hungarian Culture (Tolna County)

0. Introduction

Contemporary proverb scholarship faces an increasing interest in what has recently been termed 'empirical paremiology' (cf. Grzybek/Chlosta 1993). 'Empirical paremiology' is a particular approach in the general field of proverb studies, which ultimately aims at empirically-based answers to two major questions:

(a) how many proverbs and which proverbs are (still) commonly familiar in a given culture?
(b) what does proverb knowledge in this culture (and in general) depend upon?

As has been shown elsewhere (cf. Grzybek 1991b), an adequate analysis of proverbs may serve as a prototype for cultural semiotic studies in general. It goes without saying that the two questions mentioned above can, logically speaking, be answered with regard to a single culture, alone. The relevance of these findings, however, can only be estimated with regard to a comparative basis from other cultures as well. In this sense, 'empirical paremiology' turns out to be a more or less strictly defined concept; it definitely goes beyond those empirical approaches in general, which have, from time to time, found their way to proverb studies over the last decades.

'Empirical paremiology' is mainly based on G.L. Permjakov's (1919-1983) pioneering studies on Russian proverbs and on what he called a "paremiological minimum" of the Russian language. In the mid 70s, Permjakov conducted two paremiological experiments in Moscow

county. A major purpose of these studies was to establish the so-called Russian paremiological minimum, i.e., the most familiar Russian proverbs, proverbial expressions, proverbial comparisons, riddles, slogans, weather signs, etc.

Permjakov published several articles and books about this topic,[1] but his death prevented him from finishing his work. Specifically, Permjakov could not present any detailed explanation of the methods of his empirical investigation, analyze the results, or calculate statistics.[2]

Subsequent to the presentation of Permjakov's approach to the Western paremiological audience (cf. Grzybek 1984b), experts in paremiology and phraseology have shown an increasing interest in this topic (cf., e.g., Mieder 1986, 1990; Schellbach-Kopra 1987; Ruef 1989). Summarizing this discussion, Mieder (1990: 141f.) claims:

> Similar paremiological minima of the most frequently used phraseological units of all the national languages should now be established by paremiographers [...]. In any case, it would benefit foreign language instruction tremendously if paremiographers would establish paremiological minima for many individual languages.

Meanwhile, a number of similarly oriented studies have been initiated. Most progress in this direction is being provided by studies focusing on German and Croatian. Starting from a cultural semiotic perspective (Grzybek 1984a, 1991a), a group of German researchers[3] attempts to systematically find out the proverbs generally known in each of these two cultures, to subsequently analyze the two corpora obtained, and to study the factors influencing knowledge, from both an intracultural and an intercultural perspective (cf. Grzybek 1991a, Grzybek et al. 1994). Meanwhile, the results of pilot studies for both cultures have been published (Baur/Grzybek 1990, Grzybek 1991a, Grzybek et al. 1993). More importantly, these studies have served as a methodological orientation for scholars conducting comparable research in other cultures: thus, methodologically based on the German studies, comprehensive experiments have been undertaken by Schindler (1993) with regard to Czech proverbs, and by Tóthné Litovkina (1992, 1993) with regard to Hungarian proverbs.

As can be seen from these studies, empirical paremiology will always be, in one way or another, confronted with statistical problems. Questions as to (a) – how many and which proverbs are known? – may still be answered with the help of basic statistical tools (arithmetic means, percentages, etc.); questions as to (b), however, focusing on the study of which factors influence proverb knowledge – be it on an intracultural or on an intercultural level – necessarily involve a more complex statistical approach.

The objective of the present article is to discuss the relevance, i.e., the efficiency and the limits of various statistical methods which either have been or which might be employed in empirical paremiology. In this sense, the present article is not so much (or not only) meant to be a statistical analysis of a particular Hungarian paremiological study; rather, it is intended to be a methodological discussion relevant for any future empirically-based paremiological research.

It is important to keep in mind, therefore, that no representative conclusions as to Hungarian language and culture in general will be possible on the background of this article. Still, in addition to the general methodological discussion, some major tendencies will be outlined on the basis of the results obtained.

1. Material and Background of the Study

The material of the present study is based on a paper presented by Tóthné Litovkina at the Third International Congress of Hungarology (Szeged, Hungary, Aug. 12-16, 1991) in the Hungarian language, and its translation into English (Tóthné Litovkina 1992), presented at the 3rd Austriaco-Hungarian Semiotics Colloquium at Szombathely/Velem, Hungary, April 21-23, 1992.

On the basis of sociolinguistic research conducted all over Hungary in 1991/92, the knowledge of proverbs among Hungarian residents is analyzed. In the present paper, only the data of informants from the area of Tolna county are taken into account ($N=71$).[4] The informants were asked to complete a list of the beginnings of 378 Hungarian proverbs[5]; thus, the first part of each proverb was eliminated, and the informants' task was to fill in the missing text. As far as we know, this is the first study of this kind with regard to the Hungarian language.

While the above-mentioned papers by Tóthné Litovkina (1992, 1993) provided only preliminary results, this article contains more solid statistical analyses which allow a re-evaluation of the data presented before. The statistical methods used include basic univariate and bivariate

analysis, ANOVA and ANCOVA, and regression analysis on the basis of the General Linear Model (GLM). It will become clear that the less complex statistical analyses usually employed in the field of paremiological research may lead to erroneous conclusions. Since the purpose of this paper is, first and foremost, a methodological one, the statistical methods used will be discussed at some length.

2. The Design of the Study

2.1. Subjects

For the present study we used the data of 71 Hungarian residents of Tolna county (Southern Hungary).[6] In selecting the subjects, an attempt was made to include an approximately equal number of both male and female, as well as younger and older, respondents. – As already pointed out above, the sample is relatively small and cannot, in a statistical sense, be regarded as representative of Hungary as a whole.

2.2. Task

Each participant in the study received a list consisting of 378 Hungarian proverbs, or rather the beginnings of them, e.g.: *(a) Addig jár a korsó a kútra...* [The pitcher goes so long to the well... – cf. Engl.: The pitcher which goes often to the well will be broken at last]; *Addig üsd a vasat, ...* [Strike the iron while... – cf. Engl.: Strike while the iron is hot]. Additionally, the questionnaire contained questions concerning informants' sex, age, education, place of residence, place of residence until age 18, etc. The respondents were asked to anonymously answer these questions and to complete the presented beginnings of the proverbs. They were instructed not to ask for anyone else's help, and not to use any dictionaries of proverbs or similar reference sources.

2.3. Selection of the Material

One of the most difficult tasks in an attempt to establish a paremiological minimum is to provide the proper items (i.e., proverbs) for the questionnaire. Mostly recently, Grzybek/Chlosta (1993: 112ff.) have claimed that not only the empirical study itself, but also the achievement of the experimental corpus has to be based on empirical research. As will be seen below, this is not entirely the case with the Hungarian experiment reported on in this article; instead, quite a number of subjective factors come into play. However, this point is not crucial with regard to the conclusions of the present article.

The first step of the study was intended to reduce the vast and mainly obsolete material contained in traditional Hungarian proverb collections. According to the popular Hungarian proverb *Három a magyar igazság* [*The Hungarian justice* (i.e., custom) *is three*; cf. Engl.: *All good things come in threes*], the above-mentioned list of the 378 proverbs was developed in three steps:

1. At the first stage, a list of about 4,000 proverbs was derived. This list included:

(a) all proverbs for which illustrative quotations from Hungarian fiction (17th to 20th centuries) and contemporary journalistic literature could be found. Thus, in more than 10 years, about 4,000 examples of proverbial usage were found, among them about 1,000 different proverbs;

(b) about 100 proverbs heard from radio or television transmissions, or in talks with Hungarian native speakers over a period of five years;[7]

(c) all proverbs were taken from Erdélyi's (1851), Bálint's (1972), and Paczolay's (1987; 1989) collections of Hungarian proverbs and proverbial sayings; from O. Nagy's (1976) collection, those proverbs were taken which are marked with the sign "still popular"[8]; from Margalits (1897), those proverbs were taken selectively, which appeared to be still familiar; and from Vöő (1989), the 200 most popular proverbs were taken.[9]

As to these 4,000 proverbs, 30-50 respondents were asked to mark the proverbs familiar to them with the sign "+", and those proverbs not familiar to them with the sign "–".[10] As a result, there remained 930 proverbs familiar to at least 50% of the informants.

2. At the second stage of reducing the overall material, the 930 proverbs were used which were familiar to at least 50% of the participants from stage 1. A questionnaire consisting of the *beginnings* of these 930 proverbs and of 12 additional questions was given to 22 respondents. The respondents were asked to complete the presented proverb texts and to answer the additional questions. As a result, 378 proverbs were obtained which were, more or less, filled in in the expected form by 50% or more of the 22 informants.[11]

3. At the third stage, the 378 proverbs obtained at stage 2 were taken as the items for the final version of the list of proverbs, which was used for the present study. Of the 650 questionnaires distributed all over Hungary, 430 questionnaires were returned; 12 of them had to be eliminated from analysis, either because the informants were not native speakers of Hungarian, or due to missing answers being in the majority.

The completion of a proverb involved in phases 2 and 3 requires a real (active) knowledge of it, as opposed to the first stage, where the participants might have marked a proverb as "familiar" just because it seemed to be logical, or true. This assumption is supported by the fact that after the second stage, only 378 proverbs out of 930 were completed by more than 50% of the participants (for a methodological discussion of "measuring" acquaintance of proverbs, cf. Chlosta/Grzybek 1994).

2.4. Coding of the Data and Statistical Analyses

A simple coding system was employed for analysis, based on the categories mentioned above (cf. note 11). Statistical analyses were carried out using SPSS/PC+ 4.01.

3. Results

3.1. Demographic Variables

Since the number of respondents was relatively small, an exhaustive statistical analysis of the entire demographic information asked for in the questionnaire was not possible. Some information had to be aggregated, other information had to be excluded. The following variables were finally used in the analysis:

AGE (numerical values ranging from 18 to 74),
SEX (male=0, female=1),
EDUCATION,
RESID-YOUTH (= residence till the age of 18),
RESID-PRES (= present place of residence),
COUNTY-YOUTH (= county of residence till the age of 18).

The variables RESID-YOUTH and RESID-PRES had the levels 'town/city' (coded as 0) and 'village' (coded as 1).[12] COUNTY-YOUTH comprised the levels 'Tolna' (coded as 0) and 'Non-Tolna' (coded as 1). The variable EDUCATION was coded as follows:

primary: subjects who had finished primary school (code=1);
secondary: subjects who had finished secondary school (code=2);
university: subjects who had graduated from university or college (code=3).

In Grzybek's (1991a) pilot study on German proverbs, the age of the respondents turned out to be the only variable with a significant impact on knowledge of German proverbs. For the present Hungarian sample, *Table 1* (see page 280) shows the means, standard deviations and sample sizes for AGE cross-classified with EDUCATION on the one hand, and SEX, RESID-YOUTH, RESID-PRES and COUNTY-YOUTH on the other hand.

From *Table 1* it can be seen that 29 (40.8%) of the 71 subjects were male and 42 (59.2%) female. The average age was 44.4 with a standard deviation of 17.3 and a median of 45. The youngest informant was 18, the oldest 74. All subjects were living in Tolna county at the time of the study; 46 (64.8%) of them lived in town, the remaining 25 (35.2%) in a village. 44 subjects (62.0%) spent both their childhood and youth in Tolna county, 27 (38.0%) in other Hungarian counties. 40 (56.3%) spent both childhood and youth in a village, the remaining 31 (43.77%) in a town or in a city. 16 subjects (22.5%) finished primary school, and 39 (54.9%) secondary school; 16 (22.5%) graduated from university or college.

The variables in *Table 1* appear to be interrelated in various ways. To examine the interrelations more closely, correlation coefficients were calculated taking the level of measurement of the variables into account. To provide for comparability, only the Pearson product-moment correlations will be presented; they correspond to the Phi-coefficient in the case of two binary variables, and to the point-biserial coefficient in the case of a binary and a quantitative variable. The correlations are presented in *Table 2*.

Table 1
Cross-Classification of EDUCATION with SEX, RESID-YOUTH, RESID-PRES and COUNTY-YOUTH for AGE as Dependent Variable: Means, Standard Deviations and Sample Sizes

Variables		EDUCATION			Total
		Primary	Secondary	University	
SEX		56.8	42.8	43.4	45.5
	Male	13.1	16.5	9.4	13.9
		5	12	12	29
		58.7	38.3	40.2	43.7
	Female	16.6	19.1	13.4	19.5
		11	27	4	42
RESID-YOUTH		46.2	39.4	44.5	41.6
	Town/City	16.7	18.3	8.7	16
		4	19	8	31
		62.1	39.6	40.7	46.6
	Village	10.6	18.6	11.7	18.2
		12	20	8	40
RESID-PRES		52.0	41.0	38.8	42.8
	Town/City	12.7	16.9	9.3	14.6
		4	27	15	46
		60.2	36.2	25	47.3
	Village	14.0	21.4	0	21.5
		12	12	1	25
COUNTY-YOUTH		58.7	35.6	42.5	41.2
	Tolna	15.9	16.9	11.0	18.3
		9	29	6	44
		57.4	50.9	42.7	49.6
	Non-Tolna	11.6	17.8	10.2	14.6
		7	10	10	27
		58.1	39.5	42.6	44.4
	Total	13.7	18.2	10.1	17.3
		16	39	16	71

Table 2
Pearson Correlations*

Variables	AGE	EDUC	SEX	RESID-Y	RESID-P	COUNTY-Y
AGE	–	-.302	-.052	.145	123	.234
	–	(.005)	(.334)	(.114)	(.153)	(.025)
EDUC	-.302	–	-.299	-.169	-.483	.130
	(.005)	–	(.006)	(.079)	(.000)	(.141)
SEX	-.052	-.299	–	.019	.253	-.293
	(.334)	(.006)	–	(.436)	(.017)	(.007)
RESID-Y	.145	-.169	.019	–	.471	-.188
	(.114)	(.079)	(.436)	–	(.000)	(.058)
RESID-P	.123	-.483	.253	.471	–	-.213
	(.153)	(.000)	(.017)	(.000)	–	(.037)
COUNTY-Y	.234	.130	-.293	-.188	-.213	–
	(.025)	(.141)	(.007)	(.058)	(.037)	–

* Numbers in brackets are one-tailed probabilities; $N = 71$ in all cases.

In addition to simple Pearson correlations, multiple regression was used to measure the strength of the relationship between EDUCATION and the remaining variables. To this end, EDUCATION was split into two binary variables (dummy coding), and the multiple correlation between the two binary variables as predictors and the remaining variables was calculated. The multiple correlations were always higher than the corresponding Pearson correlations. The same holds for the adjusted coefficient of determination provided by SPSS; for an interpretation of the coefficient of determination cf. Bliesener (1992) and Grotjahn (1992). Particularly striking was the increase in correlation with the variable COUNTY-YOUTH. The value obtained is $R = .310$, which corresponds to a probability of $p = .032$. The amount of common variance is 7% (after adjustment).

A number of interesting patterns emerge from Tables 1 and 2. For example:

1. The level of education of males tends to be higher than that of fe-

males; specifically, 41% of the males, but only 9.5% of the females have graduated from university or college.
2. The level of education of respondents presently living in a town or city appears to be higher compared to those living in a village.
3. AGE seems to be related to EDUCATION. Specifically, primary education tends to be associated with the older age group.
4. Respondents who lived in a village until the age of 18 tend to be older than those who spent their childhood and youth in a town or in a city.
5. Respondents who lived outside of Tolna county until the age of 18 tend to be older than those who spent this period of life in Tolna county; in addition, these respondents tend to be male.
6. Respondents presently living in a town or city appear to have lived in a town or city during their childhood and youth.
7. Respondents who lived outside of Tolna county until the age of 18 appear to have a higher level of education than those who spent this period of life in Tolna county. (This interpretation is based on results of the multiple regression analysis).

In particular, the first three patterns are in line with sociological theory and research. This is very comforting, since the sampling of respondents has not been strictly controlled in the present study, and results may therefore be biased. However, the simple bivariate correlation between two variables may also be influenced by their correlation with the remaining variables (as well as by their correlation with variables not included in this study). This might be another source of bias; the above interpretations should therefore be viewed with caution.

In Section 3.3.6. a multiple regression approach will be used to tackle this problem. It will become clear that calculating simple bivariate correlations or simple *t*-tests of means is an inadequate approach to study the complex influence of AGE, EDUCATION, SEX, RESIDENCE, and other demographic variables on proverb familiarity.

3.2. General Knowledge of Proverbs

On average, of the 378 proverbs presented, 311.92 (82.52%) were filled in "properly", and 6.92 (1.83%) were filled in "incorrectly";[13] 57.86 (15.30%) were left totally unanswered; 1.30 (0.34%) were completed as a saying, but not as a proverb. The general knowledge of proverbs thus seems to be relatively high. On the one hand, this high level of proverb

familiarity among the respondents can be interpreted as an indication of reliability in the selection of the material (cf. Section 2.2., above); on the other hand, this a-priori selection of the test material has to be adequately taken into account when analyzing factors influencing proverb knowledge.

Only 42 proverbs, that is only 11.11% of the entire test material, were completed "correctly" by each participant: 162 proverbs (42.86%) were known by 90% or more of the respondents. The 42 proverbs known by all informants, together with an exact translation into English and, where necessary, with a short explanation, are listed in the Appendix.

3.3. What Does Knowledge of Proverbs Depend On?

3.3.1. Variables

Let us now look at the variables possibly influencing knowledge of proverbs. In the following, these variables will also be referred to as predictor variables, predictors, independent variables or factors. The predictor variables considered are: AGE, EDUCATION, SEX, RESID-YOUTH, RESID-PRES, and COUNTY-YOUTH (see Section 3.1. for more information on these variables). As opposed to the results reported by Tóthné Litovkina (1992, 1993), only subjects with a minimum age of 18 were included into the present analysis. This minimizes possible developmental influences, which will be discussed elsewhere; at the same time, this makes results comparable to those obtained in the German pilot study (Grzybek 1991a) and in the Czech experiment (Schindler 1993).

To measure knowledge of proverbs, the total number of proverbs filled in "correctly" has been counted and, for ease of interpretation, standardized by dividing it through 378, i.e., the total number of proverbs presented to the subjects. This number has then been multiplied by 100, which resulted in a possible range of [0; 100]. The corresponding variable has been designated as PROV-PROP: the proportion (%) of proverbs filled in "correctly"; in the following analyses, it constitutes the dependent variable.

3.3.2. Some Descriptive Statistics and some General Tendencies

To convey an impression of the intricate interrelations among the variables examined, we have calculated the means, standard deviations and sample sizes of PROV-PROP for various cross-classifications of the independent variables, splitting AGE at its median. The data are presented in *Tables 3* and *4*.

Table 3 (see next page)
Cross-Classification of AGE with EDUCATION, SEX, RESID-YOUTH, RESID-PRES and COUNTY-YOUTH for PROV-PROP as Dependent Variable: Means, Standard Deviations and Sample Sizes

Variables		AGE		Total
		≤45	>45	
EDUCATION	Primary	73.5	80.2	79.4
		7.9	10.2	10.0
		2	14	16
	Secondary	75.9	89.0	80.9
		10.0	6.7	10.9
		24	15	39
	University	86.3	87.5	86.7
		7.5	12.7	9.4
		10	6	16
SEX	Male	82.4	86.7	84.5
		7.6	10.4	9.2
		15	14	29
	Female	75.9	84.3	80.1
		11.2	9.8	11.2
		21	21	42
RESID-YOUTH	Town/City	80.1	87.6	83.0
		10.7	7.7	10.2
		19	12	31
	Village	77.0	84.0	81.0
		9.8	10.7	10.9
		17	23	40
RESID-PRES	Town/City	80.9	88.1	84.0
		9.3	8.7	9.6
		26	20	46
	Village	72.7	81.5	78.0
		10.7	10.7	11.4
		10	15	25
COUNTY-YOUTH	Tolna	76.9	82.5	79.4
		10.3	10.6	10.7
		24	20	44
	Non-Tolna	82.1	89.0	85.9
		9.7	8.0	9.3
		12	15	27
Total		78.6	85.2	81.9
		10.3	10.0	10.6
		36	35	71

Table 4
Cross-Classification of EDUCATION with AGE, SEX, RESID-YOUTH, RESID-PRES and COUNTY-YOUTH for PROV-PROP as Dependent Variable: Means, Standard Deviations and Sample Sizes

Variables		EDUCATION			Total
		Primary	Secondary	University	
AGE	≤45	73.5	75.9	86.3	78.6
		7.9	10.0	7.5	10.3
		2	24	10	36
	>45	80.2	89.0	87.5	85.2
		10.2	6.7	12.7	10.0
		14	15	6	35
SEX	Male	84.7	82.8	86.0	84.5
		11.4	8.1	9.8	9.2
		5	12	12	29
	Female	77.0	80.1	88.9	80.1
		8.8	12.0	8.9	11.2
		11	27	4	42
RESID-YOUTH	Town/City	80.4	80.2	91.0	83.0
		12.4	10.0	5.2	10.2
		4	19	8	31
	Village	79.1	81.6	82.5	81.0
		9.7	12.0	10.9	10.9
		12	20	8	40
RESID-PRES	Town/City	87.4	81.4	87.9	84.0
		8.2	9.8	8.5	9.6
		4	27	15	46
	Village	76.7	80.0	69.6	78.0
		9.3	13.5	0	11.4
		12	12	1	25
COUNTY-YOUTH	Tolna	74.9	80.0	83.3	79.4
		7.6	10.9	13.0	10.7
		9	29	6	44
	Non-Tolna	85.3	83.5	88.8	85.9
		10.1	11.2	6.3	9.3
		7	10	10	27
Total		79.4	80.9	86.7	81.9
		10.0	10.9	9.4	10.6
		16	39	16	71

If we consider only the bivariate relationships between the variables, the independent variables relate to proverb knowledge as follows:

(a) AGE

Knowledge of proverbs increases with age. This appears to be a rather strong tendency and seems to hold for all levels of the remaining variables.

(b) EDUCATION

The higher the level of education, the higher the knowledge of proverbs. This seems to hold for all levels of the remaining variables with the exception of RESID-PRES. (The result for RESID-PRES is probably due to a sampling error.)

(c) SEX

Males know more proverbs than females.

(d) RESID-YOUTH

Respondents who spent their youth in a town or city know slightly more proverbs than those who spent their youth in a village. However, as will be shown below, this difference is so small that it has to be ascribed to chance.

(e) RESID-PRES

Respondents presently living in a town or a city know considerably more proverbs than those from a village.

(f) COUNTY-YOUTH

Respondents who spent their youth outside of Tolna county know considerably more proverbs than those who spent their youth in Tolna county.

3.3.3. Pearson Correlations and Analysis of Variance and Covariance

In order to be able to compare the results of our analyses to those obtained by Grzybek (1991a) and Schindler (1993), the significance of the difference among the means of the five independent variables (see *Tables 3* and *4*) will now be tested by way of an analysis of variance (ANOVA). As in the studies mentioned, multiple *F*-tests will be used

without a corresponding adjustment of the significance level. As will be argued below, this approach is not very satisfactory for various reasons.

The results of the *F*-tests are shown in *Table 5*; note that the probabilities (*p*) are two-tailed.

In addition, Pearson correlations between PROV-PROP, on the one hand, and EDUCATION, SEX, RESID-YOUTH, RESID-PRES and COUNTY-YOUTH, on the other hand, were computed. The results are shown in *Table 6*.

Table 5
ANOVA: *F*-Tests of Main Effects

Variables	F	d.f.	p
AGE	7.54	1; 69	.008
EDUCATION	2.35	2; 68	.103
SEX	2.99	1; 69	.088
RESID-YOUTH	0.61	1; 69	.438
RESID-PRES	5.64	1; 69	.020
COUNTY-YOUTH	6.81	1; 69	.011

Table 6
Pearson Correlations*

Variables	AGE	EDUC	SEX	RESID-Y	RESID-P	COUNTY-Y
PROV-PROP	.392	.233	-.204	-.093	-.275	.300
	(.000)	(.025)	(.044)	(.219)	(.010)	(.006)

* Numbers in brackets are one-tailed probabilities; $N = 71$ in all cases.

According to the calculations presented in *Tables 5* and *6*, all variables except for RESID-YOUTH seem to have some impact on the knowledge of proverbs. AGE appears to be most influential, a finding which is in line with the results reported in Grzybek (1991a), Schindler (1993), and Tóthné Litovkina (1992). AGE is followed by COUNTY-YOUTH and RESID-PRES, whereas the influence of EDUCATION and SEX appears to be rather weak and might be due to chance.[14]

However, if the independent variables are interrelated, the analysis of simple main effects and of the bivariate relationship of each independent variable with the dependent variable may be quite misleading (cf. also the discussion in Section 3.3.4. above).

To control for the effect of the quantitative variable AGE on EDUCATION, Grzybek (1991a) used analysis of covariance (ANCOVA). To obtain comparable results, we shall now use ANCOVA as well, extending its application to *all* independent variables. To carry out the calculations, the procedure ANOVA provided by SPSS was used, as in Grzybek (1991a). Since the data used in the present study have not been obtained on the basis of a true experiment, the classical experimental approach to ANCOVA was not used, but a regression model (provided by SPSS as an option in the ANOVA procedure). As a consequence, all effects are assessed simultaneously, with each effect adjusted for all other effects in the model (cf. Norušis 1990b, chap. B4).

Table 7 shows the results of the analysis of covariance with age as covariate. The effect of the covariate AGE is highly significant in all five analyses. With regard to the other independent variables, however, the results of ANCOVA differ considerably from those of simple ANOVA (cf. *Tables 7* and *5*). Most interestingly, the effect of EDUCATION is now highly significant. In addition, the effect of RESID-PRES appears to be even stronger than in *Table 5*.

Table 7
Analysis of Covariance with Age as Covariate

Variables	F	d.f.	p
AGE	21.78	1; 67	.000
EDUCATION	6.66	2; 67	.002
AGE	12.24	1; 68	.001
SEX	2.82	1; 68	.097
AGE	13.90	1; 68	.000
RESID-YOUTH	1.90	1; 68	.172
AGE	16.95	1; 68	.000
RESID-PRES	9.75	1; 68	.003
AGE	9.33	1; 68	.003
COUNTY-YOUTH	3.88	1; 68	.053

In Table 3, some indication can be found as to why there is a highly significant effect of EDUCATION after partialling out the effect of AGE: the impact of EDUCATION appears to be more pronounced when we split the sample at the median of AGE (Med = 45).

Our results for EDUCATION stand in sharp contrast to Grzybek's (1991a) analyses. Grzybek found that the simple main effect of EDUCATION on the knowledge of German proverbs was significant. This effect disappeared, however, when AGE was included as a covariate in the analysis of variance.

Unfortunately, also the results of the ANCOVA may still be biased. In ANCOVA only the influence of AGE has been controlled for when analyzing, for example, the effect of EDUCATION or RESID-PRES. However, in Section 3.1. it has been shown that there is a set of complex interrelations among the variables studied. Without taking all the complex interrelations into account, our conclusions may be flawed or even completely wrong.

For the analysis of the joint effect of a set of interrelated predictor variables, regression analysis on the basis of the General Linear Model (GLM) is a more adequate and much more powerful instrument than ANOVA or ANCOVA. The GLM can be used for any combination of qualitative and quantitative predictors, including ANOVA/ANCOVA as a special case. Furthermore, it provides us with additional tools for the measurement of the (joint) effect of several independent variables.[15]

3.3.4. Regression Analysis on the Basis of the General Linear Model

The statistical model of linear multiple regression is based on a number of assumptions. Violation of assumptions may partially or even completely invalidate the model and, as a consequence, the results of analysis as well; see Tabachnick/Fidell (1989, chap. 5), Norušis (1990b, chap. B6), and Grotjahn (1992) for brief overviews. For example, the following problems can arise in regression analysis: nonnormality (of residuals), heteroskedasticity, dependence of error, multicollinearity, interaction, outliers, measurement error. An attempt was made to check for possible violations of assumptions, outliers, and other conceivable problems. To this end, the procedure "EXAMINE" and the subcommand "Residuals" within the procedure "REGRESSION" provided by SPSS were used; see Norušis (1990a, chap. B9; 1990b, chap. B4; 1991, chaps. 13, 25, 26).

In this article, we can neither describe the various procedures used nor present the results of these analyses in detail; a brief discussion will be found in Grotjahn (1994). The analyses showed that there were no serious violations of assumptions. There were some outliers, but they were not so extreme that we had to consider the possibility of excluding them from further analysis. Multicollinearity proved to be sufficiently low for all models examined.[16] Note that the relationship between AGE and PROV-PROP turned out to be linear and that in particular the distribution of PROV-PROP slightly departs from normality.

Table 8 presents the mean, median, standard deviation (SD), mode, skewness, kurtosis, minimum, and maximum for AGE and PROV-PROP.

Table 8
Descriptive Statistics for AGE and PROV-PROP

Variables	Mean	Median	SD	Mode	Skew	Kurt	Min	Max
AGE	44.41	45	17.33	49	-.06	-1.19	18	74
PROV-PROP	81.90	83.33	10.61	88.09	-.45	-.59	54.50	97.88

The distribution of AGE is symmetric, but somewhat flatter than a normal curve (this is indicated by the negative value for the kurtosis). However, since the standard error for the kurtosis is .56 and thus quite high, this departure from normality is probably due to chance. However, as both the plot of the raw frequencies and the normal plot provided by SPSS show, there are pronounced irregularities in the left tail of the distribution: 14 informants fall into the age interval 18 to 22, but only 4 informants into the interval 23 to 31 – a fact probably due to poor sampling (cf. Section 2.3.). These and other irregularities are reflected by the results of the Lilliefors test – a modified Kolmogorov-Smirnov test for normality provided by SPSS (cf. Norušis 1990a, B104). The test yields a probability of .059; the fit to a normal distribution is thus not very good.[17] This should be kept in mind when interpreting results which involve the variable AGE.

The distribution of PROV-PROP is somewhat skewed to the left. This means that knowledge of proverbs is higher than expected for a normal distribution. However, since the standard error for the skewness is .28 and thus quite high, and since furthermore the Lilliefors test for nor-

mality yields a probability greater than .20, this and other irregularities are probably due to chance.

Yet, the observed skewness might also be explained by the fact that the 378 proverbs analyzed in the present study have been selected on the basis of their familiarity to two pilot samples of informants (cf. Section 2.2.). Consequently, the task might have turned out to be relatively easy[18] for the present sample. Due to this ceiling effect, the present set of proverbs is not optimally suited for discriminating among informants with a relatively high knowledge of proverbs.

We have also checked the assumption of normality of AGE and PROV-PROP for the different levels of the remaining variables. On the whole, the fit to normality was quite satisfactory.

Subsequent to the examination of assumptions, the predictors have been checked for possible interactions by help of the procedures "ANOVA" and "MANOVA" provided by SPSS. Due to empty cells, only two-way interactions could be calculated. These analyses yielded no significant results ($.21 < p < .99$); furthermore, the overall F-value for interaction among the factors was not significant ($p = .36$). A similar result was obtained for the interaction between the factors, on the one hand, and the covariate AGE, on the other hand (assumption of homogeneous slopes).[19] Therefore, no interaction terms were included in the regression model.

Next, various regression analyses were carried out to select the variables for the model. Since selection of variables may depend on the selection method used, different methods were applied and the final decision taken on the basis of the overall results. The selection methods used were (with default options in SPSS): a) forward selection; b) stepwise entry and removal; and c) backward elimination. In these analyses EDUCATION was treated as a three-level numerical variable. The methods a) and b) yielded the same results: first AGE and then EDUCATION were selected. Backward elimination pointed to a weak additional effect of RESID-PRES.

Table 9 presents for two different models the following basic regression statistics: the multiple correlation coefficient (R); the adjusted multiple determination coefficient (R^2_a); the F-test for change F_{ch} and its probability (p_{ch}); the semipartial correlation referred to as part correlation (*Part-C*) in SPSS; the t-test for the regression coefficient b and its two-tailed probability (p_t); the F-Test (ANOVA) for overall regression (F_{ano}) and its probability (p_{ano}). Note that the values for *Part-C*, t, and p_t listed refer to the models with three and two independent variables respectively. Further, note that transforming EDUCATION into two dummy variables leads to almost identical results.[20]

Table 9
Results of Multiple Regression Analyses

Variables	R	R^2_a	F_{ch}	p_{ch}	Part-C	t	p_t	F_{ano}	p_{ano}
AGE	.39	.14	12.56	.0007	.49	4.77	.0000	12.56	.0007
EDUCATION	.54	.27	13.07	.0006	.37	2.43	.0179	13.91	.0000
RESID-PRES	.57	.29	2.91	.0925	-.17	-1.71	.0925	10.50	.0000
AGE	.39	.14	12.56	.0007	.43	4.12	.0001	12.56	.0007
RESID-PRES	.51	.24	9.75	.0026	-.33	-3.12	.0026	11.95	.0000

From *Table 9* it can be seen that AGE is the best predictor. As the value for R^2_a shows, AGE alone accounts for 14% of the variance. When we add EDUCATION, this amount rises to 27%. The inclusion of RESID-PRES adds only another two percent – the small amount being probably due to the high intercorrelation between EDUCATION and RESID-PRES ($r=.48$; see *Table 2*). Inclusion of the remaining variables SEX, RESID-YOUTH and COUNTY-YOUTH does not lead to any further increase in the amount of variance accounted for; the adjusted multiple determination coefficient even drops to .27.

We have also checked the effect of an exclusion of EDUCATION from the model. The F-value for the model containing only AGE and RESID-PRES is highly significant as well. This model accounts for 24% of the variance, that is, slightly less than a model consisting of AGE and EDUCATION (cf. bottom of *Table 9*).

As the data in *Table 3* show, there is almost no difference between the levels 1 and 2 of EDUCATION. The decisive fact is whether an informant graduated from a university/college or finished only primary/secondary school. We can thus expect the highest level of proverb knowledge with older informants who graduated from university or college and live in a city/town. It must be stressed again that in the light of the sampling procedure used and the small number of subjects involved any generalization to the whole of Tolna county is highly problematic.

Since the effect of RESID-PRES is very small, we eventually decided to use the more parsimonious model containing only AGE and EDUCATION as predictors.[21] This model yields the following values for *Part-C*, *t*, and p_t : .50, 4.75, and .0000 in the case of AGE, and .40, 3.61, and .0006 in the case of EDUCATION. Exclusion of RESID-PRES thus leads to an increase of the importance of the predictor EDUCATION – a result expected in view of the correlation between EDUCATION and RESID-PRES.

If we use the subscript *est* for *estimated*, the regression equation for the model containing only AGE and EDUCATION as predictors reads:

$$\text{PROV-PROP}_{est} = .31\ \text{AGE} + 6.08\ \text{EDUCATION} + 55.88. \quad (1)$$

In Table 10, the values for the variables used in equation (1) and the (standardized) residuals calculated on the basis of equation (1) are listed for our first 10 subjects. For subject #1, for example, insertion of 40 for AGE and 3 for EDUCATION in equation (1) yields an estimated or predicted value of 86.60 for PROV-PROP. If we subtract this value from the actual value for PROV-PROP, i.e., 93.65, we get a residual of 7.05 (and after additonal calculations a standardized residual of .79). It can be seen that except for subject #8 the model predicts the knowledge of proverbs quite well.[22]

Table 10
First 10 Subjects: Values of Variables and (Standardized) Residuals

Variables	Subject #									
	1	2	3	4	5	6	8	9	10	11
AGE	40	32	26	41	62	49	41	40	49	37
EDUCATION	3	3	2	3	2	2	2	2	2	1
PROV-PROP	93.6	89.4	73.3	93.9	92.6	95.0	59.8	75.7	76.5	79.1
Residuals	7.1	5.3	-2.9	7.0	5.2	11.6	-21.1	-4.9	-6.9	5.6
Stand. Res.	.79	.60	-.32	.79	.58	1.29	-2.34	-.54	-.76	.63

We now searched for an explanation as to why predictions are quite inaccurate for some subjects. To this end, we selected the 10 subjects with the largest standardized residuals and examined their values in *all* variables under investigation. These data are presented in *Table 11*.

Table 11
Subjects with the Largest Standardized Residuals

Variables	Subject #									
	8	14	32	44	57	59	60	68	70	77
AGE	41	20	73	46	65	19	64	50	25	18
EDUCATION	2	2	1	3	1	2	1	1	3	2
SEX	1	1	1	0	1	1	0	0	0	1
RESID-Y	0	1	1	1	1	1	1	0	1	0
RESID-P	0	1	1	0	1	1	1	0	1	0
COUNTY-Y	1	0	0	0	0	0	1	1	0	0
PROV-PROP	59.8	89.9	68.8	65.08	61.8	54.5	97.3	96.3	69.6	87.6
Residuals	-21.1	15.7	-16.0	-23.4	-20.9	-19.5	15.4	18.7	-12.3	13.9
Stand. Res.	-2.34	1.77	-1.82	-2.65	-2.36	-2.20	1.74	2.11	-1.40	1.57

Codes:
EDUCATION: 1 = primary; 2 = secondary; 3 = university/college
SEX: 0 = male; 1 = female
RESID-YOUTH: 0 = town/city; 1 = village
RESID-PRES: 0 = town/city; 1 = village
COUNTY-YOUTH: 0 = Tolna; 1 = non-Tolna

There appears to be no conspicuous pattern in *Table 11*. Perhaps it is worthwhile noting that the proportion of subjects presently living in a village is comparatively high: 60% in *Table 11*, but to only 35% in the entire sample. However, with only 10 subjects, the higher number in *Table 11* might be due to chance. So far, we have no plausible explanation as to why the model should work less well in the case of informants living in a village. Nevertheless, it would be interesting to examine the subjects in Table 11 more closely as to the effect of variables not taken into account in the present article.

4. Discussion and Summary

Based on field research in the area of Tolna county, the knowledge of proverbs among 71 Hungarian residents of Tolna county was analyzed in the present study. The following factors were examined for their potential influence on proverb knowledge: 'age', 'level of education',

'sex', 'present residence' (city/town vs. village), 'residence up to the age of 18' (city/town vs. village); 'county of residence up to the age of 18' (Tolna vs. non-Tolna). To analyze the joint effect of these variables, regression analysis on the basis of the General Linear Model was used.

An attempt was made to show that the less complex statistical analyses usually employed in paremiological research may lead to erroneous conclusions. Nevertheless, we are well aware of the fact that an exploratory regression approach as used in the present study is fraught with problems as well. A causal modeling approach would have been more satisfactory, for example on the basis of LISREL (for a brief overview of structural equation modeling cf. Pedhazur/Pedhazur Schmelkin 1991, chap. 24). However, since, at present, our understanding of the various factors influencing knowledge of proverbs is still very poor, an exploratory approach was deemed more appropriate.

The analyses showed that many proverbs are still generally known in contemporary Hungarian culture: on average, the participants of the study completed 311.92 (i.e., 82.5%) of the 378 proverbs presented. It should be kept in mind, however, that the present sample of proverbs was not optimally suited for discriminating between informants with a higher knowledge of proverbs. This fact might also have influenced the results of our analyses. Therefore, in future sociolinguistic research of the present kind, more 'difficult' items, that is, proverbs known to only a small number of informants, should be included as well.

With regard to the factors possibly influencing proverb knowledge, the results are as follows:

(1) Age

The decisive factor turned out to be 'age'. This corresponds to the results previously obtained for German (Grzybek 1991a: 249f.), Croatian (Grzybek et al. 1993) and Czech (Schindler 1993) proverbs. Thus, there seems to be the following general tendency: with an increase in age, the amount of "correctly" completed proverbs increases in a linear way.[23]

(2) Level of Education

The second most important factor of influence is the educational level of the respondents. However, in contrast to the results obtained by Grzybek (1991a) for German proverbs, where the effect of the educational level was significant when assessed by a simple ANOVA, but

became nonsignificant when the factor 'age' was introduced in the analysis, in the present study, the impact of 'education' *increased* when considered jointly with 'age'. More specifically, respondents who graduated from university or college know more proverbs than those who attended only primary or secondary schools. Similarly, Schindler (1993: 142f.), in his work on Czech proverbs, found higher and university education to be a significant factor. Therefore, if the re-analysis of Grzybek's data currently undertaken by Grotjahn (cf. Grotjahn 1994) confirms Grzybek's results, we are in need of an explanation for why 'education' does not have an impact in the case of German proverbs.

(3) Present Place of Residence

The third factor identified in the present study as having an impact on the knowledge of proverbs was the present residence of the informants. Respondents currently living in a town or city appear to know more proverbs than those from a village. However, the effect of the variable 'present residence' appeared to be pronounced only when 'education' was *not* taken into account; it drastically decreased when 'present residence' was considered jointly with 'education' and 'age' in the regression analysis. This latter result can be explained by the fact that 'present residence' and 'education' are intercorrelated: people from a town or a city have a higher level of education than those from a village. Thus, ultimately, 'education' appears to be the decisive factor. Nevertheless, there seems to be a slightly independent effect of the factor 'present residence' – at least if poor sampling could be ruled out as a potential source of bias. With regard to 'present residence' the findings of the present study contradict both Grzybek's (1991a: 248) results for German and Schindler's (1993: 144) for Czech proverbs, according to which the factor 'present residence' does not have any significant influence on proverb knowledge. This divergence points to the necessity of replication and further analysis.

(4) Sex

In her preliminary presentation of the Hungarian data, Tóthné Litovkina (1992: 298f.) reported that men obviously knew more proverbs than women. This in line with the German pilot study where men knew more proverbs than women. However, the effect found in the German pilot study (cf. Grzybek 1991a: 247) was not statistically significant ($F_{1;123} = 2.63$, $p=0.11$), although the probability nearly reached the conventional significance level of .10 often used in pilot studies. Similarly,

Grzybek et al. (1994) and Schindler (1993: 141) found no effect of the variable 'sex' for the Croatian and Czech data, respectively. In the present Hungarian sample the simple main effect of 'sex' is significant at the .10 level ($p=0.09$; cf. *Table 5*). Furthermore, if the *F*-test is viewed as one-tailed, the effect is significant even at the .05 level. However, at least in the present study 'sex' is interrelated with 'education' (males are better educated than females), and the effect of 'sex' vanishes when the variable is included in the regression analysis together with 'education'. Thus 'education' appears again to be the decisive variable.

(5) Place of Residence until the Age of 18

The effect of the place of residence until the age of 18 turned out to be nonsignificant in all our analyses. Furthermore, this variable is strongly interrelated with the present place of residence: respondents currently living in a city or a town tend to have lived in a city or town in their childhood and youth as well. The nonsignificant result for this variable is in line with that obtained for German (Grzybek 1991a: 248), Croatian (Grzybek et al. 1993: 90f.), and Czech proverbs (Schindler 1993: 144).

(6) County of Residence until the Age of 18

In our analyses, the simple main effect of the variable 'county of residence in youth' proved to be highly significant (cf. *Table 5*). However, the variable is strongly interrelated with the remaining variables (cf. *Table 2*). If the intercorrelations are accounted for in the regression analysis, the effect of 'county of residence in youth' vanishes. This again demonstrates that a statistical analysis which does not take into account the joint influence of variables may lead to erroneous conclusions.

To sum up:

According to the regression analyses, knowledge of proverbs seems to depend first of all on the age of the respondents, then on their educational attainment. Furthermore, there seems to be a slightly independent effect of the variable 'present place of residence'. The remaining variables, that is the sex of the informants and their place of residence until the age of 18 (Tolna county vs. non-Tolna county; town/city vs. village), did not appear to have any impact. However, in view of the various methodological shortcomings of the present study, further research is needed before any definite conclusions can be drawn with regard to the factors influencing proverb knowledge.

To many Hungarians at least, and to many proverb scholars and folklorists, too, the results of this study will probably appear unbelievable: they spoil their stereotype of the "ideal knower of proverbs" as an old, uneducated peasant woman, spouting proverbs, "the people's wisdom, the unencased diamonds, the golden apple on a silver saucer" (*A magyarság szellemi néprajza*, 1935: 402).

Acknowledgements

We wish to express our gratitude to everybody who has helped us with our research. First of all, we want to express our appreciation to all those colleagues, friends, students, and unknown people, who devoted their time to filling in the questionnaires; without their kind help, this study would never have come to light.

Notes

1 See the bibliographies in Grzybek (ed.) (1984) and in Permjakov (1988).

2 After Permjakov's death, some statistical analyses of the Russian data were done by the Estonian paremiologist Arvo Krikmann (1986). In fact, Krikmann's statistical analyses have to be considered as a pioneering attempt in establishing statistical methods in empirical paremiology. Still, the results of his analyses are limited in relevance, due to various restrictions, which cannot be discussed here in detail (cf. Grzybek 1991a).

3 This project, entitled "Sprichwörter-Minima im Deutschen und Kroatischen", was financially supported by the Deutsche Forschungsgemeinschaft (Bonn) from 1991-93.

4 As compared to the two presentations mentioned above (Tóthné Litovkina 1992, 1993), the number of respondents has been changed from 69 to 71: eight subjects were added, six subjects with less than 18 years of age were excluded. The present version does not include chapters about the coding system of the data, about who knows the most and the fewest proverbs. An updated list of the most popular Hungarian proverbs is included in the Appendix (see also Section 3.2. below).

5 In fact, a list of 380 proverbs was presented to the subjects. In two cases, different numbers were given to two variants of a proverb, so that two proverbs were included twice; these two variants were counted as one proverb after a cross-check of the answers given to any of the two variants.

6 In 1991, 253.700 people lived in Tolna county; among them, there were 131.350 females (51.8%) and 122.350 males (48.2%). – For the sources of these data see: *Tolna megye statisztikai évkönyve 1990*. Szekszárd, 1990.

7 With only a few exceptions, almost all proverbs of (a) and (b) were also contained in traditional proverb collections.

8 Unfortunately, Nagy did not give any reason for how he arrived at this classification; probably, it is based on his personal intuition.

9 Vöő's collection is based on items sent to her by Hungarian residents in Romania; since the number of respondents who sent in a given proverb is indicated, the 200 most frequent proverbs were taken.

10 Not all 4,000 proverbs were given to an identical number of respondents; furthermore, no respondent received more than 2,000 items.

11 A simple categorization system was used for analysis (see Tóthné Litovkina 1992), based on the suggestions made by Grzybek (1991a). This system does not pay attention to verbal variations of the proverbial surface structure. Meanwhile, a comprehensive classificational system has been developed by Grzybek et al. (1994) in order to categorize proverb variants which are obtained during empirical research.

12 In Hungarian, there is only one word ("város") to denote both towns and cities. Although there is, in principle, a distinction between locations with either more or less than 100,000 inhabitants, both settlement types have been grouped together for the purposes of the present study, in contradistinction to villages. In this understanding, there are seven towns in Tolna county (the biggest is Szekszárd with 36,943 inhabitants), and 101 villages. – (For the sources of this information, see note 6.)

13 Strictly speaking, it is not justified to speak of "correct" or "proper" completions, on the one hand, and "incorrect" completions, on the other hand. If at all, such classifications may be employed only with regard to a particular expected form (as in the Hungarian experiment). In a strict empirical approach, all variants given by the respondents would have to be equally treated as to their "correctness"; then, only more or less *frequent* completions may be distinguished from each other.

14 If we split EDUCATION into two dummy variables and calculate a multiple regression analysis, we obtain a multiple correlation (R) of .254, and an adjusted R^2 of .037. (The value for the overall F-test for regression is the same as that obtained through ANOVA.)

15 Good and rather elementary introductions to regression analysis are Norušis (1990b, chap. B6) and Tabachnik/Fidell (1989, chap. 5). Much more comprehensive treatments of regression and GLM are, e.g., Pedhazur (1982) or Darlington (1990), the latter being more difficult to read. See also Kockläuner (1988), Pedhazur/Pedhazur Schmelkin (1990, chaps. 17-21), and Grotjahn (1992) for further information.

16 For the regression model including all independent variables the values for the index of tolerance provided by SPSS as a measure of collinearity range from .59 to .81. When only AGE and EDUCATION are included, the tolerance is .91 for both variables. When only AGE and RESID-PRES are included, it rises to .98. Note that the higher the multiple correlation of a variable with the rest of the independent variables, the closer to 0 is the tolerance. (The tolerance of a variable is defined as 1 minus the squared multiple correlation of the corresponding variable with the remaining predictors.)

17 To increase the fit to normality we transformed the variable AGE in various ways. None of the transformations led, however, to a better fit.

18 The term "easy" has to be understood here as it is used in classical test theory.

19 Note that for the model eventually adopted, containing only EDUCATION and AGE as predictors, the check for homogeneity of slopes yielded the following nonsignificant result: $F_{2;\,65} = 1.60;\ p = .21$.

20 Working with dummy variables has the advantage that the regression analysis also provides information about the effect of the different levels of EDUCATION.

21 We are well aware of the fact that the exclusion of RESID-PRES may be tantamount to committing a specification error, and hence to estimating and testing a mispecified model; cf. the discussion in Grotjahn (1992).

22 Note that the subject identification numbers are not consecutive because subjects aged less than 18 (e.g., subject #7) have been excluded from the analyses. Further, note that a standardized residual of, say, + 2.0 indicates an inaccurate prediction.

23 It would be important to control the degree of literacy, especially among the older population; it might well be the case, for example, that literate people are overrepresented in the sample as compared to general standards. Unfortunately, no relevant information as to this topic has been available to us.

References

Bálint, Sándor (1972). *Szegedi példabeszédek és jeles mondások*. Budapest

Baur, R.S.; Grzybek, P. (1990). "Empirische Untersuchungen zu einem Sprichwörter-Minimum des Deutschen". In: Spillner, B. (ed.). *Interkulturelle*

Kommunikation. Kongreßbeiträge zur 20. Jahrestagung der GAL. Frankfurt/M., 220-223

Bliesener, T. (1992). "Korrelation und Determination von Konstrukten. Zur Interpretation der Korrelation in multivariaten Datensätzen". *Zeitschrift für Differentielle und Diagnostische Psychologie* 13: 21-33

Chlosta, Ch.; Grzybek, P. (1994). "Zur Methodologie der Untersuchung der Bekanntheit von Sprichwörtern". In: Grzybek, P.; Chlosta, Ch. (eds.). *Empirische Sprichwortforschung*. Bochum [in preparation]

Darlington, R.B. (1990). *Regression and Linear Models*. New York

Erdélyi, J. (1851). *Magyar közmondások könyve*. Pest

Grotjahn, R. (1992). "Evaluating the adequacy of regression models: Some potential pitfalls". In: Rieger, B. (ed.). *Glottometrika* 13. Bochum, 121-172

Grotjahn, R. (1994). "Determinanten der Bekanntheit deutscher Sprichwörter: Eine Re-Analyse der Daten von Grzybek (1991)". In: Grzybek, P.; Chlosta, Ch. (eds.). *Empirische Sprichwortforschung*. Bochum [in preparation]

Grzybek, P. (1984a). "Überlegungen zur semiotischen Sprichwortforschung". In: Grzybek 1984: 215-249. [Engl.: "Foundations of Semiotic Proverb Study". In: *Proverbium. Yearbook of International Proverb Scholarship*, vol. 4/1987: 39-85. – Repr. in: Mieder, W. (ed.). *Modern Proverb Research*. New York 1993 (in print)]

Grzybek, P. (1984b). "How to do things with some Proverbs. Zur Frage eines parömischen Minimums". In: Grzybek (ed.) 1984: 345-350

Grzybek, P. (ed.) (1984). *Semiotische Studien zum Sprichwort. Simple Forms Reconsidered I.* Herausgegeben von Peter Grzybek unter Mitarbeit von Wolfgang Eismann [= Special Issue of: *Kodikas/Code – Ars Semeiotica. An International Journal of Semiotics* (3/4) 1984]

Grzybek, P. (1991a). "Sinkendes Kulturgut? Eine empirische Pilotstudie zur Bekanntheit deutscher Sprichwörter". *Wirkendes Wort* 41(2): 239-264

Grzybek, P. (1991b). "Einfache Formen der Literatur als Paradigma der Kultursemiotik". In: Grzybek, P. (ed.). *Cultural Semiotics – Facts and Facets / Fakten und Facetten der Kultursemiotik*. Bochum, 45-61

Grzybek, P.; Chlosta, Ch. (1993). "Empirische Sprichwortforschung". In: *Proverbium. Yearbook of International Proverb Scholarship*, vol. 10. The University of Vermont, 89-128

Grzybek, P.; Chlosta, Ch; Roos, U. (1994). "Ein Vorschlag zur Klassifikation von Sprichwortvarianten bei der empirischen Sprichwortforschung". In: Sandig, B. (ed.). *Tendenzen der Phraseologieforschung. Europhras 1992*. Bochum [in print]

Grzybek, P.; Škara, D.; Heyken, Z. (1993). "Der Weisheit der Gasse auf der Spur. Eine empirische Pilotstudie zur Bekanntheit kroatischer Sprichwörter". *Zeitschrift für Balkanologie* 29(2): 85-98

Kockläuner, G. (1988). *Angewandte Regressionsanalyse mit SPSS*. Braunschweig

Krikmann, A.A. (1986). *Paremiologičeskie èksperimenty G.L. Permjakova*. Tallin

A magyarság szellemi néprajza (1935), v. III

Margalits, E. (1897). *Magyar közmondások és közmondásszerü szólások*. Budapest

Mieder, W. (1985). "Neues zur demoskopischen Sprichwörterkunde". In: *Proverbium. Yearbook of International Proverb Scholarship*, vol. 2. The Ohio State University, 307-328

Mieder, W. (1986). [Rez.:] Grzybek, P. (Hrsg.) (1984): *Semiotische Studien zum Sprichwort. Simple Forms Reconsidered I.* [= Special Issue of: *Kodikas/Code – Ars Semeiotica. An International Journal of Semiotics*, (3/4) 1984.] In: *Proverbium. Yearbook of International Proverb Scholarship*, vol. 3. The Ohio State University, 311-323.

Mieder, W. (1990). "Prolegomena to Prospective Paremiography". In: *Proverbium. Yearbook of International Proverb Scholarship*, vol. 7. The University of Vermont, 133-144

Norušis, M.J. (1990a). *SPSS/PC+ 4.0 Base Manual for the IBM PC/XT/AT and PS/2*. Chicago

Norušis, M.J. (1990b). *SPSS/PC+ Statistics 4.0 for the IBM PC/XT/AT and PS/2*. Chicago

Norušis, M.J. (1991). *The SPSS Guide to Data Analysis for SPSS/PC+ 4.0*. Chicago (2nd ed.)

O. Nagy, G. (1976). *Magyar szólások és közmondások*. Budapest.

Paczolay, G. (1987). *Magyar-észt-német-anglo-finn-latin közmondások és szólások – cseremisz és zürjén függelékkel*. Veszprém

Paczolay, G. (1989): *Magyar közmondások és szólások – Hungarian Proverbs*. Veszprém

Pedhazur, E.J. (1982). *Multiple Regression in Behavioral Research. Explanation and Prediction*. Fort Worth (2nd ed.)

Pedhazur, E.J.; Pedhazur Schmelkin, L. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, N.J.

Permjakov, G.L. (1971). *Paremiologičeskij èksperiment. Materialy dlja paremiologičeskogo minimuma*. Moskva

Permjakov, G.L. (1973). "K voprosu o paremiologičeskom urovne jazyka". In: *Sbornik statej po vtoričnym modelirujuščim sistemam*. Tartu, 26-33

Permjakov, G.L. (1982). "K voprosu o russkom paremiologičeskom mini-

mume". In: *Slovari i lingvostranovedenie. Pod red. E.M. Vereščagina.* Moskva, 131-137.

Permjakov, G.L. (ed.) (1984). *Paremiologičeskie issledovanija.* Moskva

Permjakov, G.L. (1985). *300 allgemeingebräuchliche russische Sprichwörter und sprichwörtliche Redensarten (300 obščeupotrebitel'nych russkich poslovic i pogovorok).* Moskva-Leipzig

Permjakov, G.L. (1988). *Osnovy strukturnoj paremiologii.* Moskva

Ruef, H. (1989): "Zusatzsprichwörter und das Problem des parömischen Minimums". In: Gréciano, G. (ed.) (1989). *Europhras 88. Phraséologie Contrastive.* Strasbourg, 379-385

Schellbach-Kopra, I. (1987). "Parömisches Minimum und Phraseodidaktik im finnisch-deutschen Bereich". In: Korhonen, J. (ed.). *Beiträge zur allgemeinen und germanistischen Phraseologieforschung.* Oulu, 245-255

Schindler, F. (1993). *Das Sprichwort im heutigen Tschechisch. Empirische Untersuchung und semantische Beschreibung.* München

Tabachnick, B.G.; Fidell, L.S. (1989). *Using Multivariate Statistics.* New York (2nd ed.)

Tóthné Litovkina, A. (1989). "Kísérlet a magyar és orosz közmondások nyelvi elemzésére". *Magyar Nyelv* 1: 58-69

Tóthné Litovkina, A. (1989a). "Analiz jazyka vengerskich poslovic (Russkie ekvivalenty vengerskim poslovicam). A magyar közmondások elemzése nyelvészeti szempontból (A magyar közmondások orosz megfelelői)". In: *Proverbium Paratum* 4. Budapest

Tóthné Litovkina, A. (1992). "The Use of Proverbs In Contemporary Hungarian Society. A Study of National Identity Among Hungarians of Tolna County Through Proverbs". *S – European Journal for Semiotic Studies* 4(1,2): 289-316

Tóthné Litovkina, A. (1993). "Milyen és mennyi közmondás él a Magyar köztudatban". In: *Régi és új peregrináció Magyarok külföldön. Külföldiek Magyarországon (A III. Nemzetközi Hungarológiai Kongresszuson – Szeged, 1991 augusztus 12-16 – elhangzott előadások).* Budapest–Szeged, III. kötet, 1452-1465

Vöő, G. (1989). *Igaz ember igazat szól. Közmondások a romániai magyar folklórból.* Bucureşti

Appendix: The Best-Known Hungarian Proverbs

1. *Addig jár a korsó a kútra, míg el nem törik. [The pitcher goes so long to the well until it breaks.]*

2. *Addig nyújtózkodj, ameddig a takaród ér. [Stretch yourself as far as your cover reaches.]*
3. *Ahány ház, annyi szokás. [So many houses, so many customs.]*
4. *Ajándék lónak ne nézd a fogát. [Don't look at the tooth of a gift horse.]*
5. *Aki kíváncsi, az hamar megöregszik. [Who is curious, will soon get old.]*
6. *Aki másnak vermet ás, maga esik bele. [He who digs a pit for another, falls into it himself.]*
7. *Aki mer, az nyer. [He who ventures, wins.]*
8. *Aki a virágot szereti, rossz ember nem lehet. [Who loves a flower, can't be a bad person.]*
9. *Az alma nem esik messze a fájától. [The apple does not fall far from its tree.]*
10. *Ami késik, nem múlik. [What is delayed is not lost.]*
11. *Amilyen az adjonisten, olyan a fogadjisten. [Like greeting (literally: God should give!), like answer (literally: God should accept it!).]*
12. *Amit ma megtehetsz, ne halaszd holnapra. [What you can do today do not put off till tomorrow.]*
13. *Egy fecske nem csinál nyarat. [One swallow does not make a summer.]*
14. *Egyszer volt Budán kutyavásár. [There has been a dog-market in Buda once. (= A favourable opportunity comes only once).]*
15. *Evés közben jön meg az étvágy. [Appetite comes while eating.]*
16. *Gyakorlat teszi a mestert. [Practice makes the master.]*
17. *Ha nincs ló, jó a szamár is. [If there is no horse, a donkey is good as well.]*
18. *Jobb félni, mint megijedni. [It is better to fear than to get frightened.]*
19. *Jobb későn, mint soha. [Better late than never.]*
20. *Jóból is megárt a sok. [Even of a good thing too much is harmful.]*
21. *Kettőn áll a vásár. [The agreement depends on two (persons).]*
22. *Ki mint vet, úgy arat. [As one sows, so one reaps.]*
23. *Ki mint veti ágyát, úgy alussza álmát. [As one makes one's bed, so one 'sleeps one's dream'.]*
24. *Kicsi a bors, de erős. [Pepper is small but hot.]*
25. *Lassan járj, tovább érsz. [Go slowly, you get farther.]*
26. *Minden csoda három napig tart. [Every wonder lasts three days.]*
27. *Minden jó, ha a vége jó. [All is good if the end is good.]*
28. *Minden zsák megtalálja a maga foltját. [Every sack will find its patch.]*
29. *Ne igyál előre a medve bőrére. [Don't drink in advance on the hide of a bear.]*

30. *Ne szólj szám, nem fáj fejem. [Don't speak my mouth, I won't have a headache.]*
31. *Nem mind arany, ami fénylik. [All is not gold that glitters.]*
32. *Nem a ruha teszi az embert. [Clothes do not make the man.]*
33. *Néma gyereknek az anyja sem érti a szavát. [Even a mother can't understand the word of a dumb child.]*
34. *Nyugtával dicsérd a napot. [Praise the day at sunset.]*
35. *Rend a lelke mindennek. [Order is the soul of everything.]*
36. *Sándor, József, Benedek zsákban hozzák a meleget. [Alexander, Joseph, Benedict will bring warm (weather) in a sack.]*
37. *Sokat akar a szarka, de nem bírja a farka. [The magpie wants too much, but her tail is unable to carry it.]*
38. *A szegény embert még az ág is húzza. [Even the branch of a tree pulls a poorman.]*
39. *Tévedni emberi dolog. [To err is a human matter.]*
40. *Tisztaság fél egészség. [Cleanliness is half of health.]*
41. *Többet ésszel, mint erővel. [More by wisdom than by force.]*
42. *A tűzzel nem jó játszani. [It is not good to play with fire.]*