========== /// ==========

A POPULARIDADE DOS PROVÉRBIOS
UM ESTUDO DE CASO DA RELAÇÃO FREQUÊNCIA-FAMILIARIDADE PARA OS ALEMÃES

*Peter **GRZYBEK**, Institut für Slawistik, Universität Graz, **Austria***

Resumo

Durante estes últimos anos, algumas abordagens empíricas e experimentais sobre os provérbios tornaram-se uma cada vez mais importante área de investigação. A sua relevância pode encontrar-se num material empiricamente avaliado como uma base para finalidades a definir futuramente, desde o ensino de língua estrangeira à paremiografia e paremiologia (comparativas). A vasta gama de interesses e objectivos - variando da ("mínima") necessidade para um número limitado de provérbios familiares até ao ("máximo") interesse num dado "provérbio mínimo" de uma cultura (i.e., uma alegada "totalidade"dos seus provérbios mais comuns) – resume-se à pergunta, numa dada cultura, de quem sabe quais os provérbios, em que formas verbais e do que depende este conhecimento. Esta formulação inclui o estudo de factores possíveis correlacionando com o conhecimento dos provérbios, incluindo as características pessoais e sociológicas (sexo, idade, educação do falante, etc.) assim como as características linguísticas dos itens proverbiais (complexidade sintáctica, imaginação, ritmo, tamanho das frases, etc.). A apresentação levanta sérios problemas metodológicos para os estudos comparativos, concentrando-se principalmente na inter-relação entre os factores baseados nos temas e os factores linguísticos.

---------- ||||| ----------

THE POPULARITY OF PROVERBS
A CASE STUDY OF THE FREQUENCY-FAMILIARITY RELATION FOR GERMAN

*Peter **GRZYBEK**, Institut für Slawistik, Universität Graz, **Austria***

Abstract

Over the last years, empirical and experimental approaches to proverbs have become an increasingly important research area. Their relevance is in empirically evaluated material as a basis for further purposes, from foreign language instruction to (comparative) paremiography and paremiology. The broad spectrum of interests and claims – ranging from the („minimal") need for a limited number of familiar proverbs up to the („maximal") interest in a given culture's 'proverb minimum' (i.e., an alleged „totality" of its most common proverbs) – is covered by the maximally condensed question of who, in a given culture, knows which proverbs in which verbal form and what does this knowledge depend upon? This formulation includes the study of possible factors correlating with proverb knowledge, including personal and sociological features (speakers' sex, age, education, etc.) as well as linguistic characteristics of the proverbial items (syntactic complexity, imagery, rhythm, sentence length, etc.). The presentation raises important methodological problems for comparative studies, particularly focusing the interrelation between subject-based and linguistics factors.

1. Introduction

The popularity of proverbs is an important issue not only for paremiography and paremiology, but for any research based on proverbs – be that of sociological, pedagogical, psychological, linguistic, or any other orientation. However, in defining popularity, and in tackling the specific (and, in fact, quite different questions) related, scholars have applied different methods. Comparing these methods, one should expect that each of them has been selected with regard to the specific research interest, and that they all have specific advantages and disadvantages. Apart from this fact, it is obvious that the results obtained may differ, more or less, depending on the concrete method chosen. As a consequence, in comparing the results, it should be interesting to see in how far they converge or diverge from each other – however, such a systematic comparison of results (and, by way of this, of the methods) has practically never been undertaken till today.

Such a comparison, which would ask for some kind of "meta-analysis" involving an external perspective, turns out to be quite difficult. One of the reasons for this is the fact that, more often than not, popularity as a concept has been differently understood or, to put it more directly, it has been ill defined. In particular, possibly (and probably) related, though essentially distinct categories have not been clearly distinguished: 'frequency' of proverbs, on the one hand, and 'knowledge' or 'familiarity' of proverbs, on the other. This lack of distinction is caused either by some kind of theoretical ignorance of definition, or by the assumption that both categories yield more or less identical results, an assumption which may be either implicitly contained or explicitly stated in the relevant studies.

It seems reasonable therefore, before any comparison of results can be undertaken, to analyze the concepts of familiarity and frequency in detail, specifically paying attention to possible relations or interactions between both concepts. Generally and strictly speaking, we are concerned here with basically different concepts which, in a first approach, have been termed *'usage-oriented'* vs. *'knowledge-oriented'* by Čermák (1997). By way of a general characterization, one may say that the first line of research is *frequency-oriented* and primarily *text-based*, whereas the second is *familiarity-oriented* and thus *knowledge-based*

and *subject-dependent*. With regard to the latter distinction, an additional terminological differentiation seems to be necessary and appropriate: whereas (individual or collective) **proverb knowledge** refers to a person's (or a group's) subjective acquaintance with proverbs, **proverb familiarity** refers to either individual proverbs or to a group of proverbs, in any case denoting average familiarity in a given collective. Regardless of the fact that, within a given group, the result of collective knowledge of a given proverb material thus coincides with collective familiarity, we are concerned with two essentially different perspectives. Simplifying matters, one may thus say that frequency-oriented studies are based on the analysis of text sources, either written or spoken, whereas familiarity-oriented and knowledge-oriented methods are based on asking persons, in one way or another.

Of course, the distinction of text and usage, on the one hand, and subject and knowledge, on the other, is rather rough and may be differentiated in more detail; in fact, within both approaches, there are a variety of further distinctions. Thus, with regard to *text-oriented* approaches, the following major methods have been hitherto predominantly applied in paremiology:

1. Documentation of the frequency of proverbs in "everyday life".– This method asks for comprehensive longitudinal field studies, such as, e.g., Hain's famous study from the 1950s – studies which will hardly ever be repeated in our time.

2. Analyses of written paremiographical sources, mainly of

(a) proverb collections, or

(b) archive material.

The analysis of frequency data in such sources seems to be quite common: the more often a proverb is documented (in a given geographical area, in a number of different sources, etc.), the more widespread and, as a consequence, the more familiar it is considered to be. Whereas in case of archive material this assumption still might be reasonable, its justification with regard to proverb collections clearly depends on the quality of these collections, since compilers, as we know, have always tended to simply copy proverbs from previous collections, or even translate them from other languages.

3. Analyses of mass media, mainly printed media.– Obviously, this method, which has been repeatedly applied since the 1970s, can grasp only a particular segment of language use, namely, that of public speech from journalistic discourse. A more modern (and increasingly favored method) is the analysis of internet sources, which again is problematic, since the quality of the sources can hardly be controlled.

4. Corpus-based analyses.– With the option to chose between different discourse types (such as literary, journalistic, or even spoken language, etc.), such analyses have the advantage of electronic search and retrieval strategies. As compared to unspecified internet (re-)search, the sources are rather well-defined, but it is a necessary condition, of course, that one previously knows and exactly defines one's search strategies...

As compared to this, *subject-based* approaches might be differentiated as follows:

1. In attempts to simply find some well-known proverbs, it has been regarded to be sufficient to ask a particular number of subjects to write down those proverbs which spontaneously come to their minds. In fact, this method may suffice for this specific interest. Yet, we know today that this approach usually yields only a limited number of prototypical proverbs (usually not more than 30-50 proverbs per person). The reason for this limitation seems to be that proverbs, by definition, tend to be used only in and with reference to specific situations and, as a consequence, are remembered only with regard to or "triggered" by these specific situations.

2. As an alternative, more or less comprehensive lists of selected proverbs have been presented to subjects asking them for their introspective intuition if the presented proverbs were familiar to them or not. This method – which shall be called 'full text presentation' (FTP) in this article – thus demands subjects to make a clear binary YES-NO decision. It has the advantage that people will recognize and recall also such proverbs which do not spontaneously come to their minds; on the other hand, this method has two obvious disadvantages: first, results of introspection-based studies may be misleading since subjects may only think that they know a given proverb (or appreciate it as being 'correct'), but in fact do not; and secondly, subjects may know a given proverb in a more or less divergent verbal form (i.e., some kind of variant or variation), and therefore, in this case, may be uncertain as to a correct decision.

3. In principle, the same objections hold true for scaling techniques when subjects are presented with a list of proverbs, the task providing a scale (e.g., from 1 to 7), on which individual proverb familiarity must be rated. This method shall be termed 'full text rating' (FTR), in this article; in addition to the problems listed above (2), individual differences in rating may come into play, and they must additionally be very carefully controlled.

4. A method which tries to avoid the problems outlined includes the presentation of only the beginning of a given proverb; the subjects' task then is to complete the text (e.g.: *Out of sight* ...). This method, which has been increasingly favored in empirical and experimental paremiology over the last years, shall be called 'partial text presentation' (PTP), here.

Of course, both perspectives (i.e., orientation on either frequency or familiarity) are likely to interact, in one way or another; they may even yield partly converging results. Yet, both approaches should methodologically be clearly kept apart. To put it as clear as possible: it is one thing to study the frequency occurrence of proverbs, and another thing to ask a person if s/he considers a proverb to be frequently used...

The objective of the present article is to present some comparative results based on methods including either frequency or familiarity orientation. In pursuing this question, it is desirable and necessary, of course, to have frequency and familiarity studies with results based on one and the same proverb material. However, there are only few systematic studies on proverb familiarity, there are not really more on proverb frequency, and there are hardly any which have ever attempted to relate the results obtained to each other – and this is the state of the art, whatever language may be concerned. Thus, it is not surprising that one is faced with the situation that in some studies there are proverbs for which frequency data (but not familiarity data) are available, and in others there are proverbs with familiarity data, but no frequency data. There are, of course, proverbs for which both kinds of data are available, but these proverbs have to be carefully selected item per item from one of the mentioned studies, and they can only then be compared with regard to the frequency-familiarity relation. Under these circumstances, at the present state of the art, it is hardly possible of course, to take into consideration the possible influence of the specific methods differentiated above within either frequency or familiarity orientation; rather, some general ideas can be presented providing a methodological basis for future research.

Assuming both categories, frequency and familiarity, do not simply reflect one and the same state of affairs and, as a consequence, do not yield identical results, a first decision has to be made as to the direction of dependence: does frequency (FRQ) depend on familiarity (FAM), or is familiarity dependent on frequency? In this respect, it has recently been argued (Grzybek 2008) in favor of the notion of some kind of self-regulating circle: an increase of proverb usage (i.e., heightened frequency) leads to an increase of individual and/or collective perceptibility of these items, which is the basis for higher familiarity, which then, in turn, finally results in an increase of usage, again. We would thus be concerned not with a simple unidirectional dependence, but rather with a more complex control cycle including something like a priming and an additional loop effect. Anyway, this would imply a primary dependence of FAM on FRQ; in mathematical terms, we would thus have FAM as a function of FRQ or, to put it differently, FAM as the dependent, and FRQ as the independent variable:

(1) $FAM = f(FRQ)$.

With this perspective, let us now turn to an analysis of German. Despite the overall lack of relevant studies, there is, interestingly enough, one study which has directly attempted to deal with the FAM-FRQ relation, with specific regard to German proverbs. Ďurčo (2005) analyzed 151 German proverbs, for which familiarity data were available. [1] For the sake of comparison with these familiarity data, Ďurčo conducted a corpus analysis of the well-known Mannheim *COSMAS II* corpus, thus establishing frequency occurrence for each of the 151 proverbs.

For the sake of better exemplification, the original frequency data were transformed into percentages, the largest absolute frequency being equaled with 100%. Familiarity was also calculated in percent, here ranging from 86.57% to 100%. Figure 1 illustrates the results obtained.

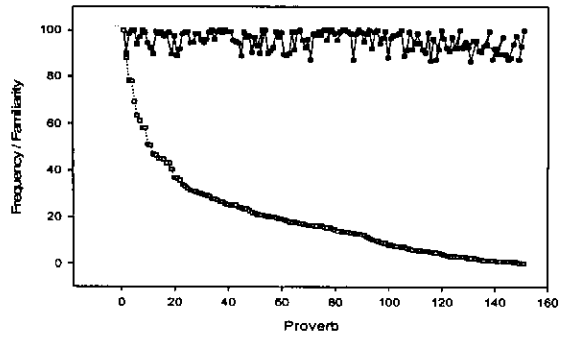

Fig. 1: Frequency (light squares) and familiarity (full black squares) data for 151 German proverbs

As can be seen, the results clearly differ for frequency and familiarity. In fact, Ďurčo (2005: 91), as a result of his comparison, arrived at the conclusion „[...] that the degree of familiarity and frequency in texts are no correlated entities".

At first sight, Ďurčo's conclusion seems to be reasonable – no functional relation can be detected between FAM and FRQ. However, as a re-analysis of Ďurčo's data or, in fact, a simple closer look at Figure 1 shows, this result is not really surprising. Rather, the reason for this outcome is quite obvious: analyzing highly familiar proverbs only, we get nothing to know about less familiar proverbs and, as a consequence, nothing substantial about the FAM-FRQ relation. In other words: analyzing familiar proverbs only, it will hardly be possible to find any reliable insight into the FAM-FRQ relation. The only conclusion to be drawn from Ďurčo's study – and this conclusion is valuable indeed –, is the fact that highly familiar proverbs may, but need not occur frequently. [2]

With this first result in mind, it turns out necessary to tackle the FAM-FRQ question anew, on a more systematic basis. Taking into account the fact that one needs both FAM and FRQ data for

one and the same proverb material, it might be worthwhile, for the sake of a first approximation, to re-analyze earlier data from the well-known Intermarket study (Hattemer/Scheuch 1983). In this study, 20 German proverbs were presented to 404 subjects, who were then asked which of these proverbs they know (FTP) and which of these proverbs they would use. It goes without saying that, statistically speaking, a sample of 20 items is a rather spare data base to arrive at far-reaching conclusions; additionally, as has been indicated above, both kinds of question are rather based on introspection than empirical testing and thus imply a large portion of subjectivity. Anyway, since knowledge and usage are generally kept apart, it seems worthwhile testing the FAM-FRQ relation with these data, which are presented in detail in Table 1.

Table 1: Percentages for 'Knowledge' and 'Usage' of 20 German proverbs

Proverb	**Knowledge**	**Usage**
	$N = 404$	$N = 404$
Zeit ist Geld.	88	41
Wenn einer eine Reise tut...	79	18
Sicher ist sicher.	81	45
Spare in der Zeit, dann hast du in der Not.	77	19
Doppelt gibt, wer gleich gibt.	19	4
Freunde in der Not...	42	10
Abends wird der Faule fleißig.	74	30
Durch Schaden wird man klug.	88	38
Was man Schwarz auf Schwarz besitzt...	53	12
Bei Geldsachen hört die Gemütlichkeit auf.	52	11
Ein Unglück kommt selten allein.	87	38
Haste was, biste was.	79	12
Wer Brot hat, dem gibt man Brot.	20	2
Nachher ist man immer klüger.	59	26
Morgenstund hat Gold im Mund.	90	31
Die dümmsten Bauern haben die dicksten Kartoffeln.	84	42
Schadenfreude ist die reinste Freude.	73	18
Über Geld spricht man nicht.	75	19
Geteiltes Leid ist halbes Leid.	81	21
Was man hat, das hat man.	72	21
	68.65	22.90

As becomes evident as first sight, the average values for knowledge and usage clearly differ, those for usage being definitely lower. This clear difference of results may in fact be interpreted in terms of a clear distinction of both categories on behalf of the subjects. Fig. 2 illustrates the result, with the percentages for usage on the horizontal axis, for familiarity on the vertical.

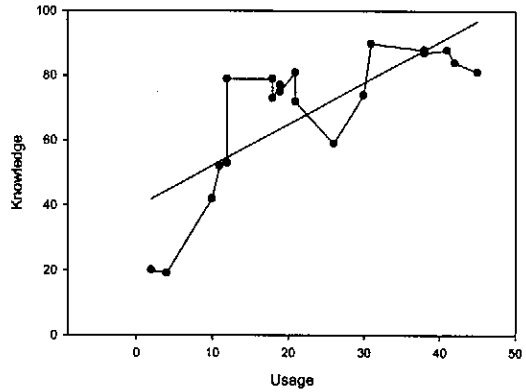

Fig. 2: Usage and knowledge data for 20 German proverbs

As can be seen from Fig. 2, there indeed seems to be some general tendency, saying that there is some increase in knowledge with an increase of usage. Trying to interpret this tendency by way of a functional relation between both categories, it seems obvious that this relation is unlikely to be of a linear kind. Such a linear model would explain only the linear proportion of the relation to be explained; likewise, the calculation of a correlation coefficient would be inadequate. Rather, it seems necessary to find a nonlinear model which might explain the specific FAM-FRQ relation.

Attempting to find such a nonlinear model, one might proceed by way of mere induction and empirical trial-and-error testing; this can easily be done with the help of modern software tools and iterative methods. Of course, one would tend to prefer a model with a better fitting result; usually goodness of fit is referred to by the determination coefficient R^2, with $0 \leq R^2 \geq 1$, a larger R^2 value indicating a better fit of the model.

Table 1 lists some relevant one- or two-parameter models; Table 1 presents some very common models, with y being the dependent variable (i.e. FAM, in our case), and x the independent variable (i.e., FRQ). Also given are the corresponding parameter values (a and b)

for each model, and with the resulting R^2 values; the determination coefficient for the linear model, which is incorporated in Fig. 2, would be $R^2 = 0.60$, for the sake of comparison.

Table 2: The FAM-FRQ relation: some selected nonlinear functions

No.	Function	*a*	*b*	R^2
I	$y = a \cdot x^b$	22.55	0.37	0.75
II	$y = a \cdot ln(x)$	23.64	–	0.81
III	$y = a \cdot exp(-b/x)$	97.74	5.72	0.82
IV	$y = a + b \cdot ln(x) / x$	110.32	-254.01	0.84

As can be seen, all of these models result in a better fitter as compared to the linear model; by way of a synopsis, Fig. 3 illustrates the results. Interestingly enough, model (I) fails to be convincing, and has the lowest determination coefficient ($R^2 = 0.75$); this is rather surprising since, in a recent study on the FAM-FRQ relation of American proverbs (cf. Grzybek/Chlosta 2009), this model turned out to be very good, with $R^2 = 0.91$, though after some data pooling.

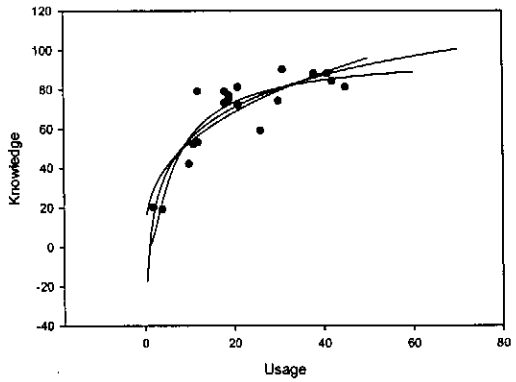

Fig. 3: Fitting nonlinear functions (cf. Table 2) to 20 German proverbs (cf. Table 1)

With the exception of the logarithmic model (II), they all have two parameters. It goes without saying that, theoretically speaking, many more models with even more parameters might be taken into consideration; in order to arrive at a qualitative interpretation one would prefer, however, a model with less parameters. Taken together, these observations ask for

some explanation – the crucial question is if any of these models lends themselves to some interpretation.

Obviously, FRQ is not the only factor influencing FAM. If FRQ were the only factor, we would obtain differential equations like

(2) $$\frac{dy}{dx} = \frac{1}{x}$$

or

(3) $$\frac{dy}{dx} = \frac{a}{x}$$

saying that the change of FAM is proportional to the relative rate of change of FRQ. The solution of differential equation (2) results in $y=ln(x)+C$, that of (3) in $y=a \cdot ln(x)+C$, or in the logarithmic model (II) with $C = 0$; C is the integration constant, here. However, as has been shown above, such a model turns out not to be adequate enough. Obviously, we need at least some kind of "brake" in our differential model; this might either be due to the fact that FAM does not increase ad infinitum, but is limited in extent on the given scale, or to further influencing factors.

With this in mind, let us consider the relative rate of change of y, i.e., dy/y; let us define this rate proportionally to the relative rate of change of x, from which we subtract some "brake factor"; the latter is similar in form, but has an additional constant b.

Thus elaborating (2), we obtain the differential equation

(4) $$\frac{dy}{y} = \left(\frac{1}{x} - \frac{1}{(b+x)}\right)dx,$$

the solution of which is obtained by integration on both sides, thus resulting in

(5) $$y = \frac{C \cdot x}{(b+x)},$$

which is also known as the Tornquist curve whose asymptote is C. And elaborating (3) we obtain

(6) $$\frac{dy}{y} = \left(\frac{a}{x} - \frac{1}{(b+x)}\right)dx,$$

the solution of which yields

(7) $$y = \frac{C \cdot x^a}{(b+x)}.$$

As can be seen, the numerator of (7) corresponds to model (1), which turned out to be adequate for the FAM-FRQ relation of American proverbs (see above), but (7) is additionally "slown down" by our braking component.

Let us, for the sake of simplicity, first test (5), which is also a special case of (7), with $a = 1$. If this model is satisfying enough, we will leave an analysis with (7) for future research. [3]

With parameter values for $C = 107.74$ and $b = 9.96$ we obtain a determination coefficient of $R^2 = 0.82$. Interestingly enough, R^2 is only slightly worse if we set $C = 100$, which is the maximum of FAM. In this case, only one parameter remains to be estimated; with $b = 7.90$, we thus obtain

$$(7a) \quad FAM = \frac{100 \cdot FRQ}{(7.90 + FRQ)}$$

which yields a result of $R^2 = 0.81$. [4] The result is graphically presented in Figure 4.

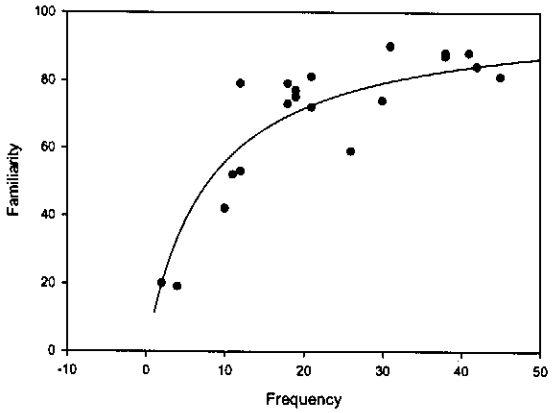

Fig. 4: 20 German proverbs from the Intermarket study: the dependence of FAM on FRQ according to (5) with $C = 100$ and $b = 7.90$

Now, let us test this model with the very same proverbs, but based on other methods of establishing frequency and familiarity. First, all twenty proverbs have been checked for frequency in the above-mentioned COSMAS II corpus. Frequency thus is occurrence-based, not intuitively estimated, as above. [5] The frequency data for all twenty proverbs are given in Table 3.

Table 3: 'Familiarity' and 'Frequency' Data for 20 German proverbs

Proverb	**Familiarity** $N = 125$	**Frequency** [abs.]
Zeit ist Geld.	85.48	303
Wenn einer eine Reise tut...	82.40	320
Sicher ist sicher.	–	475
Spare in der Zeit, dann hast du in der Not.	–	67
Doppelt gibt, wer gleich gibt.	12.80	8
Freunde in der Not...	27.20	9
Abends wird der Faule fleißig.	57.60	16
Durch Schaden wird man klug.	95.20	42
Was man Schwarz auf Schwarz besitzt...	–	21
Bei Geldsachen hört die Gemütlichkeit auf.	–	13
Ein Unglück kommt selten allein.	99.20	194
Haste was, biste was.	–	20
Wer Brot hat, dem gibt man Brot.	–	0
Nachher ist man immer klüger.	–	120
Morgenstund hat Gold im Mund.	99.20	120
Die dümmsten Bauern haben die dicksten Kartoffeln.	–	15
Schadenfreude ist die reinste Freude.	84.00	47
Über Geld spricht man nicht.	–	73
Geteiltes Leid ist halbes Leid.	81.60	97
Was man hat, das hat man.	–	32
	72.47	

Figure 5 offers a comparison between corpus-based frequencies (Table 3) and intuitively estimated frequencies of the twenty above-mentioned German proverbs (cf. Table 2).

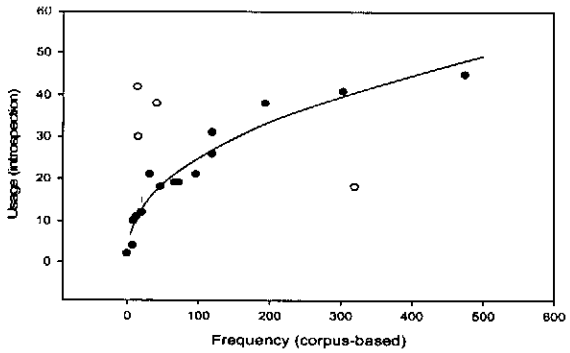

Fig. 5: Corpus-Based frequency of occurrence and intuitive frequency of usage

As can be seen, there is no complete functional relation for all data points; anyway, there seems to be a general non-linear tendency for the bulk of the data, four data points (marked by light circles in Fig. 5) can be regarded as outliners. Excluding these four data points from the analysis, the overall tendency can very well be grasped by the simple power function $y = a\,x^b$ (yielding a value of $R^2 = 0.94$). In one case ("*Wenn einer eine Reise tut* ..."), the observed frequency is much higher than expected; in the other three cases, frequency is lower – this may well be due to the fact that only "full proverbial quotations" were accepted as occurrences, and the lower frequency may well be a consequence of this decision. In any case, more detailed and systematic analyses will be necessary to arrive at reliable conclusions as to the relation between introspectively estimated usage and frequency of occurrence.

Interestingly enough, the relation between FTP and PTP methods to determine familiarity seems to be much clearer. Of the twenty proverbs presented above (cf. Table 1), ten were also part of a PTP study (Grzybek 1991. The results, based on the data represented in Table 3 (see above), are represented in Figure 6. Although we are concerned here with only ten items, the overall tendency seems to be quite clear: with an almost identical average of x = 72.10 (s = 22.15) for the FTP, and of x = 72.47 (s = 28.74) for the PTP – the difference is not significant–, there is a clearly expressed linear relation between the results from both methods ($r = 0.96$, $p < 0.001$). [6]

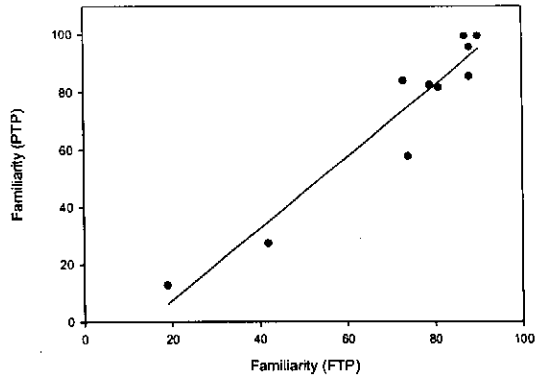

Fig. 6: Results for Familiarity – Comparison of FTP and PTP

With these results obtained, we can return to the FRQ-FAM relation, now with regard to the data based on corpus analysis and PTP. As can be seen from Table 3, there are 10 proverbs for which both frequency and PTP familiarity data are available. Fitting model (5) outlined above to these data, yields a value of $R^2 = 0.81$ (with parameter values $C = 101.12$ and $b = 15.66$), a result which is almost identical with the one obtained above. Setting $C = 100$, which is the maximum of FAM, and with parameter $b = 15.17$, this result remains unchanged. The result is graphically presented in Figure 7.

(7b) $$FAM = \frac{100 \cdot FRQ}{(15.17 + FRQ)}$$

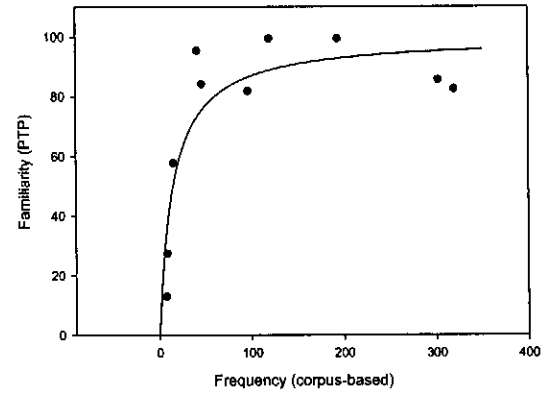

Fig. 7: The dependence of FAM (PTP) on FRQ (corpus) according to (5) with $a = 100$ and $b = 15.17$ for ten German proverbs.

Conclusions

Due to the lack of previous research in this field, the present contribution cannot arrive at far-reaching contributions – it will be an important task for future research to provide and analyze further (and sufficient) data, which may serve as material to test the hypotheses brought forth here. Notwithstanding these lacunae, some general conclusions may be drawn:

1. The notion of proverb popularity should be more consequently differentiated in paremiology, clearly keeping apart *'usage-oriented'* vs. *'knowledge-oriented'* approaches. One line of research is rather *frequency-oriented* and primarily *text-based*, the other one is *familiarity-oriented*, *knowledge-based* and *subject-dependent*.

2. Familiarity and frequency do not denote the same; both concepts cannot simply be replaced one by another; they do not represent two faces of one and the same coin, rather two different coins joint together, resulting in a value calculated according to a specific exchange rate (simple correlation coefficient or linear models are not at stake here).

3. There is clear evidence for a specific relation between proverb frequency and proverb familiarity:

a. Familiar proverbs may occur frequently

b. Familiar proverbs need not occur frequently

c. Frequent proverbs tend to be familiar

4. For any study of the FAM-FRQ relation of proverbs, data must not only be sufficient, they must also represent the whole spectrum of familiarity, and must not concentrate on familiar proverbs, only.

5. The FAM-FRQ relation of proverbs is of a specific non-linear form:

a. For German proverbs, this relation can adequately be modeled by the function $y = (C \cdot x) / (b + x)$, parameter a possibly being equaled with max_{FAM}. This function seems to be adequate for different kinds of determining frequency or usage, on the one hand, and knowledge or familiarity, on the other.

b. This function seems to be different from the one found far American proverbs, where $y = a \cdot x^b$ has turned out to be more adequate. Future studies will have to show, if language-specific effects are at work, if differences are motivated by different methods in establishing proverb popularity. It also seem reasonable to sub-summarize both models under a common, more general model, from which the two mentioned can be derived as special cases. Such a general model has been presented in this contribution – cf. (7) –, but it remains to be tested for both (and, wishfully, many other) languages in a follow-up study.

Notes

1. Both proverbs and familiarity data were basically taken from Grzybek's (1991) pilot study on German proverb knowledge, and additionally complemented by further empirical studies.

2. The very same holds true, by the way, for Čermák's (1998, 2003) analyses of the frequency-familiarity for Czech data. Relating familiarity to frequency data from the Czech National corpus, Čermák arrived at the conclusion that there is no significant correlation between both categories. Čermák, too, however, analyzed only highly familiar proverb material; thus no reliable information as to the FAM-FRQ relation is likely to be obtained.

3. A first version of this approach has been presented at the 2008 Europhras Conference in Helsinki (Grzybek 2008).

4. By the way, the result for model (III), too, is almost unchanged, if one sets a at the maximum of FAM at $a = 100$; in this case, with $b = 6.11$, the result is $R^2 = 0.82$. However, for this model, the differential equation for this model is $dy/y = (-a/x^2)\,dx$, which does not lend itself to interpretation as easily as does function (5).

5. Due to the variation of proverbs, particularly in journalistic texts, this kind of corpus-based (re)search is much more demanding than one might expect and asks for decisions as to inclusion or exclusion of varied, partly quoted, etc. items; for the present study, only full-text references were counted, allusions and modifications not being taken account of.

6. Again, the modest number of 10 items can be no solid basis for far-reaching conclusions; yet, this should suffice for the sake of methodological demonstration. As to the linear trend, the results obtained here have been meanwhile confirmed on a much larger data base. However, FTP results seem to be on a higher level as compared to PTP; for details see: Grzybek et al. (2009).

References

ČERMÁK, František, 1998. Usage of Proverbs: What the Czech National Corpus Shows. In Peter Ďurčo (ed.), Europhras '97, 37-59. Bratislava: Akadémia PZ.

ČERMÁK, František, 2003. Paremiological Minimum of Czech: The Corpus Evidence. In Harald BURGER, Annelies HÄCKI BUHOFER, Gertrud GRÉCIANO (eds.), Flut von Texten – Vielfalt der Kulturen, 15-32. Hohengehren: Schneider Verlag. [= Phraseologie und Parömiologie; 14]

ĎURČO, Peter, 2005. Sprichwörter in der Gegenwartssprache. Trnava: Univerzita sv. Cyrila a Metoda.

GRZYBEK, Peter, 1991. Sinkendes Kulturgut. Eine empirische Pilotstudie zur Bekanntheit deutscher Sprichwörter. Wirkendes Wort, 2: 239-264.

GRZYBEK, Peter, 2008. Proverb Frequency and Proverb Familiarity: Two Faces of One Coin or False Friends? Paper presented at the International Congress of the European Society for Phraseology, Helsinki 2008.

GRZYBEK, Peter; CHLOSTA, Christoph, 2009a. Some Essentials on the Popularity of (American) Proverbs. In Kevin J. McKENNA (ed.), The Proverbial "Pied Piper". A Festschrift Volume of Essays in Honor of Wolfgang Mieder on the Occasion of His Sixty-Fifth Birthday. p. ?-?. New York: Peter Lang.

GRZYBEK, Peter; CHLOSTA, Christoph; GROTJAHN, Rüdiger, 2009b. Empirische Sprichwortforschung: Methoden – Status – Perspektiven. Hohengehren: Schneider Verlag. [= Phraseologie und Parömiologie; in prep.]

HATTEMER, K.; SCHEUCH, E.K., 1983. Sprichwörter. Einstellung und Verwendung. Düsseldorf: Intermarket

========== /// ==========

A COMIDA E O COMER REPRESENTADOS NOS PROVÉRBIOS FINLANDESES

*Pekka **HAKAMIES**, Department of Folkloristics, School of Cultural Research, University of Turku, **Finland***

Conclusions

Due to the lack of previous research in this field, the present contribution cannot arrive at far-reaching contributions – it will be an important task for future research to provide and analyze further (and sufficient) data, which may serve as material to test the hypotheses brought forth here. Notwithstanding these lacunae, some general conclusions may be drawn:

1. The notion of proverb popularity should be more consequently differentiated in paremiology, clearly keeping apart *'usage-oriented'* vs. *'knowledge-oriented'* approaches. One line of research is rather *frequency-oriented* and primarily *text-based*, the other one is *familiarity-oriented*, *knowledge-based* and *subject-dependent*.

2. Familiarity and frequency do not denote the same; both concepts cannot simply be replaced one by another; they do not represent two faces of one and the same coin, rather two different coins joint together, resulting in a value calculated according to a specific exchange rate (simple correlation coefficient or linear models are not at stake here).

3. There is clear evidence for a specific relation between proverb frequency and proverb familiarity:

a. Familiar proverbs may occur frequently

b. Familiar proverbs need not occur frequently

c. Frequent proverbs tend to be familiar

4. For any study of the FAM-FRQ relation of proverbs, data must not only be sufficient, they must also represent the whole spectrum of familiarity, and must not concentrate on familiar proverbs, only.

5. The FAM-FRQ relation of proverbs is of a specific non-linear form:

a. For German proverbs, this relation can adequately be modeled by the function $y = (C \cdot x) / (b + x)$, parameter a possibly being equaled with max_{FAM}. This function seems to be adequate for different kinds of determining frequency or usage, on the one hand, and knowledge or familiarity, on the other.

b. This function seems to be different from the one found far American proverbs, where $y = a \cdot x^b$ has turned out to be more adequate. Future studies will have to show, if language-specific effects are at work, if differences are motivated by different methods in establishing proverb popularity. It also seem reasonable to sub-summarize both models under a common, more general model, from which the two mentioned can be derived as special cases. Such a general model has been presented in this contribution – cf. (7) –, but it remains to be tested for both (and, wishfully, many other) languages in a follow-up study.

Notes

1. Both proverbs and familiarity data were basically taken from Grzybek's (1991) pilot study on German proverb knowledge, and additionally complemented by further empirical studies.

2. The very same holds true, by the way, for Čermák's (1998, 2003) analyses of the frequency-familiarity for Czech data. Relating familiarity to frequency data from the Czech National corpus, Čermák arrived at the conclusion that there is no significant correlation between both categories. Čermák, too, however, analyzed only highly familiar proverb material; thus no reliable information as to the FAM-FRQ relation is likely to be obtained.

3. A first version of this approach has been presented at the 2008 Europhras Conference in Helsinki (Grzybek 2008).

4. By the way, the result for model (III), too, is almost unchanged, if one sets a at the maximum of FAM at $a = 100$; in this case, with $b = 6.11$, the result is $R^2 = 0.82$. However, for this model, the differential equation for this model is $dy/y = (-a/x^2)\,dx$, which does not lend itself to interpretation as easily as does function (5).

5. Due to the variation of proverbs, particularly in journalistic texts, this kind of corpus-based (re)search is much more demanding than one might expect and asks for decisions as to inclusion or exclusion of varied, partly quoted, etc. items; for the present study, only full-text references were counted, allusions and modifications not being taken account of.

6. Again, the modest number of 10 items can be no solid basis for far-reaching conclusions; yet, this should suffice for the sake of methodological demonstration. As to the linear trend, the results obtained here have been meanwhile confirmed on a much larger data base. However, FTP results seem to be on a higher level as compared to PTP; for details see: Grzybek et al. (2009).

References

ČERMÁK, František, 1998. Usage of Proverbs: What the Czech National Corpus Shows. In Peter Ďurčo (ed.), Europhras '97, 37-59. Bratislava: Akadémia PZ.

ČERMÁK, František, 2003. Paremiological Minimum of Czech: The Corpus Evidence. In Harald BURGER, Annelies HÄCKI BUHOFER, Gertrud GRÉCIANO (eds.), Flut von Texten – Vielfalt der Kulturen, 15-32. Hohengehren: Schneider Verlag. [= Phraseologie und Parömiologie; 14]

ĎURČO, Peter, 2005. Sprichwörter in der Gegenwartssprache. Trnava: Univerzita sv. Cyrila a Metoda.

GRZYBEK, Peter, 1991. Sinkendes Kulturgut. Eine empirische Pilotstudie zur Bekanntheit deutscher Sprichwörter. Wirkendes Wort, 2: 239-264.

GRZYBEK, Peter, 2008. Proverb Frequency and Proverb Familiarity: Two Faces of One Coin or False Friends? Paper presented at the International Congress of the European Society for Phraseology, Helsinki 2008.

GRZYBEK, Peter; CHLOSTA, Christoph, 2009a. Some Essentials on the Popularity of (American) Proverbs. In Kevin J. McKENNA (ed.), The Proverbial "Pied Piper". A Festschrift Volume of Essays in Honor of Wolfgang Mieder on the Occasion of His Sixty-Fifth Birthday. p. ?-?. New York: Peter Lang.

GRZYBEK, Peter; CHLOSTA, Christoph; GROTJAHN, Rüdiger, 2009b. Empirische Sprichwortforschung: Methoden – Status – Perspektiven. Hohengehren: Schneider Verlag. [= Phraseologie und Parömiologie; in prep.]

HATTEMER, K.; SCHEUCH, E.K., 1983. Sprichwörter. Einstellung und Verwendung. Düsseldorf: Intermarket