The Diversification of Meaning of Hungarian Verbal Prefixes
I. meg

1. The loss of independence of a morpheme results frequently in changes of its meaning or function. In most cases it becomes more abstract and according to circumstances — which have not been examined yet — it becomes either narrower or more diffuse, i.e. it modifies the class of independent morphemes to which it is attached, in different ways. The more words/stems there are with which it tends to be connected, the greater the chance of its meaning being diversified. But if these words occur rarely, diversification need not take place. It takes place necessarily if there are many words with which it is connected and if these words are used frequently. This process is in agreement with a hypothesis of Zipf according to which an increase of frequency of occurrence results in expansion of meaning (Zipf 1972).

The diversification is measurable and consequently it can be quantitatively evaluated. It is directly proportional to the entropy of meaning which can for the same reason be calculated, too. The aim of these articles is the examination of the diversification of the meaning of Hungarian verbal prefixes and the setting up some hypotheses about its course. The examination has a methodological character and we claim that it can be generalized to any kind of morphemes or words. Nevertheless, this approach must be tested on more voluminous data.

2. There are about 100 verbal prefixes in Hungarian. They are all derived from actual Hungarian adverbs. The most ancient verbal prefixes are *be-* "in", *el-* "away", *fel-* "up", *ki-* "out" and *meg-*. Meg is used nowadays to express the completion of an action or motion; its original locative meaning "back, re-" now occurs only in a few compound verbs.

Originally, the verbal prefixes in Hungarian express *direction; aspect* developed later. The third function is to *modify* or *change* the meaning of the verb. This function developed especially after the 15the century.

In most cases though, the lines between these function fields are not clear-cut: they often overlap each other. Another difficulty is presented by the fact that most verbal prefixes cannot form a compound with all verbs and, compounded with certain verbs, they can assume different meanings. When translating such compound verbs, one has to consult the dictionary, as well as consider the context, so as to obtain an appropriate answer concerning the meaning and aspect of the compound verb.

By analyzing the 141 compound verbs with the verbal prefix *meg-*, we observed 9 categories concerning verbal aspects and meaning modifications of the base verb. This classification will remain arbitrary. Although the lexicon and context help us to reach an appropriate definition, the decision will always contain a certain amount of subjectivity of the translator/examiner. For example the categories "Completed action" and "completed action + Meaning modification" are very close to each other. Neither is it easy to make a strict dictinction between the categories "Instantaneous action" and "Instantaneous action + Completed action".

In the following we will give a few examples of each category:

ért 'understand' (uncompleted action)
megért 'understand' (completed action)
ismer 'to know'
megismer 'to recognize' (completed action + modified meaning)
lenni 'to be'
meglenni 'to get along' (entirely new meaning)
figyel 'to pay attention' (uncompleted action)
megfigyel 'to observe' (completed action + new meaning)
áll 'to stay'
megáll 'to stop' (completed action + instanteneous action)
marad 'to stay, to remain, to fall, to leave'
megmarad 'to continue the action" (durative action)
talál 'to find, to hit, to happen'
megtalál 'to find' (completed action + its result)
csavar 'to twist, to turn'
megcsavar 'to turn (toward me)' (completed action + one time effect).

The examination of a relatively restricted corpus, namely the novel "A rózsakiállítás" by I. Örkény, has shown that the modifying function of *meg-* was diversified in the course of time. The new functions with their frequencies of occurrence are presented in table 1. They are ordered according to their frequency.

Seeing a frequency distribution like the given one we can ask at once whether it follows a stochastic law. It is easy to see that it does not follow the Zipf—Mandelbrot law used for rank-frequency distributions in the domain of the lexicon (cf. Orlov & Boroda & Nadarejšvili 1982), because the fall of frequency at the rank 2 is too abrupt. Nevertheless it is plausible to assume

Table 1

Functions of *meg-* and their frequencies

Rank	Meaning modifications of verbs by means of *meg-*	Frequency
1	Completed action	107
2	Completed action + meaning modification	8
3	Entirely new meaning	8
4	Completed action + new meaning	5
5	Completed action + its result	5
6	Instantaneous action	4
7	Completed action + instantaneous action	2
8	Completed action + one-time effect	1
9	Durative action	1

that one can set up a model of the distribution by means of a so called 'birth process'.

We proceed as follows. Let the probability that an entity has x meanings (functions) at time t be $P_x(t)$ $(t \geq 1)$. At time $t = 0$, where the examination starts, it has exactly 1 meaning,

$$\text{i.e.} \qquad P_1(0) = 1$$

(1)

$$\text{and} \qquad P_x(0) = 0 \quad \text{for } x = 2, 3, \ldots$$

Let us consider the meaning increase in a short interval $(t, t + dt)$. We assume that the following holds:

(1) The probability of the "birth" of one new meaning/function in this interval is proportional to the length of the interval, i.e. $f_x(t)\, dt$. The proportionality function will be determined below.

(2) The probability of emerging of two or more new meanings/functions is very small, $o(dt)$, and can be neglected.

(3) The probability of no new meaning emerging is then $1 - f_x(t)\, dt$.

(4) The events in distinct intervals are independent.

Under these conditions the probability that until time $t + dt$ the entity has exactly one meaning is given as

$$P_1(t + dt) = P_1(t)\,[1 - f(t)\, dt] \tag{2}$$

i.e. the product of the probability that until time t there is exactly one meaning and the probability that in dt no new meaning emerged. Analogically we have

$$P_x(t + dt) = P_x(t)\,[1 - f_x(t)\, dt] + P_{x-1}(t)\, f_{x-1}(t)\, dt \tag{3}$$

for $x = 2, 3, 4, \ldots$ The first expression on the right side is the probability that until time t, there were x meanings and in dt no new meaning emerged. The second expression is the probability that until time t, there were $x - 1$ meanings and in dt one new meaning emerged.

Subtracting $P_x(t)$ from the both sides of (2) and (3) and dividing them by dt we obtain

(4)

$$\frac{P_1(t + dt) - P_1(t)}{dt} = - P_1(t)\, f_1(t)$$

$$\frac{P_x(t + dt) - P_x(t)}{dt} = - P_x(t)\, f_x(t) + P_{x-1}(t)\, f_{x-1}(t)$$

Taking limits for $dt \to 0$ we obtain

(5)

$$\frac{dP_1(t)}{dt} = P_1'(t) = - P_1(t)\, f_1(t)$$

$$\frac{dP_x(t)}{dt} = P_x'(t) = - P_x(t)\, f_x(t) + P_{x-1}(t)\, f_{x-1}(t)\,.$$

In order to determine the proportionality function we assume that in the language community there is a struggle between the tendency of the speakers to amplify the meaning of an entity and the tendency of the hearers to reduce it (cf. Zipf 1972). We approximate it in the first run by means of a linear function $a + bx$ where a is the propensity to spontaneous changes and $b = d - c > 0$, d and c being constants representing the hearer's and the speaker's efforts respectively. Since the hearer controls the activity of the speaker we assumed $d > c$. Of course, there are other possibilities, too, which can be tried out in this general model. Using

$$f_x(t) = a + bx \tag{6}$$

we can write (5) as

$$\begin{aligned} P_1'(t) &= -P_1(t)\,(a + b) \\ P_x'(t) &= -P_x(t)\,(a + bx) + P_{x-1}(t)\,[a + b(x - 1)] \end{aligned} \tag{7}$$

This differential-difference equation can be solved e.g. stepwise using (1). We obtain

$$P_1(t) = e^{-(a+b)t}$$

$$P_2(t) = \frac{a + b}{b} e^{-(a+b)t} (1 - e^{-bt})$$

$$P(t)_3 = \frac{(a + b)(a + 2b)}{2b^2} e^{-(a+b)t} (1 - e^{-bt})^2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$P_x(t) = \frac{(a + b)(a + 2b) \ldots [a + (x - 1)\,b]}{(x - 1)!\, b^{x-1}} e^{-(a+b)t} (1 - e^{-bt})^{x-1}.$$

Substituting $a/b = k$ and $a = bk$ we obtain

$$P_x(t) = \frac{(k + 1)(k + 2) \ldots (k + x - 1)}{(x - 1)!} e^{-b(k+1)t} (1 - e^{-bt})^{x-1}$$

and, finally, writing

$$k + 1 = r;\; e^{-bt} = p;\; 1 - e^{-bt} = q$$

we obtain for any t

$$P_x = \frac{r(r + 1) \ldots (r + x - 2)}{(x - 1)!} p^r q^{x-1} = \binom{r + x - 2}{x - 1} p^r q^{x-1},\quad x = 1, 2, \ldots \tag{8}$$

which is the probability function of the displaced negative binomial distribution.

This result says that the most frequent meaning/function of the given entity e.g. in a given text occurs with frequency $NP_1 = Np^r$, where N is the number of all occurrences of the entity in that text. The meaning having the rank 2 has the frequency $NP_2 = Nrp^rq$ etc.

Let us test whether our hypothesis can be accepted in the case of *meg-*. We use the data from Table 1. Since the frequency of the most frequent meaning is substantial for the use of all other ones — the more conservatively it is used by the community, the smaller the chances for it to acquire an other meaning — we use it for the estimation of the parameters. As we have two parameters (p and r) we also use the first moment about the origin μ_1' that can be estimated by means of the average $\bar{x}$.

Because of $P_1 = p^r$ and $\mu_1' = 1 + rq/p$ we obtain

$$\frac{\bar{x} - 1}{\ln (f_1/N)} = \frac{1 - \hat{p}}{\hat{p} \ln \hat{p}} \tag{9}$$

From this formula, $\hat{p}$ can-be computed iteratively. Using the Newton procedure we can easily obtain $\hat{p}$ from

$$\hat{p}_{i+1} = \hat{p}_i - \frac{[\hat{q}_i \ln (f_1/N) - (\bar{x} - 1)\, \hat{p}_i \ln \hat{p}_i]\, \hat{p}_i \ln \hat{p}_i}{(\hat{p}_i - \ln \hat{p}_i - 1) \ln (f_1/N)} \tag{10}$$

Here f_1 is the (absolute) frequency of the most frequent meaning, i means the i-th iteration, $\bar{x}$ is the arithmetical mean. The second parameter follows from

$$\hat{r} = \frac{(\bar{x} - 1)\, \hat{p}}{\hat{q}}\,.$$

For our data we obtain

$$\hat{p} = 0.1736; \quad \hat{r} = 0.1602$$

which enables us to compute the theoretical frequencies (see Table 2, column NP_x)

Table 2

Observed and computed frequencies of individual meanings/functions of *meg-*

x	f_x	NP_x
1	107	107.00
2	8	13.92
3	8	6.66
4	5	3.96
5	5	2.59
6	4	1.78
7	2	1.26
8	1	0.92
9	1	2.92

Pooling the classes x = 8 and x = 9 we obtain a chisquare $X^2 = 9.39$ with 5 degrees of freedom corresponding to $P = 0.08$. Therefore we accept the null hypothesis that our model is in agreement with the data. The problems arising in connection with this new model will be discussed in another place.

The result shows that the diversification of the meaning of an entity is not arbitrary but follows a stochastic law. The boundary conditions relevant for the determination of the parameters are not yet known; they present an interesting research problem.

3. Let us now examine the problem of translating a verbal prefix that merely modifies the meaning of the verb, into another language not having a corresponding lexical device.

Two circumstances are to be considered here: (a) the translator is confronted with the diversity of *meg-* and obtains a more exact information only when hearing the adjoined verb (and the whole context); (b) even with this improvement he has no equivalent at his disposal in the target language and is forced to use heterogeneous means.

Several questions arise at once which are relevant both for the grammar and the semantics of Hungarian as well as for the contrastive linguistics and translation science:

(i) What is the uncertainty/information conveyed by a „meg-" + verb?

(ii) What is the entropy of the complete usage of *meg-*?

(iii) What is the uncertainty in translating *meg-* in general?

(iv) What is the association of a given meaning of *meg-* with the means of the target language?

(v) What is the uncertainty of drawing conclusions from the translation means to the meaning of *meg-*?

Table 3

Frequency of Dutch equivalents to the Hungarian verbal prefix *meg-*

	Dutch means															
Meaning	Paraphrase	No prefix	ver-	Other verb	werden	op-	be-	toe-	af-	her-	na-	terug-	ont-	over-	aan-	
1	26	31	14	6	7	5	5	3	2	2	1		1	1	1	105
2	2	2	1	1					1			1				8
3	7			1												8
4	3	1			1											5
5	1	1		2						1						5
6	1	1		1	1											4
7	2															2
8	1															1
9	1															1
	44	36	15	11	9	5	5	3	3	1	1	1	1	1	1	139

Several other questions can be asked. We shall treat only some of them here. We start from the frequency of translations of the particular meanings of *meg-* into Dutch in the work mentioned above. The data are presented. in table 2. In two cases the phrases containing *meg-* + verb were not translated at all so that the first row of the table has the marginal sum 105 (instead of 107) The numbers in the table show how many times the given meaning of *meg-* has been translated with the aid of the given Dutch lexical devices. As we can see the translations are relatively heterogeneous and consist not only of Dutch verbal prefixes but also of paraphrases, verbs without prefixes and verbs with special meanings.

Here we want to show that information theoretical measures can be reasonably used and interpreted in diversification and translation problems. We shall use here the following symbols:

X — variable: meanings of *meg-*

Y — variable: Dutch translation means

n_{ij} — numbers in table 2: frequency of x_i translated as y_j

n_i — marginal sums, i.e. $n_i = \sum_j n_{ij}$: frequency of the i-th meaning of *meg-*

n_j — marginal sums, i.e. $n_j = \sum_i n_{ij}$: frequency of the j-th translation mean

n — number of occurences of *meg-* (here 139)

$p_{ij} = n_{ij}/n$

$p_i = n_i/n$

$p_j = n_j/n$

The following measures will be used:

(i) The entropy of the source, i.e. the uncertainty adhering to the meaning of the Hungarian verbal prefix because of diversification

$$H(X) = -\sum_j p_i \,\mathrm{ld}\, p_i = \mathrm{ld}\, n - \frac{1}{n}\sum_i n_i \,\mathrm{ld}\, n_i . \quad (11)$$

This is, as a matter of fact, a measure of diversification.

(ii) The entropy of the receiver, i.e. the uncertainty with which the meaning of a Hungarian verbal prefix is interpreted

$$H(Y) = -\sum_j p_j \,\mathrm{ld}\, p_j = \mathrm{ld}\, n - \frac{1}{n}\sum_j n_j \,\mathrm{ld}\, n_j . \quad (12)$$

(iii) The entropy of the communication system as a whole, i.e. the joint uncertainty of the system

$$H(X, Y) = -\sum_i \sum_j p_{ij} \,\mathrm{ld}\, p_{ij} = \mathrm{ld}\, n - \frac{1}{n}\sum_i \sum_j n_{ij} \,\mathrm{ld}\, n_{ij} . \quad (13)$$

(iv) The conditional entropy of the receiver, i.e. the uncertainty of choosing a translation means when we know which meaning of the verbal prefix is involved

$$
\begin{aligned}
H(Y|X) &= -\sum_i \sum_j p_{ij} \operatorname{ld} (p_i/p_{ij}) \\
&= \frac{1}{n}\sum_i n_i \operatorname{ld} n_i - \frac{1}{n}\sum_i \sum_j n_{ij} \operatorname{ld} n_{ij} \qquad (14) \\
&= H(X, Y) - H(X).
\end{aligned}
$$

(v) The conditional entropy of the source, i.e. the uncertainty of identifying the meaning of the prefix if we know the translation

$$
\begin{aligned}
H(X|Y) &= -\sum_i \sum_j p_{ij} \operatorname{ld} (p_{ij}/p_j) \\
&= \frac{1}{n}\sum_i n_i \operatorname{ld} n_i - \frac{1}{n}\sum_i \sum_j n_{ij} \operatorname{ld} n_{ij} \qquad (15) \\
&= H(X, Y) - H(Y).
\end{aligned}
$$

(vi) Synentropy being that part of the information of the source that reaches the receiver. It is usually called 'transinformation'

$$
\begin{aligned}
I(X:Y) &= \sum_i \sum_j p_{ij} \operatorname{ld} \left(p_{ij}/(p_i p_j)\right) \\
&= \sum_i \sum_j p_{ij} \operatorname{ld} p_{ij} - \sum_i p_i \operatorname{ld} p_i - \sum_j p_j \operatorname{ld} p_j \qquad (16) \\
&= H(X) + H(Y) - H(X, Y) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X).
\end{aligned}
$$

It can be interpreted as the gain of information of the target language reader, if he knows the probability of the i-th meaning of the prefix in the source language (p_i) and the conditional probability of the i-th meaning being used given the particular means in the target language (p_{ij}/p_j).

The entropy of the source minus transinformation, i.e. $H(X) - I(X; Y) = H(X|Y)$ is that part of the information sent by the source that does not reach the receiver; therefore $H(X|Y)$ is called 'equivocation'. On the other hand $H(Y) - I(X; Y) = H(Y|X)$ is that part of the information received that does not come from the source but say from the context. It is called 'irrelevance'.

When translating a verb with the prefix *meg-*, the translator is confronted not only with *meg-* itself but also with the whole verb which reduces its uncertainty. Without the verb (and the context) the meaning of *meg-* is diffuse, having perhaps a relatively solid core but somewhat vague peripheries. In order to grasp this vagueness (diversification, diffusion) numerically we use formula (11)

$$H(X) = \text{ld } 139 - \frac{1}{139}(105 \text{ ld } 105 + 8 \text{ ld } 8 + 8 \text{ ld } 8 + 5 \text{ ld } 5 +$$
$$+ 5 \text{ ld } 5 + 4 \text{ ld } 4 + 2 \text{ ld } 2 + 1 \text{ ld } 1 + 1 \text{ ld } 1) = 1.4627 .$$

The computations can also be performed with decadic or natural logarithms.

The translator has two problems: to operate with his own devices so that the text will be stylistically correct, and at the same time to convey the meaning of the Hungarian compound verb of which *meg-* is a part. If the concentrates on *meg-* he has to choose his means with uncertainty H(Y); if he concentrates on the information not contained in this source but in the context, he has to consider the quantity $H(Y|X)$. Since the context specifies the meaning, it must hold that $H(Y) \geq H(Y|X)$ which can be shown mathematically, too (cf. Kämmerer 1971, Peters 1967, Reza 1961).

H(Y) can be computed from the marginal frequencies at the bottom ot table 2:

$$H(Y) = \text{ld } 139 - \frac{1}{139}(44 \text{ ld } 44 + 36 \text{ ld } 36 + \ldots + 1 \text{ ld } 1 + 1 \text{ ld } 1) =$$
$$= 2.8815.$$

In order to compute $H(Y|X)$ we first obtain $H(X, Y)$ as

$$H(X, Y) = \text{ld } 139 - \frac{1}{139}(26 \text{ ld } 26 + 31 \text{ ld } 31 + 14 \text{ ld } 14 + \ldots +$$
$$+ 2 \text{ ld } 2 + 1 \text{ ld } 1 + 1 \text{ ld } 1) = 4.0352$$

from wich according to (14)

$$H(Y|X) = 4.0352 - 1.4627 = 2.5725$$

which is smaller than H(Y).

Even using the context, part of the information reaching the translator cannot be translated, e.g. connotations, word plays etc. and gets lost. This loss can be expressed with

$$H(X|Y) = H(X, Y) - H(Y) = 4.0352 - 2.8815 = 1.1537 .$$

In general, the smaller this quantity the "better" the translation in information theoretical sense, and it holds that $H(X) \geq H(X|Y)$. If there is a one-to-one translation i.e. a total dependence of translation means on the particular meanings of *meg-* then $H(X|Y) = 0$; in case of total independence $H(X|Y) = H(X)$.

The measure of mutual information $I(X; Y)$ can be computed simply as

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = 1.4627 + 2.8815 - 4.0352 =$$
$$= 0.3090 .$$

It can be shown that this quantity measures the mutual dependence of the source meaning and translation means since

(17) $$2nI\,(X; Y) \ln_{e^2} = \chi^2_{(M-1)(N-1)}$$

i.e. the given transformation yields approximately a chi-square with (M—1) (N—1) degrees of freedom, where M is the number of translation means and N the number of meanings of the verbal prefix.

Analyses of this kind could be performed in any domain of language in order to measure the role of the context, the contrast or similarity of two languages etc. This had only been done qualitatively until now.

E. Beöthy—G. Altmann

References

Kämmerer, W. (1971), Einführung in mathematische Methoden der Kybernetik. Akademie-Verlag, Berlin
Orlov, Ju. K., Boroda, M. G., Nadarejšvili, I. Š. (1982), Sprache, Text, Kunst. Quantitative Analysen. Brockmeyer, Bochum
Örkény, I. (1977), Rózsakiállítás. Szépirodalmi, Budapest
Örkény, I. (1979), Rozententoonstelling. Meulenhoff, Amsterdam
Peters, J. (1967), Einführung in die allgemeine Informationstheorie. Springer, Berlin
Reza, F. M. (1961), An introduction to information theory. McGraw-Hill, New York
Zipf, G. K. (1972), Human Behavior and the Principle of Least Effort. Hafner, New York