QUANTITATIVE LINGUISTICS

Vol. 32

GLOTTOMETRIKA 8

edited

by

I. Fickermann

Studienverlag Dr. N. Brockmeyer
Bochum 1987

QUANTITATIVE LINGUISTICS

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Glottometrika. – Bochum: Studienverlag Brockmeyer
8. – 1987. – II, 211 S.
(Quantitative linguistics; Vol. 32)
ISBN 3-88339-559-5

ISBN 2-88339-559-5

Querenburger Höhe 281, 4630 Bochum 1
Druck: Thiebes GmbH & Co. Kommanditgesellschaft Hagen

We would like to express our gratitude
to the

STIFTUNG VOLKSWAGENWERK

a generous grant from which made possible
the edition of this book

CONTENTS

REVIEW

NEW PROJECTS

A THEORY OF DISTANCES BETWEEN LIKE ELEMENTS IN A SEQUENCE

P. Zörnig, Hagen

0. According to a linguistical hypothesis identical text elements occur in texts more often a short distance from one another then a long one (cf. ZÖRNIG, 1984I).

In order to prove this hypothesis we consider a text only in view of the arrangement of certain entities, i.e. we interprete a text as a sequence F of entities, chosen from a suitable set $\mathbb{F}$ of sequences (cf. sections 1 and 4). For every $F \in \mathbb{F}$ we consider the numbers d_i or c_i of i-distances between like elements, i = 0,..., n-2 (cf. section 1).

We want to test for a given text $F \in \mathbb{F}$ if the observed numbers $\hat{d}_i$ or $\hat{c}_i$ significantly diverge from the expected numbers $\bar{d}_i$ or $\bar{c}_i$, provided that F is chosen randomly from $\mathbb{F}$ (cf. section 4). To this end we need the means and variances of the random variables d_i and c_i. The means $E(d_i)$ and $E(c_i)$ were already computed in two previous works (ZÖRNIG, 1984, I, II). In the present article we compute the variances $V(d_i)$ and $V(c_i)$.

The theory presented here is a generalization of the theory of runs in so far as the latter one confines to a consideration of 0-distances.[1)]

However the restriction to 0-distances is not sufficient since the unbroken sequence of entities rarely occurs in a language, especially in the central disciplines of linguistics. The general case is the repetition of identical entities in certain distances.

I. DEFINITIONS

Let us briefly recall the definitions of the previous works of ZÖRNIG (1984). We have defined $\mathbb{F} := \mathbb{F}_{k_1, k_2, \ldots, k_p}$ as the set of all finite sequences consisting of n elements out of the set $\{1,\ldots,p\}$, were the element r occurs exactly k_r times (r=1,...,p

and $k_1,\ldots,k_p$ are natural numbers with $k_1+\ldots+k_p = n$). Let $F = (a_1,\ldots,a_n)$ be a sequence out of $\mathbb{F}$. For any μ,ν with $1 \le \mu \le \nu \le n$ the quantity $c(\mu,\nu) := \nu - \mu - 1$ is called the distance between a_μ and a_ν. Now for any $F \in \mathbb{F}$ the quantities $d_i(F)$ and $c_i(F)$ represent the frequency of occurrence of the distance i between arbitrary like elements in F or between adjacent like elements in F respectively. I.e. $d_i(F)$ is the number of all pairs of indices (μ,ν) with the properties

$$a_\mu = a_\nu$$

and

$$c(\mu,\nu) = i$$

and $c_i(F)$ is the number of all pairs of indices (μ,ν) with the properties

$$a_\mu = a_\nu$$
$$a_\sigma \neq a_\mu \text{ for all } \sigma \text{ with } \mu < \sigma < \nu$$

and $$c(\mu,\nu) = i$$

$(i = 0,\ldots,n-2;\ 1 \le \mu < \nu \le n;\ F = (a_1,\ldots,a_n) \in \mathbb{F})$. The sums D_i and C_i are defined by

$$D_i := \sum_{F\in\mathbb{F}} d_i(F) \tag{1}$$

and

$$C_i := \sum_{F\in\mathbb{F}} c_i(F). \tag{2}$$

In ZÖRNIG (1984 I and II) we derived

$$D_i = (n-i-1)\,\frac{(n-2)!}{k_1!\ldots k_p!}\sum_{r=1}^{p} k_r(k_r-1) \tag{1a}$$

$$C_i = \frac{(n-1-i)!}{k_1!\ldots k_p!}\sum_{r=1}^{p} k_r(k_r-1)(n-k_r)_{(i)}. \tag{2a}$$

We introduce the new definitions

$$D_i^{(2)} := \sum_{F\in\mathbb{F}} d_i^2(F) \tag{3}$$

and

$$C_i^{(2)} := \sum_{F\in\mathbb{F}} c_i^2(F) \tag{4}$$

(cf. table 1).

Finally we define for integers n,k:

$$n^+ := \begin{cases} n \text{ for } n > 0 \\ 0 \text{ else} \end{cases},$$

$$n! := 0 \text{ for } n < 0,$$

$$\binom{n}{k} := 0 \text{ for } k < 0 \text{ or } k > n,$$

$$n_{(k)} := k!\binom{n}{k}.$$

2. DERIVATION OF FORMULAE FOR $D_i^{(2)}$ AND $C_i^{(2)}$

Obviously the problem of the evaluation of the variances can be reduced to the evaluation of the numbers $D_i^{(2)}$ and $C_i^{(2)}$ respectively, $i = 0,\ldots,n-2$ (cf. section 3).

In order to do this we must first prove two Lemmata.

Lemma 1: Let μ,ν,i be natural numbers with $0 \le i \le n-2$; $1 \le \mu < \nu \le n-1-i$. Let $N(\mu,\nu,i)$ be the number of all sequences $F = (a_1,\ldots,a_n) \in \mathbb{F}$ with the properties

$$a_\mu = a_{\mu+i+1},\quad a_\nu = a_{\nu+i+1}. \tag{5}$$

Then it holds:

$$N(\mu,\nu,i) = \begin{cases} N_1 := \frac{(n-3)!}{k_1!\ldots k_p!} \sum_{r=1}^{p} k_r(k_r-1)(k_r-2) & \text{for } \nu=\mu+i+1 \\ N_2 := \frac{(n-4)!}{k_1!\ldots k_p!} \left[\left(\sum_{r=1}^{p} k_r(k_r-1) \right)^2 \right. & \\ \left. + 2 \sum_{r=1}^{p} k_r(k_r-1)(3-2k_r) \right] & \text{else} \end{cases}$$

Proof: Case 1: $\nu = \mu + i + 1$.

The sequences $F \in \mathbb{F}$ with the properties (5) are of the form

$$F = \ldots r \ldots\ldots\ldots r \ldots\ldots\ldots r \ldots \quad \text{(r arbitrary)}$$

Positions: μ, $\mu+i+1 = \nu$, $\nu+i+1$

If r is fixed there are evidently

$$\binom{n-3}{k_1,\ldots,k_{r-1},k_r-3,k_{r+1},\ldots,k_p} \text{ sequences of this form,}$$

i.e.

$$N(\mu,\nu,i) = \sum_{r=1}^{p} \binom{n-3}{k_1,\ldots,k_{r-1},k_r-3,k_{r+1},\ldots,k_p} =$$

$$= \frac{(n-3)!}{k_1!\ldots k_p!} \sum_{r=1}^{p} k_r(k_r-1)(k_r-2).$$

Case 2: $\nu \neq \mu + i + 1$.

The sequences $F \in \mathbb{F}$ with the properties (5) are of the form

$$F = \ldots q \ldots\ldots\ldots r \ldots q \ldots\ldots\ldots r \ldots \quad \text{(r,q arbitrary)}$$

Positions: μ, ν, $\mu+i+1$, $\nu+i+1$

If r and q are fixed, there exist

$$\binom{n-4}{k_1,\ldots,k_{r-1},k_r-4,k_{r+1},\ldots,k_p} \text{ or}$$

$$\binom{n-4}{k_1,\ldots,k_{q-1},k_q-2,k_{q+1},\ldots,k_{r-1},k_r-2,k_{r+1},\ldots,k_p}$$

sequences of this form if r=q or r≠q respectively; i.e.

$$N(\mu,\nu,i) = \sum_{r=1}^{p} \binom{n-4}{k_1,\ldots,k_{r-1},k_r-4,k_{r+1},\ldots,k_p} +$$

$$+ \sum_{\substack{r,q=1 \\ r\neq q}}^{p} \binom{n-4}{k_1,\ldots,k_{q-1},k_q-2,k_{q+1},\ldots,k_{r-1},k_r-2,k_{r+1},\ldots k_p}$$

$$= \frac{(n-4)!}{k_1!\ldots k_p!} \sum_{r=1}^{p} k_r(k_r-1)(k_r-2)(k_r-3)$$

$$+ \frac{(n-4)!}{k_1!\ldots k_p!} \sum_{\substack{r,q=1 \\ r\neq q}}^{p} k_q(k_q-1)k_r(k_r-1)$$

$$= \frac{(n-4)!}{k_1!\ldots k_p!} \sum_{r=1}^{p} k_r(k_r-1)(k_r-2)(k_r-3)$$

$$+ \frac{(n-4)!}{k_1!\ldots k_p!} \left[\left(\sum_{r=1}^{p} k_r(k_r-1) \right)^2 - \sum_{q=1}^{p} k_q^2(k_q-1)^2 \right]$$

This proves the Lemma for case 2.

Lemma 2: Let μ,ν,i be natural numbers with $0 \leq i \leq n-2$; $1 \leq \mu < \nu \leq n-1-i$. Let $N'(\mu,\nu,i)$ be the number of all sequences $F=(a_1,\ldots,a_n) \in \mathbb{F}$ with the properties:

$$
\begin{aligned}
a_\mu &= a_{\mu+i+1} \\
a_\sigma &\neq a_\mu \qquad \text{for } \mu<\sigma<\mu+i+1 \\
a_\nu &= a_{\nu+i+1} \\
a_\sigma &\neq a_\nu \qquad \text{for } \nu<\sigma<\nu+i+1
\end{aligned} \tag{6}
$$

Then it holds

$$
N'(\mu,\nu,i) = \begin{cases} N_1' & \text{for } \nu=\mu+i+1 \\ N_2' & \text{for } \nu>\mu+i+1 \\ N_3'(\mu,\nu) & \text{for } \nu<\mu+i+1 \end{cases}
$$

where N_1', N_2', N_3' are defined as follows:

$$
N_1' := \frac{(n-2i-3)!}{k_1!\ldots k_p!} \sum_{r=1}^{p} (n-k_r)\,(2i)^{k_r}(k_r-1)\,(k_r-2)
$$

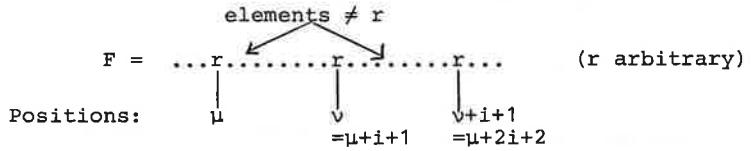
$$
N_2' := \frac{1}{k_1!\ldots k_p!}\, 2 \sum_{\substack{q,r=1 \\ q<r}}^{p} (n-k_r-k_q)!\,k_r!\,k_q! \sum_{x\geq 0} \binom{n-2i-4}{x} \binom{i}{k_q-2-x} \binom{n-4-i-x}{k_r-2}
$$

$$
+ \frac{(n-2i-4)!}{k_1!\ldots k_p!} \sum_{r=1}^{p} (n-k_r)\,(2i)^{k_r}(k_r-1)\,(k_r-2)\,(k_r-3) \qquad 2)
$$

$$
N_3'(\mu,\nu) = \frac{1}{k_1!\ldots k_p!}\, 2 \sum_{\substack{q,r=1 \\ q<r}}^{p} (n-k_r-k_q)!\,k_r!\,k_q! \sum_{x\geq 0} \binom{n-\nu+\mu-i-2}{x} \binom{\nu-\mu-1}{k_q-2-x} \binom{n-i-3-x}{k_r-2}
$$

Proof: Case 1: $\nu = \mu + i + 1$.

The sequences $F \in \mathbb{F}$ with the properties (6) are of the form



If r is fixed there are

$$
\binom{n-2i-3}{k_r-3} \binom{n-k_r}{k_1,\ldots,k_{r-1},k_{r+1},\ldots,k_p}
$$

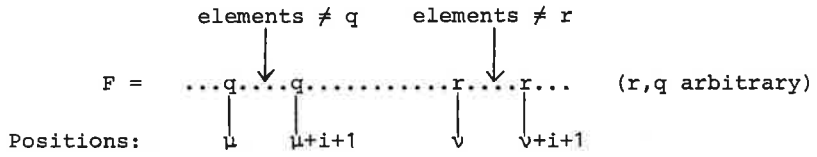
sequences of this form; k_r-3 free elements of kind r can be placed on the $(n-2i-3)$ positions $1,\ldots, \mu-1, \mu+2i+3,\ldots,n$ and $n-k_r$ elements of kinds $\neq r$ can be placed on the remaining positions. Now we have

$$
N'(\mu,\nu,i) = \sum_{r=1}^{p} \binom{n-2i-3}{k_r-3} \binom{n-k_r}{k_1,\ldots k_{r-1},k_{r+1},\ldots,k_p}
$$

$$
= \frac{(n-2i-3)!}{k_1!\ldots k_p!} \sum_{r=1}^{p} (n-k_r)\,(2i)^{k_r}(k_r-1)\,(k_r-2).
$$

Case 2: $\nu > \mu + i + 1$.

The sequences $F \in \mathbb{F}$ with the properties (6) are of the form



If r and q are fixed, there exists

$$
\binom{n-2i-4}{k_r-4} \binom{n-k_r}{k_1,\ldots,k_{r-1},k_{r+1},\ldots,k_p} \tag{7}
$$

or

$$\sum_{x\geq 0}\binom{n-2i-4}{x}\binom{i}{k_q-2-x}\binom{n-4-i-x}{k_r-2}\binom{n-k_r-k_q}{k_1,\ldots,k_{q-1},k_{q+1},\ldots,k_{r-1},k_{r+1},\ldots,k_p} \qquad (8)$$

sequences of this form if r=q or r≠q respectively.

The expression (7) for r=q can be derived quite analogically as in case 1.

We obtain expression (8) for r≠q as follows: Assume that x elements of kind q are to place outside of the intervals $[\mu,\mu+i+1]$, $[\nu,\nu+i+1]$ of positions: Now there are $\binom{n-2i-4}{x}$ ways to place these x elements of kind q on places outside of $[\mu,\mu+i+1]$ and $[\nu,\nu+i+1]$, and there are $\binom{i}{k_q-2-x}$ ways to place the remaining elements of kind q between the positions ν and ν+i+1. Then k_r-2 free elements of kind r can be placed on that positions outside of $[\nu,\nu+i+1]$ that are not already occupied by elements of type q, i.e. there are $\binom{n-4-i-x}{k_r-2}$ ways to place the elements of kind r.

Finally there are

$$\binom{n-k_r-k_q}{k_1,\ldots,k_{q-1},k_{q+1},\ldots,k_{r-1},k_{r+1},\ldots,k_p}$$

ways to place the elements of types ≠ q,r on the remaining free positions. Now (8) is evidently the number of sequences of the above form for fixed q and r with q ≠ r.

Altogether yields

$$N'(\mu,\nu,i) = \sum_{\substack{q,r=1\\ q\neq r}}^{p}\sum_{x\geq 0}\binom{n-2i-4}{x}\binom{i}{k_q-2-x}\binom{n-4-i-x}{k_r-2}$$

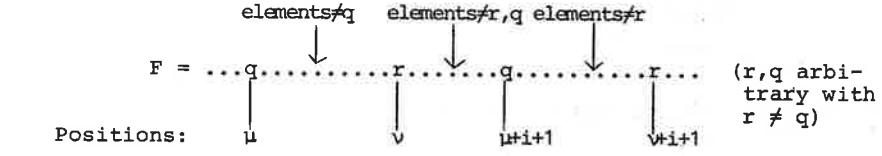
$$\cdot\binom{n-k_r-k_q}{k_1,\ldots,k_{q-1},k_{q+1},\ldots,k_{r-1},k_{r+1},\ldots,k_p}$$

$$+\sum_{r=1}^{p}\binom{n-2i-4}{k_r-4}\binom{n-k_r}{k_1,\ldots,k_{r-1},k_{r+1},\ldots,k_p}.$$

Note that the sum over x does not change if we exchange k_q for k_r. This proves the Lemma for case 2.

Case 3: ν < μ + i + 1.

The sequences F ∈ 𝔽 with the properties (6) are now of the form



Similarly to case 2 it can be shown that now exist

$$\sum_{x\geq 0}\binom{n-\nu+\mu-i-2}{x}\binom{\nu-\mu-1}{k_q-2-x}\binom{n-i-3-x}{k_r-2}\binom{n-k_r-k_q}{k_1,\ldots,k_{q-1},k_{q+1},\ldots,k_{r-1},k_{r+1},\ldots,k_p}$$

sequences of this form if r and q are fixed, r ≠ q (cf. (8)).

This proves the Lemma for case 3. (Note that the number of sequences now depends also on μ and ν).

In the special case when p = 2 we obtain from Lemma 2:

$$N_1' = \frac{(n-2i-3)!}{k_1!k_2!}\left((k_2)_{(2i)}k_1(k_1-1)(k_1-2)+(k_1)_{(2i)}k_2(k_2-1)(k_2-2)\right)$$

$$N_2' = 2\sum_{x\geq 0}\binom{n-2i-4}{x}\binom{i}{k_1-2-x}\binom{n-4-i-x}{k_2-2}$$

$$+\frac{(n-2i-4)!}{k_1!k_2!}\left((k_2)_{(2i)}k_1\ldots(k_1-3)+(k_1)_{(2i)}k_2\ldots(k_2-3)\right)$$

$$N_3'(\mu,\nu) = 2\sum_{x\geq 0}\binom{n-\nu+\mu-i-2}{x}\binom{\nu-\mu-1}{k_1-2-x}\binom{n-i-3-x}{k_2-2}$$

Example 1: For $k_1 = 2$, $k_2 = 3$; p = 2, n = 5 we obtain from this:

$$N_1' = \begin{cases}1 \text{ for } i = 0;1\\ 0 \text{ for } i = 2;3\end{cases}$$

$$N_2' = \begin{cases} 2 & \text{for } i = 0 \\ 0 & \text{for } i = 1,2,3 \end{cases}$$

$$\sum_{\substack{\mu,\nu=1 \\ \mu<\nu<\mu+i+1}}^{n-i-1} N_3'(\mu,\nu) = \begin{cases} 0 & \text{for } i = 0 \\ 4 & \text{for } i = 1 \\ 0 & \text{for } i = 2 \\ 0 & \text{for } i = 3 \end{cases}$$

Theorem 1: It holds for $0 \le i \le n-2$:

$$D_i^{(2)} = D_i + 2(n-2i-2)^+(N_1-N_2) + (n-i-1)(n-i-2)N_2$$

where N_1 and N_2 are defined as in Lemma 1 (cf. (1a)).

Proof: a) For $1 \le \mu < \nu \le n$ and $F = (a_1,\ldots,a_n) \in \mathbb{F}$ we define the auxiliary quantity

$$\alpha_{\mu,\nu}(F) := \begin{cases} 1 & \text{for } a_\mu = a_\nu \\ 0 & \text{else} \end{cases} \qquad (9)$$

Now we evidently have

$$d_i(F) = \sum_{\mu=1}^{n-i-1} \alpha_{\mu,\mu+i+1}(F), \quad (i=0,\ldots,n-2). \qquad (10)$$

Inserting (10) into definition (3) yields

$$D_i^{(2)} = \sum_{F\in\mathbb{F}} \Big(\sum_{\mu=1}^{n-i-1} \alpha_{\mu,\mu+i+1}(F) \Big)^2 \qquad (11)$$

$$= \sum_{F\in\mathbb{F}} \sum_{\mu=1}^{n-i-1} \alpha^2_{\mu,\mu+i+1}(F)$$

$$+ 2 \sum_{\substack{\mu,\nu=1 \\ \mu<\nu}}^{n-i-1} \sum_{F\in\mathbb{F}} \alpha_{\mu,\mu+i+1}(F)\cdot\alpha_{\nu,\nu+i+1}(F)$$

Since $\alpha_{\mu,\mu+i+1}(F)$ only takes the values 0 and 1 it follows from (1) and (10) that the first summand in (11) equals D_i: Evidently it is $\alpha_{\mu,\mu+i+1}(F)\cdot\alpha_{\nu,\nu+i+1}(F) = 1$ if the sequence F has the properties (5) of Lemma 1, else it is $\alpha_{\mu,\mu+i+1}(F)\cdot\alpha_{\nu,\nu+i+1}(F) = 0$. Hence the sum

$$\sum_{F\in\mathbb{F}} \alpha_{\mu,\mu+i+1}(F)\cdot\alpha_{\nu,\nu+i+1}(F)$$

in (11) is the number of sequences $F \in \mathbb{F}$ with the properties (5), i.e.

$$\sum_{F\in\mathbb{F}} \alpha_{\mu,\mu+i+1}(F)\cdot\alpha_{\nu,\nu+i+1}(F) = N(\mu,\nu,i) \qquad (12)$$

(cf. Lemma 1). Inserting (12) into (11) yields

$$D_i^{(2)} = D_i + 2 \sum_{\substack{\mu,\nu=1 \\ \mu<\nu}}^{n-i-1} N(\mu,\nu,i). \qquad (13)$$

From this follows by Lemma 1:

$$D_i^{(2)} = D_i + 2\Big((n-2i-2)^+ N_1 + \Big(\binom{n-i-1}{2} - (n-2i-2)^+\Big)N_2 \Big), \qquad (14)$$

because there are $(n-2i-2)^+$ pairs of indices (μ,ν) with $1\le\mu<\nu\le n-i-1$, $\nu=\mu+i+1$ and there are $\left(\binom{n-i-1}{2} - (n-2i-2)^+\right)$ pairs of indices (μ,ν) with $1\le\mu<\nu\le n-i-1$, $\mu\ne\nu+i+1$. The Theorem follows immediately from (14).

From Theorem 1 follows that $D_i^{(2)}$ is a quadratic function of i for fixed $p;k_1,\ldots,k_p$, because the numbers N_1 and N_2 do not depend on i (cf. the numbers N_1', N_2', N_3' used in Theorem 2).

Example 2: Let be $k_1 = 2$, $k_2 = 3$; $p = 2$, $n = 5$. We obtain

$$D_i = 4(4-i),\ N_1 = 1,\ N_2 = 2$$

from (1a) and Lemma 1. From this follows by Theorem 1:

$$D_i^{(2)} = 4\cdot(4-i) + 2\cdot(5-2i-2)^+\cdot(1-2) + (5-i-1)\cdot(5-i-2)\cdot 2$$

$$= \begin{cases} 2i^2 - 14i + 34 & \text{for } i = 0;1 \\ 2i^2 - 18i + 40 & \text{for } i = 2;3, \text{ i.e.} \end{cases}$$

$$D_i^{(2)} = \begin{cases} 34 & \text{for } i = 0 \\ 22 & \text{for } i = 1 \\ 12 & \text{for } i = 2 \\ 4 & \text{for } i = 3 \end{cases} \quad \text{(cf. Table 1).}$$

Theorem 2: It holds for $0 \le i \le n-2$:

$$C_i^{(2)} = C_i + 2\Big((n-2i-2)^+N_1' + \binom{n-2i-2}{2}N_2' + \sum_{\substack{\mu,\nu=1 \\ \mu<\nu<\mu+i+1}}^{n-i-1} N_3'(\mu,\nu)\Big),$$

where N_1', N_2', $N_3'(\mu,\nu)$ are defined as in Lemma 2 (cf.(2a)).

Proof: We define the auxiliary quantity

$$\beta_{\mu,\nu}(F) := \begin{cases} 1 & \text{for } a_\mu = a_\nu \text{ and } a_\sigma \neq a_\mu \text{ for } \mu<\sigma<\nu \\ 0 & \text{else} \end{cases}$$

for $1\le\mu<\nu\le n$, $F = (a_1,\ldots,a_n) \in \mathbb{F}$ (cf.(9)). Now it is

$$c_i(F) = \sum_{\mu=1}^{n-i-1} \beta_{\mu,\mu+i+1}(F), \quad (i=0,\ldots,n-2).$$

From this follows quite analogically to the proof of Theorem 1:

$$C_i^{(2)} = \sum_{F\in\mathbb{F}} \sum_{\mu=1}^{n-i-1} \beta^2_{\mu,\mu+i+1}(F) + 2 \sum_{\substack{\mu,\nu=1 \\ \mu<\nu}}^{n-i-1} \sum_{F\in\mathbb{F}} \beta_{\mu,\mu+i+1}(F)\cdot\beta_{\nu,\nu+i+1}(F) \quad (15)$$

(cf. equation (11)).

Now the first summand in (15) equals C_i and

$$\sum_{F\in\mathbb{F}} \beta_{\mu,\mu+i+1}(F)\cdot\beta_{\nu,\nu+i+1}(F) = N'(\mu,\nu,i)$$

(cf. Lemma 2). From this follows

$$C_i^{(2)} = C_i + 2\sum_{\substack{\mu,\nu=1 \\ \mu<\nu}}^{n-i-1} N'(\mu,\nu,i), \quad \text{(cf. (13)).} \qquad (16)$$

From this and Lemma 2 follows

$$C_i^{(2)} = C_i + 2\Big((n-2i-2)^+N_1' + \binom{n-2i-2}{2}N_2' + \sum_{\substack{\mu,\nu=1 \\ \mu<\nu<\mu+i+1}}^{n-i-1} N_3'(\mu,\nu)\Big) \qquad (17)$$

because there are $(n-2i-2)^+$ pairs of indices (μ,ν) with $1\le\mu<\nu\le n-i-1$, $\nu=\mu+i+1$ and there are $\binom{n-2i-2}{2}$ pairs of indices (μ,ν) with $1\le\mu<\nu\le n-i-1$, $\nu>\mu+i+1$.

Example 3: Let be $k_1 = 2$, $k_2 = 3$; $p = 2$, $n = 5$. From (2a) we obtain

$$C_i = (4-i)^2.$$

In example 1 we have just shown:

$$N_1' = \begin{cases} 1 & \text{for } i = 0;1 \\ 0 & \text{for } i = 2;3 \end{cases}$$

$$N_2' = \begin{cases} 2 & \text{for } i = 0 \\ 0 & \text{for } i = 1,2,3 \end{cases}$$

$$\sum_{\substack{\mu,\nu=1 \\ \mu<\nu<\mu+i+1}}^{n-i-1} N_3'(\mu,\nu) = \begin{cases} 0 & \text{for } i = 0 \\ 4 & \text{for } i = 1 \\ 0 & \text{for } i = 2 \\ 0 & \text{for } i = 3. \end{cases}$$

From this follows by Theorem 2:

$$C_i^{(2)} = (4-i)^2 + 2\Big((3-2i)^+ N_1' + \binom{3-2i}{2} N_2' + \sum_{\substack{\mu,\nu=1 \\ \mu<\nu<\mu+i+1}}^{n-i-1} N_3'(\mu,\nu) \Big)$$

$$C_i^{(2)} = \begin{cases} 16 + 2(3\cdot 1 + \binom{3}{2}\cdot 2 + 0) & \text{for } i = 0 \\ 9 + 2(1\cdot 1 + 0\cdot 0 + 4) & \text{for } i = 1 \\ 4 + 2(0\cdot 0 + 0\cdot 0 + 0) & \text{for } i = 2 \\ 1 + 2(0\cdot 0 + 0\cdot 0 + 0) & \text{for } i = 3, \end{cases}$$

i.e.

$$C_i^{(2)} = \begin{cases} 34 & \text{for } i = 0 \\ 19 & \text{for } i = 1 \\ 4 & \text{for } i = 2 \\ 1 & \text{for } i = 3 \end{cases} \qquad \text{(cf. Table 1).}$$

Tab. 1: List of the numbers $d_i(F)$ and $c_i(F)$ for $k_1=2$, $k_2=3$; $p=2$, $n=5$

$\mathbb{F}$	d_0	d_1	d_2	d_3	d_0^2	d_1^2	d_2^2	d_3^2	c_0	c_1	c_2	c_3	c_0^2	c_1^2	c_2^2	c_3^2
11222	3	1			9	1			3				9			
12122	1	2	1		1	4	1		1	2			1	4		
12212	1	1	2		1	1	4		1	1	1		1	1	1	
12221	2	1		1	4	1		1	2			1	4			1
21122	2		1	1	4		1	1	2		1		4		1	
21212		3		1		9		1		3				9		
21221	1	1	2		1	1	4		1	1	1		1	1	1	
22112	2		1	1	4		1	1	2		1		4		1	
22121	1	2	1		1	4	1		1	2			1	4		
22211	3	1			9	1			3				9			
sums	16	12	8	4	34	22	12	4	16	9	4	1	34	19	4	1

The sums D_i, $D_i^{(2)}$, C_i, $C_i^{(2)}$ are listed in the last row under the corresponding columns (cf. section 1).

3. DERIVATION OF THE VARIANCES

Let $F \in \mathbb{F}$ be a randomly chosen sequence from $\mathbb{F}$. Let d_i and c_i be the number of i-distances between like elements (or between adjacent like elements respectively) in the chosen sequence F (cf. section 1). In this section we are able to derive formulae for the variances $V(c_i)$ and $V(d_i)$ of the considered random variables c_i and d_i. The results are summarized in

Theorem 3: It holds for $i = 0,\ldots,n-2$:

$$\text{(a) } V(d_i) = \frac{1}{\binom{n}{k_1,\ldots,k_p}} \left(D_i + 2(n-2i-2)^+ (N_1 - N_2) + (n-i-1)(n-i-2)N_2\right) - \left(\frac{D_i}{\binom{n}{k_1,\ldots,k_p}}\right)^2$$

$$\text{(b) } V(c_i) = \frac{1}{\binom{n}{k_1,\ldots,k_p}} \left[C_i + 2\left((n-2i-2)^+ N_1' + \binom{n-2i-2}{2} N_2' + \sum_{\substack{\mu,\nu=1 \\ \mu<\nu<\mu+i+1}}^{n-i-1} N_3'(\mu,\nu) \right)\right] - \left(\frac{C_i}{\binom{n}{k_1,\ldots,k_p}}\right)^2$$

where N_1, N_2, N_1', N_2', $N_3'(\mu,\nu)$ are defined as in Lemma 1 or Lemma 2 respectively (cf. (1a), (2a) for the computation of D_i and C_i).

Proof: (a) Since $V(d_i) = E(d_i^2) - E^2(d_i)$, we obtain

$$V(d_i) = \frac{1}{\binom{n}{k_1,\ldots,k_p}} \sum_{F\in\mathbb{F}} d_i^2(F) - \left(\frac{1}{\binom{n}{k_1,\ldots,k_p}} \sum_{F\in\mathbb{F}} d_i(F)\right)^2$$

i.e.

$$V(d_i) = \frac{1}{\binom{n}{k_1,\ldots,k_p}} D_i^{(2)} - \left(\frac{1}{\binom{n}{k_1,\ldots,k_p}} D_i\right)^2 \qquad (18)$$

From this follows the proof by Theorem 1.

(b) The proof is analogous by using Theorem 2.

Example 4: Let be again $k_1 = 2$, $k_2 = 3$; $p = 2$, $n = 5$. From Theorem 3(a) follows (cf. (18)):

$$V(d_i) = \frac{D_i^{(2)}}{\binom{5}{2,3}} - \left(\frac{D_i}{\binom{5}{2,3}}\right)^2, \quad (i=0,1,2,3). \qquad (19)$$

In example 2 we have just found for this special case:

$$D_i = 4\cdot(4-i)$$

$$D_i^{(2)} = \begin{cases} 2i^2-14i+34 & \text{for } i = 0;1 \\ 2i^2-18i+40 & \text{for } i = 2;3. \end{cases}$$

Inserting in (19) yields:

$$V(d_i) = \begin{cases} 0.04i^2-0.12i+0.84 & \text{for } i = 0;1 \\ 0.04i^2-0.52i+1.44 & \text{for } i = 2;3, \end{cases}$$

i.e.

$$V(d_i) = \begin{cases} 0.84 & \text{for } i = 0 \\ 0.76 & \text{for } i = 1 \\ 0.56 & \text{for } i = 2 \\ 0.24 & \text{for } i = 3. \end{cases}$$

4. AN EXAMPLE OF APPLICATION

We use again the example in ZÖRNIG (1984 I and II) for an application of the variance $V(d_i)$ derived in section 3. Considering the verses 81 to 130 in Vergils Aeneis we obtain the sequence

$$\begin{array}{l} 32122\ 10232\ 11221\ 31211\ 32132 \\ 04122\ 34212\ 13123\ 03212\ 22322 \end{array} \qquad (20)$$

where the i-th element of the sequence (20) is the number of dactyls in the verse (80+i), (i = 1,...,50). The sequence (20) is an element of $\mathbb{F}_{k_1,\ldots,k_p}$, where $k_1 = 3$, $k_2 = 14$, $k_3 = 21$, $k_4 = 10$, $k_5 = 2$; $p = 5$, $n = 50$.

In this section we test whether the observed numbers $\hat{d}_i$ of i-distances in (20) significantly diverge from the theoretically expected numbers $\bar{d}_i$ of i-distances in a randomly chosen sequence F from $\mathbb{F}_{3,14,21,10,2}$ (cf. the last sections in ZÖRNIG (1984 I and II).

To this end we compute the probabilities $P_i := P(|d_i-\bar{d}_i| \geq |\hat{d}_i-\bar{d}_i|)$ that the random deviation $|d_i-\bar{d}_i|$ is greater (or equal) than the observed deviation $|\hat{d}_i-\bar{d}_i|$ (i = 0,...,n-2).

We define

$$z_i := \frac{d_i-\bar{d}_i}{\sqrt{V(d_i)}} \qquad i = 0,\ldots,n-2$$

as the standardized variable ($\hat{z}_i$ is defined analogically). Assuming that for $n \to \infty$ the limiting distribution of z_i is normal we obtain the approximation

$$\begin{aligned} P_i &= P(|d_i-\bar{d}_i| \geq |\hat{d}_i-\bar{d}_i|) \\ &= P(|z_i| \geq |\hat{z}_i|) \\ &\approx 2\,\Phi(-|\hat{z}_i|), \end{aligned} \qquad (21)$$

where Φ is the standardized normal distribution function.

The numbers $\bar{d}_i$, $\hat{d}_i$, $V(d_i)$, $\hat{z}_i$ and P_i are listed in Table 2. For the computation of $\bar{d}_i$ we use the formula $\bar{d}_i = \frac{2}{7}(49-i)$ of

ZÖRNIG (1984 I, section 3). In order to compute $V(d_i)$ we proceed as follows: from Lemma 1 and (1a) we obtain

$$\underline{N}_1 := \frac{N_1}{\binom{n}{k_1,\ldots,k_p}} = 0.0926020408$$

$$\underline{N}_2 := \frac{N_2}{\binom{n}{k_1,\ldots,k_p}} = 0.0805181647$$

$$\underline{D}_i := \frac{D_i}{\binom{n}{k_1,\ldots,k_p}} = \frac{2}{7}(49-i) \ .$$

From this follows by Theorem 3(a):

$$V(d_i) = \underline{D}_i + 2(n-2i-2)^+(\underline{N}_1-\underline{N}_2) + (n-i-1)(n-i-2)\underline{N}_2 - \underline{D}_i^2$$

$$= \frac{2}{7}(49-i) + (49-i)(48-i)\underline{N}_2 - \left(\frac{2}{7}\right)^2(49-i)^2$$

$$+ (48-2i)^+2(\underline{N}_1-\underline{N}_2)$$

$$= \begin{cases} -0.00111144883i^2 - 0.1443117673i + 8.53877551 & \text{for } i\leq 23 \\ -0.00111144883i^2 - 0.0959762628i + 7.378723404 & \text{for } i\geq 24 \end{cases}$$

$$(i = 0,\ldots,48).$$

Choosing $\alpha := 0.05$ as level of significance we can learn from Table 2 that the observed values $\hat{d}_0$, $\hat{d}_7$ and $\hat{d}_{40}$ significantly diverge from the expected values $\bar{d}_0$, $\bar{d}_7$ and $\bar{d}_{40}$, i.e. P_0 P_7, $P_{40} \leq$ ≤ 0.05. Especially $\hat{d}_0$ is smaller than expected whereas $\hat{d}_7$ and $\hat{d}_{40}$ are greater than expected (cf. Fig. 1). Thus we cannot say that the sequence (20) shows any tendencies to prefer short distances and avoid long distances, i.e. the linguistical hypothesis cited in the introduction cannot be verified by the text sample used here.

However we suppose that the test described here would yield other results by taking larger samples.

The confidence interval for $\alpha = 0.05$, i.e. that interval I symmetrical to $\bar{d}_i$ with the property $P(d_i \notin I) = 0.05$, is of the form

$$I := [\bar{d}_i - 1.96\sqrt{V(d_i)},\ \bar{d}_i + 1.96\sqrt{V(d_i)}], \text{ (cf. (21))}.$$

The confidence limits are listed in Table 3. Fig. 1 shows the corresponding confidence band.

REMARKS

1 Obviously it holds $r = n - d_0$, when r is defined as the number of runs in a sequence F∈F, i.e. counting the number of runs is equivalent to count the numbers of 0-distances.

2 For the summation over x note that $\binom{n}{k}$ is defined as 0 for $k\leq 0$ or $k\geq n$ (cf. section 1).

REFERENCES

ZÖRNIG, P., The distribution of distances between like elements in a sequence I. Glottometrika 6, 1984, 1-15

ZÖRNIG, P., The distribution of distances between like elements in a sequence II. Glottometrika 7, 1984, 1-14

Tab. 2: Probabilities P_i

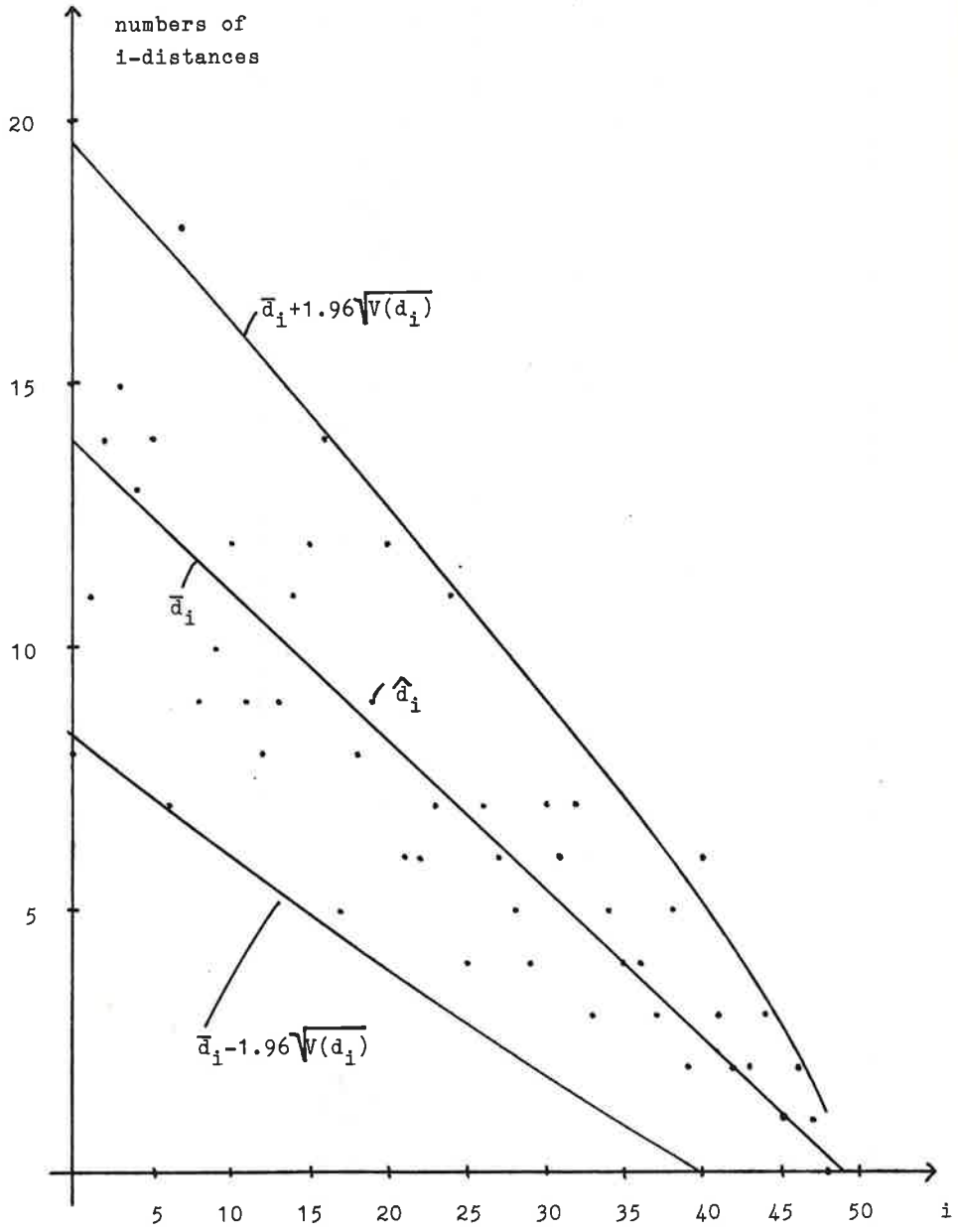
i	$\bar{d}_i$	$\hat{d}_i$	$V(d_i)$	$\hat{z}_i$	P_i
0	14	8	8.54	-2.05	0.04
1	13.71	11	8.39	-0.94	0.35
2	13.43	14	8.25	0.20	0.84
3	13.14	15	8.10	0.65	0.52
4	12.86	13	7.94	0.05	0.96
5	12.57	14	7.79	0.51	0.61
6	12.29	7	7.63	-1.91	0.06
7	12	18	7.47	2.19	0.03
8	11.71	9	7.31	-1.00	0.32
9	11.43	10	7.15	-0.53	0.60
10	11.14	12	6.98	0.32	0.75
11	10.86	9	6.82	-0.71	0.48
12	10.57	8	6.65	-1.00	0.32
13	10.29	9	6.47	-0.51	0.61
14	10	11	6.30	0.40	0.69
15	9.71	12	6.12	0.92	0.36
16	9.43	14	5.94	1.87	0.06
17	9.14	5	5.76	-1.73	0.08
18	8.86	8	5.58	-0.36	0.72
19	8.57	9	5.39	0.18	0.86
20	8.29	12	5.21	1.63	0.10
21	8	6	5.02	-0.89	0.37
22	7.71	6	4.82	-0.78	0.44
23	7.43	7	4.63	-0.20	0.84
24	7.14	11	4.43	1.83	0.07
25	6.86	4	4.28	-1.38	0.17
26	6.57	7	4.13	0.21	0.83
27	6.29	6	3.97	-0.14	0.89
28	6	5	3.82	-0.51	0.61
29	5.71	4	3.66	-0.90	0.37
30	5.43	7	3.50	0.84	0.40
31	5.14	6	3.33	0.47	0.64
32	4.86	7	3.17	1.20	0.23
33	4.57	3	3.00	-0.91	0.36
34	4.29	5	2.83	0.42	0.67
35	4	4	2.65	0	1.00
36	3.71	4	2.48	0.18	0.86
37	3.43	3	2.30	-0.28	0.78
38	3.14	5	2.12	1.27	0.20
39	2.86	2	1.94	-0.62	0.54
40	2.57	6	1.76	2.59	0.01
41	2.29	3	1.57	0.57	0.57
42	2	2	1.38	0	1.00
43	1.71	2	1.19	0.26	0.80
44	1.43	3	1.00	1.57	0.12
45	1.14	1	0.80	-0.16	0.87
46	0.86	2	0.61	1.47	0.14
47	0.57	1	0.41	0.67	0.50
48	0.29	0	0.20	-0.63	0.53

Tab. 3: Confidence limits

i	$\bar{d}_i - 1.96\sqrt{V(d_i)}$	$\bar{d}_i + 1.96\sqrt{V(d_i)}$
0	8.27	19.73
1	8.04	19.39
2	7.80	19.06
3	7.57	18.72
4	7.33	18.38
5	7.10	18.04
6	6.87	17.70
7	6.64	17.36
8	6.41	17.01
9	6.19	16.67
10	5.96	16.32
11	5.74	15.97
12	5.52	15.62
13	5.30	15.27
14	5.08	14.92
15	4.86	14.56
16	4.65	14.21
17	4.44	13.85
18	4.23	13.49
19	4.02	13.12
20	3.81	12.76
21	3.61	12.39
22	3.41	12.02
23	3.21	11.65
24	3.02	11.27
25	2.80	10.91
26	2.59	10.55
27	2.38	10.19
28	2.17	9.83
29	1.97	9.46
30	1.76	9.09
31	1.56	8.72
32	1.37	8.34
33	1.18	7.97
34	0.99	7.58
35	0.81	7.19
36	0.63	6.80
37	0.45	6.40
38	0.29	6.00
39	0.13	5.59
40		5.17
41		4.74
42		4.30
43		3.85
44		3.39
45		2.90
46		2.38
47		1.82
48		1.17

Fig. 1:

confidence band



Bio- und neurolinguistische Aspekte der dynamischen Sprachtheorie

Wolfgang Wildgen, Bremen

Ein exaktes Modell der Neurodynamik, welches dem beobachteten Sprachverhalten zugrunde liegt, wäre ein Desiderat der heutigen Sprachforschung. Mangels genauer Kenntnisse und einem nur indirekten empirischen Zugang soll versucht werden, zumindest den Rahmen für eine solche Modellkonstruktion zu entwickeln. Dies geschieht in fünf Schritten:

(1) Der Modellbildungsbereich "menschliche Sprache" wird eingegrenzt.
(2) Die verfügbaren Modellkonzepte werden verglichen.
(3) Hebbs Theorie der Neuronenverbände wird kurz beleuchtet.
(4) Topologisch-dynamische Konzepte für ein Modell der Neurodynamik werden dargestellt.
(5) Schließlich werden (in Abschnitt 5 und 6) zwei Vorschläge des Mathematikers René Thom diskutiert.

Insgesamt zeigen diese Überlegungen die Richtung einer erfolgversprechenden bio- und neurolinguistischen Theorie der Sprache an.

Vorbemerkung

Die dynamische Sprachtheorie hat zwar tiefe historische Wurzeln (vgl.Wildgen, 1985); in ihrer modernen, formalen Gestalt geht sie jedoch auf Ansätze in René Thom (1972, 1974, 1980a, b) zurück. In Wildgen (1979, 1982,1985) und in Petitot (1982) wurde sie systematisch ausgearbeitet. Das Buch von Mottron und Wildgen (1986) stellt eine systematische Erweiterung dar, bei der sowohl der breitere Rahmen der Theorie dynamischer Systeme als Modellbildungshintergrund benützt wird (vgl.Kapitel 2 in Mottron und Wildgen, 1986), als auch die biolinguistischen Ansätze René Thoms zu einer Theorie der Epigenese sprachlicher Formen dargestellt und formalisiert werden (vgl.Mottron und Wildgen, ibidem; Kap.4 und 5). In diesem Aufsatz möchte ich interessante Ansätze für eine dynamische Neurolinguistik vorstellen[1]. Da sich dieses Gebiet erst langsam, dafür aber

in interdisziplinärer Kooperation entwickelt, darf der Leser noch keine vollständige Beschreibung oder gar Erklärung neurolinguistischer Prozesse erwarten.

1. Vorüberlegung zu einer biologischen Fundierung der dynamischen Sprachtheorie

Die Neurolinguistik und in noch stärkerem Ausmaß die Biolinguistik rücken erst allmählich ins Gesichtsfeld der linguistischen Forschung. Die Forderung nach einer biologischen Fundierung der Sprachtheorie ist jedoch nicht neu (cf.Lieb, 1976: 497ff); für die statisch-klassifikatorischen Modelle wurde diese Forderung bis jetzt allerdings nicht sehr ernst genommen; man gab sich eher mit pauschalen Zusammenhängen zufrieden (cf.Chomsky, 1975 und als Kritik dazu Bauer, 1977).

In einem biologischen Kontext sind die Begriffe: Sprache, Code, Verständigung, Kommunikation und Information relativ unterbestimmt. Man spricht z.B. von der Informationsübertragung durch Viren und von genetischem Code. Unser Hauptinteresse gilt natürlich der menschlichen Verständigung (hauptsächlich der mit Sprache), trotzdem ist es biologisch nicht sinnvoll, Verständigung auf diese eine, uns am nächsten stehende Unterspezies einzuschränken. Die neueren Ergebnisse der Primatenforschung, insbesondere das Projekt in Yerk (USA) (cf.Rumbaugh, 1977), welches frühere Ergebnisse von Premack (1971) und Fouts (1975) konsolidierte, zeigen, daß die These der Abgehobenheit des Menschen wegen seiner speziellen Sprachbegabung relativiert werden muß. Insbesondere machen diese Untersuchungen auch deutlich, wie schwierig die Beurteilung der nichtmenschlichen Verständigung durch den Menschen ist. Wir wollen deshalb vorsichtig versuchen, durch die Angabe notwendiger Bedingungen für "Sprache" diesen Begriff einzugrenzen.

(1) Sprache setzt eine Organisation relativ unabhängig lebender Organismen voraus, welche einen Verband bilden, dessen Fortbestand zu sichern, im Interesse der Organismen liegt.

Diese noch recht vage Bedingung muß weiter eingeengt werden:

(2) Die einzelnen Individuen müssen Aspekte der Umgebung in sich repräsentieren können und zwar prinzipiell in gleicher Weise.

Während man bei (1) sogar noch Zellverbände zulassen könnte (allerdings nicht Kristallstrukturen), fordert (2) eine höhere Stufe der Umgebungsadaptation, was Thom auch Neurulation nennt, d.h. die Hereinnahme von Aspekten des äußeren Systems in das innere System. Die beiden Bedingungen (1) und (2) müßten nun qualitativ und quantitativ differenziert werden, so daß man schließlich den Objektbereich einer Sprachtheorie so einschränken könnte, daß:

(a) bestimmte Formen sozialer Organisation vorliegen müssen, als Vorbedingung für die Notwendigkeit und Möglichkeit "höherer" Kommunikationsmedien;

(b) bestimmte sensorisch-motorische und zerebrale Vorbedingungen erfüllt sein müssen, wodurch eine Klasse von möglichen Umweltpräsentationen im Gehirn ausgezeichnet wird. Die menschliche Form der Umweltpräsentation sollte ein zentrales Element dieser Klasse sein.

Endziel wäre eine theoretisch fundierte Abgrenzung menschlicher Sprachfähigkeit, wobei durch die Art der Abgrenzung gleichzeitig die Art der Bezogenheit auf die Kommunikationsfähigkeiten jenseits der Abgrenzung deutlich werden müßte. Nach der neueren Primatenforschung zu urteilen, wird die Abgrenzung des rezenten Menschen vom Neanderthaler und vom Primaten eine sehr differenzierte Analyse erfordern.
Ein weiteres Unterscheidungsmerkmal zwischen Mensch und höherem Säugetier ist die besonders lange Reifungs- und Prägezeit des Menschen; dabei spielen wiederum soziokulturelle Faktoren eine entscheidende Rolle.
Wir müssen unterscheiden:

(a) das Umgebungslernen: es wird stark durch die Wahrnehmungs-, Verarbeitungs- und Speicherungsfähigkeiten des Individuums gesteuert. Die Ergebnisse dieser Prozesse können deshalb biologisch (neuro- bzw.psycholinguistisch) erklärt werden.

(b) Das soziale Lernen. Erfahrungen und Wissen werden von Generation zu Generation weitergegeben oder sogar symbolisch fixiert.

Das direkte Umgebungslernen ergibt bei ähnlicher Umgebung individuell vergleichbare Resultate; in diesem Bereich können wir auch das Auftreten von Universalien erwarten. Das soziale Lernen führt jedoch zu stark differierenden Resultaten je nach sozialer Bezugsgruppe des Lernenden. Daß diese Differenzierung nicht ins Uferlose geht, dafür sind neben der partiellen biologischen Steuerung auch in diesem Bereich Stabilitätskriterien für soziale Kommunikationsmedien verantwortlich.

Damit haben wir grob den Objektbereich dieser Theorie der "Verständigung" umrissen. Er umfaßt zentral "menschliche Verständigung" und peripher die Verständigung zwischen höheren sozial organisierten Tieren. Nur der zentrale Bereich wird im folgenden näher untersucht.

Der Zusammenhang mit den niedrigeren Stufen der Verständigung ist aber nicht nur aus Gründen der Abgrenzung wichtig. Wegen der evolutionären Kontinuität ist anzunehmen, daß die einfacheren Formen der Verständigung früher Evolutionsstadien im Menschen fortbestehen und eine Art Kern unseres Verständigungsvermögens ausmachen. Ähnlich wie frühe evolutionäre Schichten in der Entwicklung des Embryos wieder auftauchen, ist es naheliegend anzunehmen, daß sowohl physiologisch, z.B. in der Struktur des Gehirns, als auch im Verhalten archaische Formen fortbestehen. Diese evolutionäre Komponente ist gehirnphysiologisch längst nachgewiesen, in der bisherigen Neurolinguistik wurde sie jedoch wenig berücksichtigt (vgl.auch Wildgen, 1983a).

Als nächstes müssen wir uns fragen, wie die biologischen Modelle wohl aussehen müßten, auf die eine linguistische Theorie aufbauen könnte. Die in diesem Abschnitt genannten biologischen Bedingungen der Verständigung zeigen:

(1) Gewisse kognitive, perzeptive und Handlungsfähigkeiten müssen zur Verfügung stehen. Dies bedeutet, daß wir ein Modell des Gehirns und seiner afferenten und efferenten Kanäle brauchen.

(2) Gewisse Fähigkeiten zur Umweltadaptation müssen gegeben sein; Wahrnehmungsresultate müssen eine stabile Repräsentation erhalten. Wir brauchen also eine Theorie des Gedächtnisses.

(3) Die evolutionäre Einbettung von Organstrukturen und biologischen Fähigkeiten muß durch die Theorie beschrieben werden. Wir brauchen eine Biogenetik der Kognition und der Kommunikation.

Eine biologische Theorie, welche diese drei Leistungen erbringen könnte, gibt es nicht. Wir werden deshalb auf Teilergebnisse und theoretische Konstrukte zurückgreifen müssen, um zumindest umrißhaft eine biologisch fundierte Verständigungstheorie angeben zu können.

2. Forschungsparadigmen in der biologischen Modellbildung

Innerhalb der Biologie gibt es eine ganze Reihe stark divergierender Forschungsrichtungen; zwei Extreme sind die physiologisch (letztlich chemisch-physikalisch) fundierten Modelle und die verhaltensbiologischen Modelle. Die physiologischen Modelle können entweder direkt chemische, elektrische oder histologische Befunde zusammenfassen oder diese generalisieren. Die Generalisierung wird häufig dadurch bewerkstelligt, daß kleinere formale Modelle konstruiert werden, welche die vorgefundenen lokalen Eigenschaften simulieren. Die Verhaltensmodelle haben dagegen eine andere empirische Basis, da das Verhalten des Gesamtorganismus bei unterschiedlichen Bedingungen beschrieben wird. Auf dieser Basis werden "integrative" Begriffe gesucht und es werden Modelle entwickelt, welche mit den gesicherten Beobachtungen kompatibel sind.

Zwischen den beiden Paradigmen biologischer Modellbildung (auch für psychologische Modelle kann eine analoge Einteilung vorgeschlagen werden) gibt es eine Beobachtungskluft, welche nur notdürftig durch chirurgische und therapeutische Befunde (z.B. aus der Aphasieforschung) geschlossen werden kann. Gerade in diesem etwas orientierungslosen Zwischengebiet können qualitative und formale Modelle sehr hilfreich sein, welche einerseits eine plausible Gesamtsicht vermitteln und anderer-

seits auf Prinzipien rekurrieren, die in ihrer Relevanz gesichert sind. Ähnliche metawissenschaftliche Überlegungen liegen der folgenden Einteilung biologischer Modelle durch Zeeman (1977:290) zugrunde.

Small - scale	Medium - scale	Large - scale
neurology	dynamic	psychology
explicit	implicit	explicit
quantitative	qualitative	quantitative

$$X \cong Y \cong Z$$

Das Problem der Beziehung zwischen $X \cong Z$ ist, daß es sich um eine grobe, strukturerhaltende Abbildung handelt, welche die quantitativen Größen verändert und qualitative Ähnlichkeiten erhält (wegen des unterschiedlichen empirischen Substrats). Ein integratives vermittelndes Modell muß deshalb topologisch und qualitativ sein, es ist indirekt über die an die Empirie angebundenen Modelle X und Z zu rechtfertigen. Die Problematik dynamisch-topologischer (z.B. katastrophentheoretischer) Modelle in der Biologie läßt sich teilweise durch diesen Zwischenstatus von medium - scale - Modellen erklären.

Welche Position nimmt in dieser Klassifikation nun die Linguistik ein? Einerseits ist sie wie die Psychologie und die Verhaltensbiologie eine large-scale-Theorie. Im Vergleich zur Psycholinguistik werden sogar relativ große Ausschnitte des Verhaltens thematisiert, d.h. die Linguistik ist globaler als die Psycholinguistik. Gleichzeitig sind die Überprüfungsverfahren relativ schwach entwickelt, d.h. die Linguistik ist eher qualitativ als quantitativ. Sie ist nämlich selbst wieder ein medium-scale-Modell zwischen Psychologie (small scale) und Soziologie (large scale).

Psychologie	Linguistik	Soziologie
explizit	implizit (relativ)	explizit (relativ)
quantitativ	qualitativ	quantitativ

Linguistische Modelle sind in der empirischen Interpretation ihrer theoretischen Begriffe ziemlich implizit. Die mentalistische Position mit der Intuition des Forschers als Beobachtungsbasis versucht z.B. diese Implizitheit zu standardisieren. Die soziologischen und psychologischen Modelle sind dagegen direkter mit experimentellen Daten korrelierbar.

Thom unterscheidet (im Rahmen katastrophentheoretischer Modelle) zwei Kategorien von Modellen:

(1) Die <u>induktiven Modelle</u>. Man verfügt über exaktes Wissen in Bezug auf das thematisierte Phänomen. Dies gilt z.B. für die klassische Mechanik und Teile der Physik. Mit Hilfe der Katastrophenmodelle und verwandter mathematischer Ansätze kann das System dieses Wissens verfeinert oder umstrukturiert werden. Das Modell hat also die Funktion, neue Begrifflichkeiten für die Theoriekonstruktion bereitzustellen.

(2) Die <u>deduktiven Modelle</u>. Thom spricht von einem "a priori " oder "metaphysical way" (in: Zeeman, 1977: 235). Hier ist die eigentliche Sachverhaltsstruktur noch recht ungeklärt. Man postuliert auf Grund bestimmter Einschätzungen, daß das dynamische Modell als Erklärungsmuster für das beobachtete Verhalten in Frage kommt, und versucht, auf dieser Basis die zugrundeliegende Dynamik und Struktur des beobachteten Verhaltens zu rekonstruieren. In einem zweiten Schritt werden diese Modelle durch Anwendungen konkretisiert. Dabei wird ihre Fruchtbarkeit überprüft; diese ergibt schließlich das wesentliche Evaluationskriterum für solche Arbeiten (vgl. die "paradigmatischen Beispiele" von Sneed; Stegmüller, 1975: 513).

Die topologischen Modelle eignen sich besonders gut zur Konstruktion von medium-scale-Theorien, da sie die Inkompatibilität der Beobachtungs- und Meßmethoden überbrücken können. In strukturreichen Anwendungsgebieten mit exakten Meßverfahren beschränkt sich ihre Anwendung mehr auf Zwischenbereiche, die relativ unstabil sind, so daß lokale Kausalzusammenhänge kompliziert oder gar chaotisch sind.

Gerade der linguistische Zugang verlangt somit in zweifacher Weise nach einer qualitativen Modellbildung. Wir können eine erste (Fundierungs-) Ebene linguistischer Modellbildung unterscheiden, bei der es darum geht, die Sprachproduktion und -rezeption durch ein qualitatives Modell zu erklären und eine zweite Ebene, welche die Regularitäten des sprachlichen Verhaltens, die geltenden Gebrauchsnormen und deren Funktionen beschreibt. Die zweite Ebene ist der traditionelle Objektbereich der Linguistik, die erste Ebene ist seine notwendige Fundierung.

Man kann die paradigmatischen Beziehungen zwischen Modellen der Verständigung allerdings auch noch anders beleuchten. Dewey & Bentley (1949) unterscheiden drei historische Stadien wissenschaftlicher Erklärung:

(a) self-action: "things are viewed as acting under their own powers. This ist the case in Aristotelian substance in which things possess being which is eternal in action according to their nature" (Livingstone, 1974: 37f).

(b) interaction: "The laws of Newton are classic examples of interactions: 'the effects of simple forces acting on unalterable bodies'. Space and time are omitted because they constitute a fixed framework within which events take place." (ibidem: 38)

(c) transaction: "The concept of immutable substance gives way to properties of fields, what were thought of as unalterable particles become recognized as mutable and dynamic loci of energy." (ibidem)

Der "abstract speaker-hearer" der Chomsky-Grammatik gehört eindeutig in die erste Kategorie. Die Archetypensemantik, welche in Wildgen (1979, 1982, 1985) entwickelt wurde, gehört zum "interaction"-Paradigma, während eine breitere Theorie der Verständigungsdynamik, welche soziale und historische Prozesse in der Sprache thematisiert, schon auf das "transaction"-Paradigma ausgerichtet ist. Die drei Ansätze sind kein Entweder - Oder; sie stellen eine Art Paradigmenenentfaltung dar, d.h. sie vervollständigen und relativieren das jeweils einfachere Paradigma.

3. Neuronenverbände als Verhaltenskorrelate

Die im folgenden diskutierten Vorschläge versuchen, physiologische mit verhaltensbiologischen Sachverhalten in einen systematischen Zusammenhang zu bringen. In Scott (1977: 3) wird eine Hierarchie von Untersuchungsebenen angegeben; die Ebenen (4) und (5) enthalten das Programm eines medium-scale-Modells.

(1) biochemistry from the chemical elements;
(2) membrane electrodynamics from biochemistry;
(3) neuron behavior from membrane dynamics and electromagnetic theory;
(4) neural network dynamics from neuron behavior;
(5) the elements of thought from network dynamics;
(6) psychology from the elements of thought;
(7) human personality and culture from psychology.

Einen ersten erfolgversprechenden Versuch, eine integrative Theorie aufzubauen, welche zwischen Neurophysiologie und Psychologie vermittelt, unternahm D.O.Hebb (1949) in seinem Buch "Organization of Behavior". Der zentrale Begriff ist der der "cell-assembly", d.h. es wird eine mittlere Ebene zwischen Neuron und Gehirn angesetzt. Die Neuronenverbände entstehen durch Anpassung von Zellen aneinander, so daß es zu abhängigen Funktionsmustern des Zellverbandes kommt. Die Zellverbände fördern intern die Reaktionen, inhibieren aber die Interaktionen zwischen Nachbarverbänden. Dieser Ansatz wurde in der Folgezeit durch eine große Anzahl von Arbeiten (cf.Scott, 1977: 13-25) differenziert und modifiziert. In Legéndy (1967) wird angenommen, daß die Neuronen des Gehirns bereits genetisch gesteuert in "subassemblies" organisiert sind, wobei Legéndy (1967) ein Ratio von 30 subassemblies/assembly annimmt. Jede "subassembly" hat (geschätzt) 10^4 Neuronen, was bei einer Gesamtzahl von etwa 10^{10} Neuronen 10^6 "subbassemblies" ergibt.

Die Neuronenverbände werden mit psychologischen Größen wie der Erkennung von Eigenschaften, Objekten usw. korreliert. In Legéndy (1974) "Can Neurons develop Responses to Objects?"

ergibt die Diskussion folgende Schlußfolgerungen (ibidem: 288f):

(1) Nicht alle Verbindungen im Cortex sind genetisch geplant, oft sind nur allgemeine Struktureigenschaften vorgegeben.

(2) Gewisse Fähigkeiten zur Objekt- bzw. Eigenschaftserkennung sind allerdigns ohne Lernprozeß vorhanden. Die Erstausstattung des Gehirns ist dazu bereits in der Lage.

(3) Zuerst werden sogenannte "local miracles", welche besonders prägnant sind und sich durch Nichtkonstanz auszeichnen, zu Schemata organisiert. Innerhalb eines Neuronennetzes sind komplexe und einfache Objekte wiedererkennbar, kontextuelle Einflüsse werden berücksichtigt,und es wird eine gewisse Intensivität gegenüber Imperfektionen und Deformationen entwickelt.

Dieser erste Ansatz ist deutlich elementaristisch; er versucht, durch das Einfügen von hierarchischen Ebenen zwischen der Funktion des Neurons und dem Outputverhalten die sukzessive Globalisierung der Funktionen zu erklären. Der Mechanismus besteht in einer Adaptation präsynaptischer Neuronen und im Aufbau von Inhibitionsmustern (cf.Legéndy, 1974: 277f). Solche Modelle sind allerdings schon ziemlich weit von empirisch gesicherten Fakten entfernt, da bereits die Dynamik von Synapsen, also die Transmission zwischen Neuronen äußerst kompliziert und nicht voll erforscht ist (cf.Bruter, 1976: 95f). Außerdem kennt man inzwischen weitere Mechanismen, welche den Aufbau von Gedächtnisstrukturen ermöglichen.

(a) molekulare Mechanismen.
Proteinstrukturen werden in Bändern gespeichert (cf.Bruter, 1976: 113) und können Informationen für längere Zeit konservieren.

(b) Es gibt Neurotransmittersubstanzen (cf.Bruter, 1976: 114), welche in Neuronennetzen zirkulieren können und so eine Spur der Erregung für kurze Zeit konservieren.

(c) Schließlich können sich neue synaptische Verbindungen ausbilden und somit die Funktionsweise von Neuronennetzen modifizieren.

Das genaue Zusammenwirken dieser Teilmechanismen des Gedächtnisses ist noch nicht bekannt; auf jeden Fall zeigt diese kurze Diskussion, daß die strukturelle Modellbildung von Hebb (1949) und seinen Nachfolgern noch zu einfach ist. Die Korrelation von Objekten und Wörtern mit Neuronenverbänden ist sicher verfrüht; das Gehirn ist kein Lexikon. Ein Kommentar von Taylor (1974: 252) beleuchtet die Situation in der Neuronennetzforschung sehr gut:

> "If we build models of nerve nets there are so many possible choices that we don't get very far and certainly the activity in nerve net theory of the last twenty to thirty years has shown that we have got nowhere".

Das zumindest partielle Scheitern des Neuronenverbandmodells beruht darauf, daß man versucht hat, von den physiologischen Daten über die Neuronen und Synapsen in Richtung auf eine Theorie des Verhaltens zu extrapolieren. Das physiologische Wissen für eine solche Extrapolation ist jedoch zu schmal, außerdem haben diese Theoretiker a priori eine strukturalistische Lösung mit hierarchischer Organisation von Zwischenebenen gesucht, welche dem Gegenstand nicht angemessen ist.

In dieser Situation scheint es fruchtbarer zu sein, für das medium-scale-Modell eine abstrakte topologische Sprache zu wählen und in Kongruenz mit gesicherten Detailanalysen die Grobstruktur der Gesamttheorie aufzubauen. Dieses zwar sehr generelle Gerüst kann dann als Rahmen für die methodologische Reflexion über weitere Schritte zur Erhellung des noch sehr unsicheren Zwischenbereiches dienen.

Ausgangspunkt sind nicht mehr wie bei Hebb (1949) die spezifischen, schon bekannten Eigenschaften einzelner Neuronen, sondern unser Wissen über das Funktionieren größerer Teilbereiche. Dieses Wissen kann teilweise aus bereits funktional erhellten neurophysiologischen Studien kommen, es kann aber auch auf Verhaltensbeobachtungen beruhen. Im Gegensatz zum elementaristischen Ansatz von Hebb (1949) kann man somit von einem holistischen Ansatz sprechen. Ein erster Versuch in diese Richtung wurde von Zeeman (1962) in der Arbeit: "The Topology of the

Brain and Visual Perception" unternommen. Die katastrophentheoretischen Ansätze durch Thom und Bruter fußen auf Zeemans Vorschlägen.

4. Topologisch-dynamische Modelle wichtiger Gehirnfunktionen

Der elementaristische Ansatz, der im vorherigen Abschnitt referiert wurde, verlangt letztlich eine Art Lokalisierung bestimmter Wahrnehmungsentitäten und gerät dadurch in Widerspruch mit den empirischen Ergebnissen der Neurophysiologie. Zeeman geht den umgekehrten Weg; er nimmt an, daß es Zustände der Gesamtdynamik des Gehirn gibt, welche abstrakte Korrelate von Verhaltensentitäten sind. Diese Position vermeidet eine explizite Aussage über das physiologische Zusammenwirken verschiedener Gehirnteile, läßt aber die Möglichkeit lokaler Konzentrationen im Gehirn zu.

Zwei Ideen sind zentral:

(a) Bei der Wahrnehmung kommt es zu einer enormen "Aufblähung". Ist das eigentliche Wahrnehmungsfeld, d.h. das teilsphärische Bild, das die Retina erreicht, noch zweidimensional, so hat die Retina selbst bereits 10^5 Neuronen und eine ähnliche hohe Dimensionalität von Reaktionen. Das Großhirn gar, in dem die Reize der Retina weiterverarbeitet werden, hat etwa 10^{10} Neuronen. Das Wahrnehmungsmodell muß dieser "Aufblähung" Rechnung tragen.

(b) Die Auflösung von Helligkeitsunterschieden im zweidimensiolen Bild ist begrenzt. Der Toleranz der Punktwahrnehmung muß eine Toleranz in der Reaktionen der Retina und im Gehirn entsprechen.

Als Grundmodell setzt Zeeman (1965: 284) deshalb die folgenden Toleranzräume an:

(1) $(X, \xi) \xrightarrow{e} (I^X, \eta) \xrightarrow{f} (R, p) \xrightarrow{g} (C, \gamma)$

X ist das zweidimensionale (schwarz-weiß) Bild, I^X die Menge der "Figuren" in X, R ist die Retina (10^5 Neurone), C ist der zerebrale Kubus (10^{10} Neuronen). Die Abbildungen e, f, g sind Toleranzeinbettungen; sie bewahren wichtige Eigenschaften des

Bildes in X, so z.B. die Kohärenz des Bildes und die Zahl der Komponenten. Daß das visuelle System tatsächlich primär so arbeitet, zeigen die Heilungsgeschichten von Blindgeborenen (von Senden, 1932). Die Patienten konnten gerade Linien nicht von krummen und Dreiecke nicht von Kreisen unterscheiden. Sie konnten diese Unterscheidung nach einigem Training jedoch lernen.

Die Zuordnung des Wahrnehmungsbereiches (X, ξ) bzw. von Figuren in ihm zur Gehirnstruktur (C, γ) betrifft nur die Wahrnehmungsfunktion des Gehirns.
Der Zustand des Gehirns zu einem bestimmten Zeitpunkt (Zeeman nennt ihn einen Gedanken) ist jedoch von einer Reihe anderer Faktoren abhängig, so z.B.
- vom vorangegangenen Zustand,
- von subcortikalen Anregungen,
- von der permanenten Struktur des Cortex (Gedächtnis),
- von chemischen Faktoren im Blut.

(cf.Zeeman, 1965: 279).

Für den Sprachgebrauch ist der Faktor "Gedächtnis" besonders wichtig. Die Wiedererkennung von Sinneswahrnehmungen verlangt die Existenz eines Vektorfeldes im Raum C. Wichtig sind dabei die Minima dieses Vektorraumes, da sich an diesen Stellen der zerebrale Fluß ("Gedankenfluß") stabilisieren kann. Der Wahrnehmungsinput r hat als grobes Korrelat einen Toleranzbereich in C und stabilisiert sich im Minimum eines Vektorfeldes $v(r)$ in diesem Toleranzbereich.[2)]

Im Gegensatz zur Theorie der "cell-assemblies" entspricht in Zeemans Theorie einem Wahrnehmungsobjekt in X (cf.(1)) ein zerebraler Zustand, der physiologisch beliebig komplex sein kann. Insbesondere können mehrere Gehirnteile zusammenwirken, um den Zustand zu definieren; andere können nur subsidiär oder gar nicht beteiligt sein. Während wir bis jetzt über den Raum C nur wenig aussagen können, ist zumindest der intermediäre Raum R so gut bekannt, daß Zeeman einige Strukturunterschiede zwischen Formwahrnehmung, Farbwahrnehmung und akustischer Wahrnehmung machen kann (cf.Zeeman, 1965: 287-290). Eine weitere Leistung dieses Modells ist die Beschreibung des Gedächt-

nisses als eines statischen Vektorfeldes, das dann die Inputdynamik reguliert. In ähnlicher Weise sind durch innere Aktivierungen aus subkortikalen Bereichen dynamische Prozesse (Gedanken in reflektorischem Sinn)beschreibbar. Das Modell leistet somit eine erste qualitative Annäherung an das Funktionieren unseres Gehirns.[3)]

Nun ist aber selbst R so hochdimensional (10^5 Neuronen), daß das ganze Modell einer genaueren Spezifikation unzugänglich zu sein scheint. Dies ist jedoch nicht der Fall, da man recht systematische Aussagen über die lokalen Eigenschaften in der Nähe einzelner Attraktoren und die dort anzutreffenden Prozesse machen kann. Die Typen dynamischer Prozesse lassen sich nämlich im Lokalen mit dem Thom'schen Klassifikationssatz als eben die in Thom (1972) und Zeeman (1977) erläuterten elementaren Katastrophen erweisen. Dies bedeutet inhaltlich, daß lokal selbst bei einem noch so hoch dimensionalem Raum (z.B.in C mit 10^{10} Dimensionen) der Prozeß durch maximal 6 (meist nur 4) Entfaltungsparameter gesteuert wird.[4)] Wir erhalten somit z.B. die Kuspenbifurkation beim Übergang der Dominanz zu einem andern Attraktor.

Diese wenigen zugrundeliegenden Faktoren bzw. die in der Elementarkatastrophe eingebetteten Attraktoren können eine Verhaltensinterpretation erhalten; sie sind der inhaltliche Kern des zerebralen Geschehens. Wir haben in Wildgen (1985) diese Seite des medium-scale-Modelles im Detail ausgearbeitet. Die physiologischen Korrelate sind z.Zt. noch nicht anzugeben, obwohl Ansätze dazu vorliegen (cf.Rössler, 1974: 411, 415, 417). Indirekt zeigen aber EKG-Messungen am Gehirn Effekte von Bifurkationen in bestimmten Verhaltenssituationen auf (cf.Zeeman, 1977: 295f).[5)] Die Bifurkationen auf der Basis einer Gradientendynamik mit Punktattraktoren, welche durch die Katastrophentheorie nahegelegt werden, sind jedoch nicht die einzige Möglichkeit. Die Schnelligkeit, mit der unser Gehirn auf äußere Einflüsse reagiert und die Lernfähigkeit des Gehirn legen es nahe, zumindest zusätzlich eine Art Schwingungsdynamik anzunehmen. Das innere dynamische System ist dann bereits durch eine schwache Koppelung mit äußeren Ereignissen in Resonanz zu bringen (cf.Zeeman, 1976: Duffings Equation in Brain Modelling). Da wir die teilweise noch nicht voll entwickelte Mathematik der Schwingungsdynamik nicht näher erläutern können, müssen die folgenden Bemerkungen sehr skizzenhaft bleiben.

(a) Sinneswahrnehmungen.

Das Erreichen des Schwellenwertes einer Wahrnehmung kann durch die Kuspenkatastrophe beschrieben werden. Die Frequenz des Feuerns der Neuronen etwa in der Retina entspricht dem Kontrollparameter v , die Amplitude der zerebralen Reaktion entspricht dem inneren Parameter x . An einem bestimmten Punkt des Weges parallel zum Parameter v (bei $u < o$) kommt es zum Katastrophensprung, d.h. der Parameter x (=Amplitude) wird drastisch erhöht, der Schwellwert der Wahrnehmung ist überschritten.

(b) Assoziation.

Werden intern zwei stabile Anziehungszyklen (=Kreisattraktoren) in Verbindung gebracht, so entscheidet die Ratio der Schwingungsfrequenzen über die Stabilität des Produktes der beiden Schwingungskreise im Produkttorus. Ist die Ratio einfach, so kommt es zu einer spitzen Resonanz. Je nach Verhältnis der Schwingungsfrequenzen entstehen jedoch weniger scharfe, d.h. vage Resonanzen. Dies Modell der Assoziation kommt ohne feed-back aus; eine Koppelung der Schwingung und zufällige Deformationen des Produkttorus führen zu einem stabilen Zustand (cf.Zeeman, 1977: 297, Thom, 1974a: 220-225).

(c) Emotionalität.

Emotionale Zustände und der emotionale Zustandwechsel gehören nach Zeeman (1977: 298) zu den wohl einfachsten gehirndynamischen Phänomenen. Sie haben physiologisch ihr Zentrum im limbischen System (speziell im Hypothalamus). Man kann davon ausgehen, daß man sich immer im Dominanzbereich nur einer Emotion (eventuell im Konfliktfeld mehrerer) befindet. Die Elementarkatastrophen beschreiben

deshalb mögliche Prozesse in diesem Bereich. Dieses Phänomen ist mit den qualitativen Interpretationen in Wildgen (1985) in Verbindung zu bringen, d.h. die Einteilung von qualitativen Skalen, welche nicht durch Wahrnehmungsmetriken differenziert sind, geschieht nach emotionalen Polaritätsprinzipien.[6)]

(d) Verhaltensregulierung.

Die Verhaltensregulierung ist eine der fundamentalen Gehirnfunktionen:
"all brain excitation has ultimately one end, to aid in the regulation of motor coordination. Its patterning throughout is determined on this principle". (Sommerhoff, 1974: 347) Ausgangspunkt ist wieder der Hypothalamus, der als Zentrum des "drive-system" (cf.Sommerhoff, 1974: 306) angesetzt wird. Die umgebungsbezogene Steuerung der Bewegung wird vom höheren Cortex koordiniert, während automatisierte Teilprozesse im Kleinhirn sowie in den Basiskernen im subkortikalen Bereich gesteuert werden. Archetypische Handlungsabläufe sind wegen ihres evolutionären Alters auch sehr tief gespeichert, etwa in der Nähe der retikulären Formation (cf.Bruter, 1976: 127-131).

Der evolutionär ältere Teil des Gehirns läßt sich schematisch in drei Basisbereiche unterteilen, die Jakovlev (1972) ectopile, mesopile und endopile nennt (cf.Bruter, 1976: 82).

(1) ectopile : Umgebungswahrnehmung : S

(2) mesopile : Körperbewegungen : M

(3) endopile : innere Wahrnehmungen und Bewegungen : I

Der Differenzierungsprozeß wird in Abb.1 schematisch wiedergegeben (cf.Bruter, 1976: 129), die Differenzierungsstadien legen eine iterierte Kuspenentfaltung nahe.

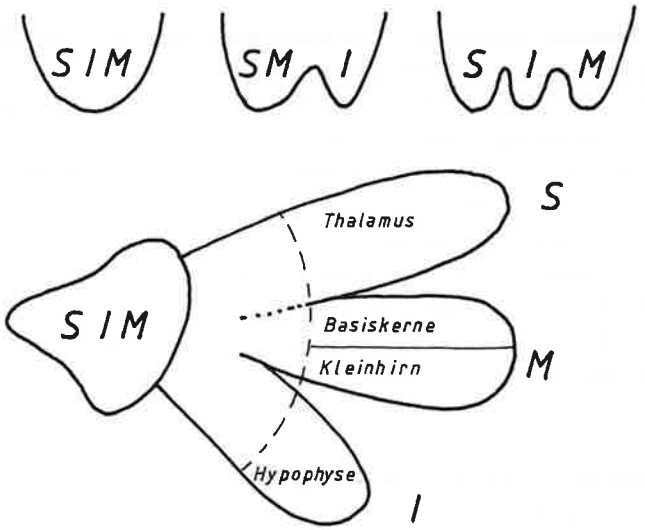


Abb. 1 (nach Bruter, 1976: 129, die Darstellung ist nicht physiologisch real, sondern funktional schematisch).

Bruter nimmt nun an, daß Verhaltensarchetypen in der Komponente I lokalisierbar sind. Abb. 2 gibt die Grobstruktur archetypischer Reaktionen an.

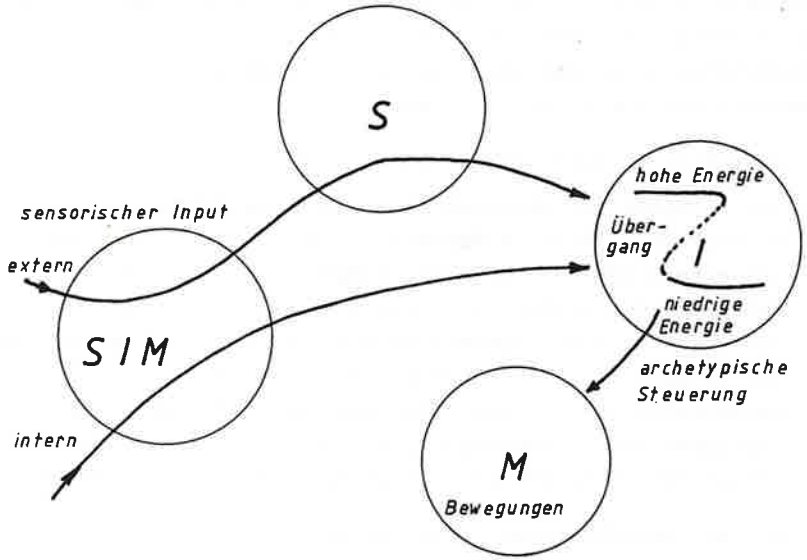


Abb. 2 (nach Bruter, 1976: 131)

In I werden dynamische Zustände aufgrund der Inputdaten aus S (bzw.SIM) stabilisiert bzw. gewechselt (Abb.2 zeigt den charakteristischen Schnitt durch die Kuspe innerhalb von I). In dieser Sichtweise gibt es für die sehr tiefen semantischen Archetypen evolutionär tiefe physiologische Korrelate. Die Differenzierung und Elaboration der semantischen Archetypen zu semantischen Strukturen, welche durch aktuelle Sätze realisiert werden, geschieht in den Assoziationsbereichen des höheren Cortex. Modellbildungen sind für diesen Bereich wesentlich schwieriger, da die zerebrale Organisation komplizierter ist als in den "alten" Gehirnteilen.[7)]

5. René Thoms Vorschläge für eine dynamische Theorie der Verständigung

René Thom hat seit 1968 eine ganze Reihe von Vorschlägen für eine psycholinguistische Theorie der Sprache entwickelt. Alle Vorschläge sondieren das Problemfeld und geben Anregungen für eine qualitative Modellbildung. Er sagt (Thom, 1977: 333): "En conclusion, on retiendra que l'analyse des structures grammaticales du langage requiert un subtil mélange d'algèbre, de dynamique et de biologie". Da die mathematischen Begriffe, die Thom einführt, teilweise den Rahmen der in diesem Aufsatz angesprochenen mathematischen Theorien sprengen, geben wir seine Grundgedanken nur informell wieder.

Zwei Metaphern durchziehen Thoms Ideen:

(1) Die Metapher der Resonanz zwischen dynamischen Systemen. So kann die zerebrale Repräsentation der Wahrnehmungswelt als eine Koppelung von äußerem System und innerem System interpretiert werden. Auch die Verständigung zwischen zwei Personen ist eine vermittelte Resonanz zwischen Denkzuständen in Sprecher 1 und Sprecher 2. Diese Metapher läßt sich mathematisch präzisieren. Sie ist auch mit gehirnphysiologischen Fakten kompatibel, so daß die Metapher bei einer detaillierten Ausarbeitung sich zum Modell entfalten kann.

(2) Die biologische Metapher der Befruchtung.
Die sprachliche Botschaft schließt die Idee des Sprechers wie ein Samen in sich ein. Kommt der Samen auf den richtigen Nährboden, bzw. verbindet er sich mit dem "Ei" im Hörer,

so entsteht ein identisches bzw. ein teilidentisches neues Wesen; die Botschaft wird verstanden. Die biologische Metapher ist zumindest ansatzweise auf die dynamische zurückzuführen. Wesentlich ist die Invarianz eines Bedeutungskernes, der in einer objektiven Struktur vom Sprecher wahrgenommen, in die Verbalisierung eingeschlossen und schließlich vom Hörer wieder aufgeschlüsselt wird.[8)]

Wir wollen die zweite Metapher, die typisch für Thoms morphogenetisches Denken ist, als solche stehen lassen und die eher mathematisierbare erste Metapher näher erläutern.

Das zentrale Element der Satzproduktion ist das Verb, das jedoch einen Interaktionsraum U (mit maximaler Dimension 3; vgl.Thom, 1974: 215) definiert. Die Festlegung der Verbdynamik (d.h. des semantischen Archetyps) führt primär zu einer Selektion möglicher Verbrealisationen und sekundär zu einer Anzahl nominaler Figuren Q_1, Q_2..., die Thom (ibidem) "figures spectrales" nennt. Ein bestimmter in Frage kommender Begriff V für eine Figur Q_i hat selbst eine Figur Q(v). Wird eine Identifikation h von V und U vorgenommen, so entscheidet die Interaktionsentropie von v und h: S (v,h) über das Zustandekommen einer Satzbedeutung bzw. einer partiellen Satzbedeutung. Die Maxima der Interaktionsentropie sind Resonanzzustände, von denen es übrigens mehrere geben kann (cf. Thom, 1974: 225f). Wenn es Diskontinuitäten der Interaktionsentropieverteilung gibt, so kommt es zu einer katastrophischen Resonanz, d.h. die Resonanz setzt mit einem Katastrophensprung ein.

Beim Sprecher hat eine solche Resonanz die Folge, daß eine geeignete Verbalisierung einer Wahrnehmung oder einer inneren Anregung gefunden wird; sie löst über efferente Mechanismen die Realisierung des Satzes aus. Der Hörer baut umgekehrt aus den Satzbestandteilen "spektrale Figuren" Q_i auf, deren Produkt hyperkomplex und unstabil ist. Erst die stabile Resonanz des Produktes der Einzelspektren führt zum Verstehen des Satzes. Tritt dies nicht ein, wird das unstabile Gebilde sehr schnell aus dem Kurzzeitgedächtnis verdrängt.

Dieser Modellansatz beleuchtet einige andere Probleme:

(1) die Verbindung von Gestik und Wort.
In Thom (1977: 315) wird ein zerebraler Konnektor ("locali-

teur - connecteur") zwischen dem semantischen Raum und der sensomotorischen Repräsentation ("carte sensori - motrice") angesetzt.
Abb.3 zeigt Thoms Vorstellung.

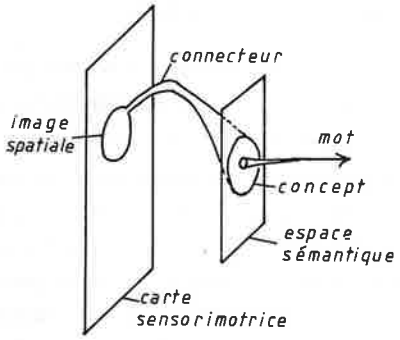


Abb. 3

Auf diese Weise ist jeder Verwendung eines Begriffs ein Bild, d.h. eine raumzeitliche Konkretisierung zugeordnet (die scharf oder vage sein kann). Im Falle des Eigennamens fallen Begriff und "Bild" zusammen. In den anderen Fällen wird die Realisierung der Referenz aufgespalten in eine semantisch abstrakte Entität, die z.B. als 'common-noun' realisiert wird, und ein Demonstrativum. Dieses kann text- bzw. situationsintern die Lokalisierung leisten oder es kann durch eine Geste ergänzt werden. Die Zeigegeste kann aber auch mit einem Pronomen zusammen die inhaltliche semantische Ausfüllung dem Hörer/Zuschauer auftragen.

(2) Die Einpassung der Figur V in den Interaktionsraum U kann dadurch erleichtert werden, daß dynamische Eigenschaften von V , welche für eine Resonanz mit U wichtig sind, eigens markiert werden. Diese Funktion füllen Kasusmarkierungen und Präpositionen aus. Sie können allerdings auch ebenso wie Verbaffixe die archetypische Dynamik elaborieren.[9)]

(3) Man kann die Begriffe, die zu verschiedenen grammatischen Kategorien gehören, nach ihrer "semantischen Dichte" ordnen (cf.Thom, 1974: 295-298). Die semantische Dichte entspricht im dynamischen Modell der Tiefe des Potentialkraters, der die Bedeutung eines Wortes stabilisiert. Je dichter ein Begriff ist, desto länger braucht die Analyse, bis sie das repräsentative Wort findet: die regulierenden Mechanismen sind komplexer.

Das Substantiv, das die reelle Substanz, die wir mit unseren Sinnen erforschen, klassifiziert, ist sicher das dichteste Element. Dabei sind die Begriffe, welche Belebtes, Menschliches umfassen, dichter als Begriffe aus dem Bereich des Unbelebten, Konkreta sind dichter als Abstrakta. In derselben Linie sind die Adjektive weniger dicht als die Substantive, da sie sich nur auf wenige qualitative Dimensionen beziehen. Die Adjektive sind jedoch wiederum dichter als die Verben. Deren raumzeitliches Korrelat ist einfacher, grundlegender als das qualitative Substrat der Substantive und Adjektive. Die Adverben modifizieren die Verben und beziehen sich oft auf eher formelle Eigenschaften (etwa auf Quantitäten), sie sind also weniger dicht als die Verben. Je äußerlicher der Gebrauch einer grammatischen Kategorie wird, desto oberflächlicher sind diese semantisch. Die oberflächlichste Ebene bilden wohl die Konnektoren der Logiker: und, oder,nicht..., da sich ihre Verwendung algebraisch beschreiben läßt:

"Au plus superficiel du langage on trouve - paradis du logicien - les copules telles que ou, et,ne... pas, etc., elles correspondent à des structures algébriques explicites qui agissent sur le matériel verbal de manière "presque" indépendante du contenu signifié". (Thom, 1974:297).

Diese Hierarchie ist für eine Elaborationssemantik, die wir nur angedeutet haben, von großer Bedeutung. Man wird allerdings weitere Informationen über die tatsächliche zerebrale

Dynamik heranziehen müssen, bis diese Mutmaßung von René Thom erhärtet ist. Eine solche Mutmaßung wäre für eine Beschreibung sprachlicher Vagheit von großer Bedeutung.

(4) Thom zieht schließlich noch Folgerungen für eine Sprachtypologie. Dabei unterscheidet er emissive Sprachtypen: Sie realisieren zuerst die semantisch weniger dichten Elemente des Satzes, sowie rezeptive Sprachtypen: Sie orientieren sich an der Verstehensdynamik beim Hörer und realisieren deshalb zuerst die semantisch dichteren Elemente. Bei den emissiven Typen müßte das Verb den nominalen Aktanten vorangehen (VSO), bei den rezeptiven müßte das Verb am Ende stehen (SOV). Wenn man das thematische Subjekt (S) ausklammert, erhält man:

emissiv	rezeptiv
Verb - Objekt	Objekt - Verb
Nomen - Adjektiv	Adjektiv - Nomen
Nomen - Genitiv	Genitiv - Nomen

Auch diese Mußmaßung Thoms ist es wert, näher geprüft zu werden. Die Morphologie der konventionellen Seite der Sprache ist jedoch sicher wesentlich komplizierter als die archetypische Morphologie, die in Wildgen (1979,1982, 1985) detailliert untersucht worden ist.

Insgesamt sind die Vorschläge von Thom interessante und tiefsinnige Anregungen für eine Semantik, die sich aus den ritualisierten Denkformen bisheriger Ansätze befreien möchte.

6. Eine dynamisch-morphologische Theorie der Bedeutung

In diesem Abschnitt soll kurz Thoms zweite Phase, seine Arbeiten zur Semiotik seit 1978, insbesondere in Thom (1980a,b) und in neueren Schriften und Vorträgen, vorgestellt werden. Wie Mottron (1983) zeigt, ist die "zweite Phase" eigentlich nur die Einlösung von Ideen, die in knappster Form Anfang der siebziger Jahre skizziert wurden (etwa in Thom, 1973b; vgl. Mottron und Wildgen, Kap.4). Für eine ausführlichere Behandlung,

welche außerdem wissenschaftshistorische Zusammenhänge und Anwendungsbereiche bzw. Nachbarfelder berücksichtigt, wird auf Mottron und Wildgen (1985/86) verwiesen.

Thoms Sprachtheorie hat zwei Angelpunkte. Erstens sein Lokalismus (vgl.Wildgen, 1985: Kap.3), dieser führt konsequent zur topologisch-dynamischen Formalisierung, die in Wildgen (1982, 1985) ausführlich dargelegt wurde. Zweitens die Kontinuität Mensch-Tier. Thom versucht, aufbauend auf verhaltensbiologischen Sachverhalten beim Tier (Säugetiere, Primaten), die spezifische Symbolkonstitution beim Menschen als Weiterentwicklung (mit grundlegender Neuorientierung) zu erklären. Diese theoretische Linie macht für Thom den Kern einer Biolinguistik aus. Sekundär ist dann diese Theorie auch auf die Mikroebenen der zerebralen Organisation auszudehnen; zuerst muß jedoch der globale Zusammenhang plausibel erklärt werden.

Im tierischen Organismus finden wir zwei Strukturquellen, die auch beim Menschen eine Rolle spielen, auch wenn durch die Sprache wesentliche Verschiebungen im Gesamtfeld eintreten:

(a) Sensorielle "Prägnanzen"; sie hängen mit der Struktur unserer und der tierischen Wahrnehmungsorgane zusammen, welche gewisse "Fenster" zur Welt definieren und auf gewisse Reize stark reagieren. Die sensoriellen Prägnanzen sind insgesamt sehr vielfältig und umgebungs- bzw. lernabhängig.

(b) Biologische "Prägnanzen"; die prägnanten Formen haben für das wahrnehmende Subjekt eine festumrissene biologische Bedeutung. Dies trifft zu auf: Nahrung, Beute, Beutetier (Räuber), Sexualpartner u.ä.. Wird eine entsprechende Form wahrgenommen, treten im Organismus des Subjektes Veränderungen von großer Stärke, Dauer und Ausbreitung ein.

Der Zusammenhang zwischen den beiden Größen, in Thom (1982) werden sie "salience" und "prégnance" genannt, erklärt die Entstehung von "bedeutsamen" Formen, von Bedeutungen. Er führt insbesondere beim Menschen, wo dieser Zusammenhang biologisch gelockert ist, zur Morphogenese von Zeichenstrukturen. Die Entfaltung der Zeichenstrukturen wird als eine Diffusion der

biologischen Prägnanz angesehen, wobei Thom mit Jakobson zwei Grundrichtungen der Diffusion annimmt:

(a) Diffusion durch Metonymie. Darunter versteht Thom eine Weitergabe der Prägnanz an raumzeitlich benachbarte Strukturen. Dies trifft z.B. auf die Übertragung der Prägnanz des Futters auf das Glockenzeichen beim Pawlowschen Hund zu. Die zufällige, bzw. experimentell arrangierte raumzeitliche Zuordnung von Glockenton und Nahrung führt zu einer Übertragung von Reaktionen des Hundes, die sich auf die Nahrung beziehen (Speichelfluß, Hinlaufen, Aufmerksamkeit), auf das Nachbarphänomen Glockenton.

(b) Diffusion durch Metaphern. Die Ähnlichkeit von prägnantem Objekt und einem anderen (nicht notwendigerweise benachbarten) Objekt führt zur Übertragung.

Damit diese Diffusion gesteuert abläuft, ist es notwendig:

(a) daß es sehr wenige und biologisch stabile "Urprägnanzen" gibt,

(b) daß die Übertragungslinien nicht zirkulär sind und die Prägnanzintensität entlang der Linie abnimmt. Die Gesamtkraft der Prägung wird also durch die Diffusion nicht vermehrt und lange Übertragungslinien führen zu geringem und instabilem "Bedeutungsgehalt" von Formen. Dies ist z.B. deutlich beim Pawlowschen Hund. In Abwesenheit des Futters wird die "Bedeutung" des Glockentones sehr schnell schwinden oder gar durch das Frustrationserlebnis zerstört werden. Auch die Kraft von Metaphern läßt bei mehrfacher Verwendung schnell nach, sie ist hauptsächlich im Prägungsakt, wo der Bezug zum Original, zur Quelle der Übertragung, lebendig ist, stark.

Beim Menschen sind die direkten Prägungsphänomene weniger stark (obgleich vorhanden) und weniger umfassend.[10] Die biologische Prägung wird vielmehr in die Sprache kanalisiert und von dieser reguliert. Nur in bestimmten Formen der Geisteskrankheit finden wir überstarke und unkontrollierte Prägnanzdiffusionen wieder, entweder als Konzentration auf wenige Träger in der Paranoia oder als Zersplitterung und Desorganisation in der Schizophrenie.[11]

In der Sprache führt die Prägnanzdiffusion zu einer sehr reichen und gleichzeitig stabilen Struktur der lexikalischen Kategorien, etwa dem Lexikon der Substantive und nominalen Einheiten. Diese sind in der Linie der Prägnanzentfaltungen von einem Feld benachbarter (vgl. die beiden Prinzipien Metonymie und Metapher) Zentren umgeben. Damit diese Entfaltung jedoch nicht zu einer willkürlichen Zersplitterung unseres Wahrnehmungs- und Lebensraumes führt, wird sie durch feste Prinzipien reguliert (so wie auch beim Tier die Notwendigkeit, eine Beute zu verfolgen, eine Stabilisierung von Prägnanzlinien erzwingt).

Diese Prinzipien sind im wesentlichen syntaktischer Natur, wobei sich die syntaktischen Randbedingungen auf raum-zeitliche Kontaktformen bzw. grundlegende qualitative Oppositionsstrukturen zurückführen lassen. Dieses Grundsystem, welches die Willkür des Bedeutungssystems verhindert und deren Entfaltung an die "objektive" Welt anbindet, ist durch die Archetypensemantik gegeben, also durch die Primitiva der dynamischen Semantik.

Der Übergang vom Tier zum Menschen, vom Instinktwesen zum Sprachwesen, ist also einerseits durch die explosionsartige Vermehrung von Prägnanzzentren (dem Lexikon) und andererseits durch deren Kontrolle durch eine ritualisierte und sehr stabile Entfaltungsstruktur gekennzeichnet.

Global gesehen bildet die rigorose und fast primitive Archetypensemantik das notwendige Gegengewicht zur zersplitterten, aufgefaserten, hochdifferenzierten Kategorisierung und kategorialen Wahrnehmung.

Die Notwendigkeit der Archetypensemantik ist damit plausibel geworden. Es bleibt aber die Frage: wo kommt diese Regulierung her, wie wird sie vererbt bzw. entwickelt oder gar welche biologisch-chemische Struktur steckt hinter der Archetypensemantik?

Diejenigen, welche sich einen Rekurs auf die Erbinformation erhoffen, müssen enttäuscht werden. Die morphogenetischen Prinzipien enthalten so grundlegende Regularitäten, daß es keiner spezifischen Kodierung bedarf, keines Homunculus in der Samenzelle. Da es sich um fundamentale raumzeitliche Prozeßstrukturen handelt, sind Spuren der Archetypensemantik wohl bis in die Evolution des Kosmos zurückzuverfolgen. Nur die spezielle Ausprägung im Menschen, der Übergang vom Tier zum Menschen, die spezielle Reorganisation, welche die Sprache entstehen ließ, bedürfen einer unmittelbaren Erklärung. In der Linie der Arbeiten von Mottron (1983) erscheint es außerdem durchaus plausibel, daß die semantischen Archetypen sich in ihrer spezifischen sprachlichen Ausprägung sekundär zur Entwicklung von Beziehungsstrukturen im primären Feld: Mutter-Welt-Kind entwickeln und damit gar nicht angeboren sind. Die Suche nach angeborenen Ideen ist eine Verzweiflungstat der algebraischen-statischen Theorien, welche, da sie keine Dynamik im Modell haben, letztlich annehmen müssen, daß alles vorgeprägt ist, entweder im Einzelnen oder in dessen Umgebung. Es darf von der Modellideologie her nichts entstehen, Komplexität darf sich nicht von selbst entwickeln; Selbsterzeugung (Autopoiese) ist eine Denkmöglichkeit, die zum Tabu wurd.

Fußnoten

1) Teile dieses Aufsatzes stimmen mit dem Kapitel 8 meiner Habilitationsschrift: Verständigungsdynamik."Bausteine für ein dynamisches Sprachmodell" überein. Dieses Kapitel wurde bei der Druckfassung der Habilitationsschrift (Wildgen, 1985) aus Platzgründen nicht berücksichtigt. Eine erste Fassung dieses Kapitels wurde 1978 im Zoologischen Kolloquium der Universität Regensburg vorgetragen. Ich danke den Prof.Altner und Boeckh (Regensburg) für die Einladung und den Gedankenaustausch. Kapitel 6 wurde 1983 hinzugefügt; dabei wurde auf neuere Arbeiten von René Thom Bezug genommen. Seit 1982 gibt es eine Kooperation mit Laurent Mottron, der in seiner Thèse d'Etat (Sorbonne,1983) Thoms neue Theorieansätze aufgegriffen hat und eine ethologisch-psychoanalytisch fundierte Theorie des frühen Spracherwerbs und der Genese autistischer Verhaltens- und Sprachstörungen entwickelt hat.

2) Die Toleranztopologie ist bei einer Präzisierung der Dynamik eliminierbar.

3) Für neuere Ansätze zur Anwendung von Toleranzräumen in der Neurobiologie cf. Dal Cin, 1974 .

4) Die anderen Variablen sind sogenannte "dummy-Variablen", die nicht entfaltungsrelevant sind.

5) Weitere Arbeiten zu diesem Problembereich sind Zeemann, 1973, 1977 und Gurel, 1973. Inzwischen gibt es allerdings eine größere Anzahl von Modellbildungen für den medium-scale-Bereich (cf.Conrad, Güttinger & Dal Cin, 1974). Keine erreicht jedoch die Anschaulichkeit und Eleganz der katastrophentheoretischen Modelle. Eine Koordination verschiedener Zugänge wird jedoch die Zukunft in diesem noch sehr offenen Bereich bestimmen.

6) Die verschiedenen Sinnesorgane sind zur unterschiedlichen Auflösung objektiver Strukturen befähigt. Cf.Zeemann (1965: 287-290); er baut eine Differenzierungsskala: Formwahrnehmung (quasimetrisch) - Hörwahrnehmung (topologisch mit Ordnungseigenschaften) - Farbwahrnehmung (nur topologisch bzw. toleranzrelational) auf.

7) Cf. Zeemann & Bunemann (1968). Dort wird ebenfalls den Mittelhirnregionen (Thalamus, Hypothalamus, Hypocampus, retikuläre Formation) eine zentrale Steuerungsfunktion zugesprochen. Er verweist die flexibleren Synapsen in diesem Bereich, welche eine rasche Adaptation ermöglichen und dann zu sekundären Adaptationen im weniger flexiblen Cortex führen können. Auch das Holographie-Modell des Gedächtnisses in: Pribram (1971) paßt in dieses Konzept. In

dieselbe Richtung verweisen die physiologisch-aphasiologischen Untersuchungen von Brunner & Wallesch (1979); ihre Ergebnisse lassen vermuten, daß die bisherigen aphasiologischen Untersuchungen die Relevanz der äußeren Gehirnschichten für die Sprache überbewertet haben, da keine differenzierten Instrumente zur Diagnose des Gehirnschadens vorhanden waren. Die moderne Computertomographie erlaubt dagegen sehr genaue Lokalisationen. Die Ursachen von Verhaltensfolgen können deshalb viel enger eingegrenzt werden.

8) Cf. Thom (1974: 217): "dans certaines conditions d'excitation, le concept fabriquera un "gammète porteur du "logos" du concept. Ce gamète n'est autre que le mot, énoncé par le locuteur. Dans l'esprit de l'auditeur, le mot véritable semence du concept, pourvu qu'il rencontre un contexte approprié, germe et éclate: le "logos" du concept se déploie, et reconstitue la figure de regulation du concept, donc sa signification".
oder Thom (1977: 314): "Le gaméte émis par le concept n'est autre que le mot (le nom correspondant). L'émission verbale apparaitainsi comme un véritable orgasme".

9) Diese Tatsache läßt es fast unmöglich erscheinen, durch eine Analyse der Kasus- bzw. Präpositionensysteme die Archetypenliste zu rechtfertigen. Diese morphologischen Subsysteme sind immer auch elaborierend.

10) Wir finden hier die Sprachursprungstheorie Herders wieder, der auf die Instinktschwäche des Menschen als Vorbedingung und Motiv der Entwicklung einer Sprache verwiesen hat. Vgl. Wildgen (1983c).

11) Zur Pathologie der Sprache vgl. Mottron (1983) und Mottron und Wildgen (1985/86).

Bibliographie

BAUER, Ch, Reflection on Universal Grammars. Synthese 37, 239-251, 1977

BRUNNER, R.J. und C.W.Wallesch, Sprachstörungen bei Basalganglien-Läsionen. Eine Kritik der klassischen Aphasietheorie. In: W.Kühlwein und A.Raasch (Ed.), Sprache und Verstehen, Bd.II,Tübingen: Narr, 127f, 1980

BRUTER, C.P., Topologie et perception, Tome 1: Bases philophiques et mathématiques, Doin/Maloine, Paris, 1974

BRUTER, C.P., Topologie et Perception, Tome II, Aspects neurophysiologiques, Paris: Maloine, 1976

CHOMSKY, N., Reflections on Language, New York: Pantheon, 1975

CONRAD, M., Güttinger, W. und Dal Cin, M. (Ed.), Physics and Mathematics of the Nervous System, Reihe: Lecture Notes in Biomathematics 4, Berlin: Springer, 1974

DAL CIN, M., Modifiable Automata with Tolerance. A Model of Learning. In: Conrad, Güttinger & Dal Cin, 1974: 442-455

DEWEY, J. und BENTLEY, A., Knowing and the Known, Boston: Beacon, 1949

DORAN, J.,Some Recent Models of the Brain. In: Meltzer, B. und Michie, D., Machine Intelligence, New York, 1971: 207-220

FOUTS, R.S., Communication with Chimpanzees. In: C.Kurth und I.Eibl-Eibesfeldt (Ed.) Hominisation und Verhalten, Stuttgart: Fischer, 1975

GUREL, O., Topological Dynamics in Neurobiology, Intern.Journal of Neuroscience 6, 1973: 165-179

HEBB, D.O., The Organization of Behavior, New York: Wiley, 1949

LEGENDY, C.R., On the Scheme by which the Human Brain Stores Information, Math.Biosciences 1, 1967: 555-597

LEGENDY, C.R.,Can Neurons Develop Responses to Objects?. In: Conrad, Güttinger & Dal Cin, 1974: 269-289

LIEB, H.H., Universals of Language: Quandaries and Prospects, Foundations of Language 12, 1975: 471-511

LIVINGSTON, R.B., Some Limitations Affecting Physics and Mathematics as Applied to Biology and Especially to the Nervous System. In: Conrad, Güttinger & Dal Cin,1974: 31-39

LORENZ, K., The Role of Gestalt Perception in Animal and Human Behaviour. In: Lancelot L.W. (Ed.), Aspects of Form, London: Lund, 1968

MOTTRON, L., Contraintes communes à l'aquisition, la théorisation et la pathologie de la deixis, thèse d' Etat, Sorbonne, Paris, 1983

MOTTRON, L. und W.WILDGEN, Dynamische Sprachtheorie.Dynamische Modelle für die Sprachstruktur, den frühen Spracherwerb und seine Pathologien, Bochum:Brockmeyer, 1986

PETITOT-COCORDA, J., Pour un schématisme de la structure. De quelques implications sémiotiques de la théorie des catastrophes, Thèse d'Etat, Sorbonne, Paris, 1982

PREMACK, D. Language in Chimpanzee?, Science 172, 808-822; deutsch: Sprache beim Simpansen?. In: I.Schmidetzky (Ed.). Über die Evolution der Sprache, Frankfurt: Fischer, 1973: 91-131

PRIBRAM, K.H., Language of the Brain, New York: Prentice Hall, 1971

RÖSSLER, O.E., Chemical Automata in Homogeneous and Reaction-Diffusion Kinetics. In: Conrad, Güttinger & Dal Cin, 1974: 399-418

RUMBAUGH, D.M., Language Learning by a Chimpanzee. The Lana Project. New York: Academic Press, 1977

SAARI, D.G., A Qualitative Model for the Dynamics of Cognitive Processes, Journal of Math.Psychology 15, 1977: 145-168

SCOTT, A.C., Neurodynamics: A Critical Survey, Journal of Mathematical Psychology 15, 1977: 1-45

SENDEN, M. von, Raum- und Gestaltauffassung bei operierten Blindgeborenen vor und nach der Operation, Leipzig: Barth, 1932

SOMMERHOFF, G., Logic of the Living Brain, London: Wiley, 1974

SPERRY, R.W., Brain Bisection and Mechanisms of Consciousness. In: Eccles, J.C. (Ed.), Brain and Conscious Experience, Berlin: Springer, 1966

STEGMÜLLER, W., Hauptströmungen der Gegenwartsphilosophie , Bd. 2, Stuttgart: Kröner, 1975

TAYLOR, J.G., Neural Networks and the Brain. In: Conrad, Güttinger & Dal Cin, 1974: 230-253

THOM, R., Stabilité structurelle et morphogénèse. Essai d'une théorie générale des modèles, New York: Benjamin, 1972

- Übersetzung: Structural Stability and Morphogenesis, New York: Addison Wesley (teilweise verändert, die linguistischen Ansätze sind verkürzt).
- Zweite überarbeitete, korrigierte und erweiterte Auflage (cf.Thom, 1977)

THOM, R., Langage et catastrophes: éléments pour une sémantique topologique. In: Peixoto, M.M. (Ed.) Dynamical Systems. Proceedings of the Symposium at Salvador, Brazil, 1971, London: Academic Press, 619-654; wiederabgedruckt (leicht verändert) in: Thom, 1974: 89-126. (Thom 1973a)

THOM, R., Sur la typologie des langues naturelles: essai d'interprétation psycholinguistique. In: Gross, M., Halle, M. und Schützenberger, M.-P. (Ed.), 1973. Formal Analysis of Natural Languages, Mouton, Den Haag; wiederabgedruckt in: Thom, 1974 : 289-313. (Thom 1973b)

THOM, R., 1974. Modèles mathématiques de la morphogénèse. Recueil de textes sur la théorie des catastrophes et ses applications, Reihe: 10-18, Union Générale d' Editions, Paris.

- Zweite, erweiterte Auflage, Paris: Bourgeois, 1980
- Englische Übersetzung (mit Veränderungen); Mathematical Models of Morphogenesis, Chichester: Horwood, 1983

THOM, R., Stabilité structurelle et morphogénèse. Essai d'une théorie des modèles. Deuxiéme Edition, revue, corrigée et augmentée, Paris: Interéditions, 1977

THOM, R., L'espace et les signes, Semiotica 29; 193-208, 1980a

THOM, R., Prédication et grammaire universelle, in: Fundamenta Scientiae, 1: 23-34,1980b

THOM,R., Psychisme animal et psychisme humain, preprint, 1982a

THOM,R., Sens de la forme et forme du sens, Vortrag beim Kolloquium: Logos et catastrophes, Cérisy, 1982b

WILDGEN, W., Verständigungsdynamik. Bausteine für ein dynamisches Sprachmodell. (Habilitationsschrift, Ms.) Regensburg , 1979

WILDGEN, W., Catastrophe Theoretic Semantics. An Elaboration and Application of René Thom's Theory, Amsterdam: Benjamins, 1982

WILDGEN, W., Skizze einer katastrophen-theoretisch fundierten dynamischen Semantik, Linguistische Berichte, 84: 33-51, 1983a

WILDGEN, W., Modelling Vagueness in Catastrophe Theoretic Semantics. In: T.Ballmer, und M.Pinkal (Ed.), Approaching Vagueness, Amsterdam: North Holland, 317-360 1983b

WILDGEN, W., Goethe als Wegebereiter einer universalen Morphologie (unter besonderer Berücksichtigung der Sprachform). In: Jahresbericht des Präsidenten der Universität Bayreuth 1982: 235-277 und in: L.A.U.T. Preprints, Trier, 1983c

WILDGEN, W., Archetypensemantik. Grundlagen für eine dynamische Semantik auf der Basis der Katastrophentheorie, Tübingen: Narr, 1985

WOODCOCK, A.E.R., Catastrophe Theory and the Modelling of Biological Systems. In: Matthews, D. (Ed.), Mathematics and the Life Sciences, Lecture Notes in Biomathematics 18: 343-385, 1977

ZEEMAN, E.C., The Topology of the Brain and Visual Perception. In: Fort, M.K. (Ed.), Topology of 3-manifolds, New Jersey: Prentice Hall, 1962

ZEEMAN, E.C., Topology of the Brain, Mathematics and Computer Science in Biology and Medecine, Medical Research Council Publications, 1965

ZEEMAN, E.C. Catastrophe Theory in Brain Modelling, Internat. Journal of Neuroscience 6: 39-41, 1973

ZEEMAN, E.C., Levels of Structure in Catastrophe Theory. In: Proceedings of the Internat.Congress of Mathematics, Vancouver, Bd.2, 533-546, 1974; wieder in: Zeeman, 65-78, 1977

ZEEMAN, E.C., Catastrophe Theory. Selected Papers, Reading (Mass.): Addison-Wesley, 1977

ZEEMAN, E.C., und D.P.BUNEMANN, Tolerance Spaces and the Brain, in: C.H. Waddington (Ed.) Towards a Theoretical Biology, Vol.1, 140-151.

ZUR SPRACHUNABHÄNGIGKEIT SUPRASEGMENTALER PHONETISCHER EIGENSCHAFTEN*

Werner Enninger/Reinhard Köhler/
*Helge Korda/Joachim Raith***

0. Gegenstand, Fragestellung, Ansatz, Ziel

Den Gegenstand der folgenden Studie bilden suprasegmentale Eigenschaften von Sprachen verschiedener Sprachfamilien und verschiedener Sprachtypen. Die übergeordnete Fragestellung ist, ob auf der phonetischen Ebene der syntagmatischen Dimension solch verschiedener Sprachen mit verschiedenen Phonemsystemen sprachunabhängige Eigenschaften welcher Art auch immer vorkommen oder nicht.

Die Anregung zu dieser Fragestellung ergab sich aus dem Konzept der design-features. Bekanntlich wurde dieses Konzept von Charles Hockett (1960a, b) zuerst formuliert. Die design-features

* Das Vorhaben wurde dankenswerterweise aus dem Forschungspool der Universität - GHS Essen gefördert.

** Die Studie ist das Resultat von Teamarbeit. Joachim Raith betreute die phonetische Komponente. Helge Korda transkribierte die Mehrzahl der Texte, erstellte die Lautmatrizes und führte die grundlegenden Rechnungen durch. Reinhard Köhler brachte die mathematischen Modelle und Verfahren ein und führte die Autokorrelationsanalysen durch. Werner Enninger koordinierte das Projekt. Das Team dankt Gothild Thomas (Essen) für das Japanische und Justus Roux (Stellenbosch, Südafrika) für das Afrikaans-Material.

waren als ein einheitlicher heuristischer Bezugsrahmen konzipiert, als Satz von Instrumenten einer komparativen Methode (1960b:89), die in der Anwendung auf verschiedenartige Zeichensysteme (Bienentanz, das Werbeverhalten des Stichlings, die Rufe von Gibbons, Instrumentalmusik, parasprachliche Phänomene, Sprachen; vergleiche auch Thorpe 1972) deren gemeinsame und spezifischen Eigenschaften entdecken sollten, oder besser: das alltagsweltlich Bekannte an diesen Eigenschaften in eine überprüfbare Systematik überführen sollten. Da die design-features der Sprache bisher am ausführlichsten untersucht wurden, geben die design-features der Sprache die umfangreichste Liste von Kriterien wieder. Von daher scheint es angezeigt, das Konzept der design-features am Beispiel der Sprache zu erläutern. Nach Hockett (1960a, b), Hockett und Altmann (1968), Thorpe (1972), und Lyons (1977:70-85), hat Sprache die folgenden design-features:

1. vocal auditory channel
2. broadcast transmission and directional reception
3. rapid fading of acoustic signal
4. interchangeability of transmitter and receiver
5. complete feedback of acoustic signal
6. specialization
7. semanticity
8. arbitrariness
9. discreteness
10. displacement
11. openness
12. tradition
13. duality of patterning
14. prevarication
15. reflectiveness/reflectivity/reflexivity und
16. learnability

Poyatos erweitert diesen Satz von design-features, um zu einem umfassenderen Instrumentarium mit einem größeren Anwendungsbereich zu gelangen. Unter den zusätzlichen sieben design-features (1981:9), sind

- shared or idiosyncratic nature of the signaling system
- its use as an interactive tool
- the verbalization versus non-verbalization feature und costructuration with preceding, simultaneous and following sign processes going on in the same or other channels

für den Vergleich von Sprache mit anderen Zeichensystemen aufschlußreich. In einer Anweisung für empirische semiotische Analysen formuliert Krampen (1981:135) Aufgaben, von denen einige auf die obengenannten design-features abheben, von denen andere aber eine echte Erweiterung des Instrumentariums darstellen. Zu solchen Erweiterungen gehören:

- unterscheide zwischen Problemen der Kommunikation und Problemen der Signifikation
- bestimme, ob das Signalsystem autonom oder nicht autonom ist, wie zum Beispiel der Morsekode oder der Flaggenkode, die nicht autonom, sondern sekundäre Ausdruckssysteme natürlicher Sprachen sind
- bestimme die geformte Substanz des Signifikanten und den Übertragungskanal sowie die geformte Substanz des Signifikats
- bestimme die Art der Beziehung zwischen Signifikant und Signifikat (intrinsische versus extrinsische Kodierung)
- beschreibe die Kompositionsregeln für die Elemente und
- bestimme, welche Kompositionen prädikative Funktionen haben

Erkennbar sind diese design-features auf einer Abstraktionshöhe angesiedelt, die die Hoffnung ausschließt, durch ihre Anwendung auf Sprache, und vor allem auf die einleitend skizzierte syntagmatische Dimension der phonetischen Ebene, Neues zu entdecken. Interessant bleibt das Konzept der design-features für die eingangs gestellte Frage insofern,

1. als sie in ihrem ursprünglichen Anwendungsbereich vom Status distinktiver Merkmale zum Status konstitutiver Merkmale von Zeichensystemen übergewechselt sind, und

2. insofern als sie nicht nur Systemeigenschaften von Zeichensystemen im Sinne von Hjelmslev thematisieren, sondern über Zeichensysteme hinausgreifen auf die Ebene der Materialisation und des Transportes von Zeichenkörpern in Kanälen und dabei die Produktion und Rezeption der Signale in Rechnung stellen (vergleiche design-features 1-5)

Beide Aspekte seien kurz erläutert:

Zu 1:
Die design-features der verschiedenen Zeichensysteme werden derzeit nicht allein als das angesehen, was die verschiedenen Zeichensysteme voneinander unterscheidet, sondern als das, was die verschiedenen Zeichensysteme jeweils als eigene Zeichensysteme konstituiert. Hieraus ergibt sich die Frage, ob in Analogie zu diesem Statuswechsel der design-features nicht überlegt werden könnte, ob das, was an lautlichen Merkmalen von Sprachen bisher unter dem Stichwort "Distinktivität" abgehandelt wurde, unter dem Stichwort "Konstitutivität" (vergleiche Altmann/Lehfeldt 1980) wieder aufgenommen werden kann.

Zu 2:
Phoneme sind, wie die Einheiten anderer linguistischer Ebenen auch, nach strukturalistischem Verständnis Zeichen, deren Stellung im System jeweils von Interesse ist, während ihre materielle Gestalt als irrelevant angesehen wird. Im Falle der Phoneme heißt dies insbesondere, daß die Formung der Zeichen aus primärem, akustischem Material nicht linguistisch untersucht wird. Nun muß aber jede Sprache ihre Zeichen, zum Beispiel die Phoneme, in irgendeiner Weise in wahrnehmbarem Material realisieren. Es läßt sich beobachten, daß die Auswahl des Materials zur Formung der Zeichen nicht arbiträr vor sich geht, sondern gewissen Regelhaftigkeiten unterliegt, die viele Sprachen in dieser Hinsicht

vergleichbar machen. Aus dieser Beobachtung ergibt sich die Überlegung, ob es außersprachliche, beispielsweise kanaltechnische, artikulationsphysiologische oder wahrnehmungspsychologische Anforderungen gibt, denen die materielle Seite der Zeichen genügen muß, damit das Zeichensystem seiner Bestimmung gerecht werden kann. Denn die Realisierung eines Phonems als Phon ist

1. an den akustischen Kanal gebunden, dessen Charakteristiken gewisse Vorkehrungen erfordern, damit die Nachrichtenübertragung funktionieren kann, und unterliegt

2. den anatomisch-physiologischen Bedingungen (möglichst unaufwendiger) Lautproduktion und muß

3. die wahrnehmungspsychologischen Bedingungen der (möglichst unaufwendigen) Diskrimination erfüllen.

Solche außersprachlichen Anforderungen an das Sprachsystem, und insbesondere die Form seiner Realisierung kann man unter systemtheoretischem Gesichtspunkt als Bedürfnisse interpretieren, die auf die in der Sprache ablaufenden Mechanismen und Prozesse einwirken, indem sie Randbedingungen setzen, denen sich optimal anzupassen das System durch Selbstregulationsvorgänge (vergleiche Altmann 1981) versucht. In diesem Fall aber müssen die Auswahlprinzipien für die Eigenschaften der äußeren Gestalt der Zeichen als allgemeine Gesetze formulierbar sein, die für alle Sprachen gelten. So erfordert die Kanaleigenschaft der Flüchtigkeit, daß wenigstens einige der Zeichen von Dauer sind, während wahrnehmungspsychologische Erfordernisse die Verwendung von Lauten verlangen, die einen ausreichenden Kontrast ermöglichen. Entsprechende konstruktive Merkmale müssen also in den Phonemsystemen aller Sprachen vorhanden sein.

Hieraus ergibt sich das Ziel dieses Beitrags. Es gilt herauszufinden, ob in der phonischen Realisation der Lautsysteme verschiedener Sprachen über die bekannten segmentalen Eigenschaften hinaus weitere einzelsprachunabhängige phonetische Eigenschaften auf der syntagmatischen Dimension erkennbar sind, und wenn ja, welche, sowie in wie starker Ausprägung. Die Untersuchung bezieht nicht nur Kontaktphänomene zwischen unmittelbar benachbarten Segmenten (wie z.B. Assimilation und Dissimilation) ein, sondern versucht auch und gerade die Rekurrenz phonetischer Merkmale in nicht unmittelbar benachbarten Segmenten zu bestimmen. Dabei geht es darum, herauszufinden, ob sich verschiedene Einzelsprachen in dieser Hinsicht so wenig divergent verhalten, daß man die rekurrenten Distributionsmuster von phonetischen Merkmalen als ein einheitliches Bauprinzip von Sprache postulieren kann. Der Versuch schließt sich damit dem Firthschen (1948) Verständnis von prosodischer Phonologie an; einerseits ist dieser Ansatz kaum ausgearbeitet (vgl. Crystal 1976:61-62), andererseits sind zur Analyse solcher Phänomene von der Psychologie und der quantitativen Linguistik gewisse Vorarbeiten geleistet (vergleiche Kapitel 2).

2. Methoden

Die hier unternommene Studie konzentriert sich also darauf, suprasegmentale syntagmatische Abhängigkeiten aufzudecken, die nicht offensichtlich (zum Beispiel deterministisch) sind. Offensichtliche und bekannte Versionen solcher Abhängigkeiten sind etwa die Vokalharmonien in verschiedenen ural-altaischen Sprachen: die Merkmale back, high, rounded sind jeweils für eine bestimmte Reichweite (Wort) auf Übereinstimmung festgelegt.

Wenig untersucht hingegen sind solche Merkmalsharmonien bzw. Disharmonien, die nicht ausnahmslos gelten bzw. deren Reichweite möglicherweise nicht so klar hervortritt. Derartige sprachliche Phänomene, die eher in Form von Wahrscheinlichkeiten auftreten, mußten von einer algebraisch orientierten formalen Linguistik zwangsläufig infolge des Mangels an Beschreibungsmitteln ignoriert werden.

In der Psychologie hat man viel früher zu stochastischen Modellen gegriffen als in der Linguistik, und zwar auch zur Beschreibung des menschlichen Sprachverhaltens. Eine Kombination von zwei mathematischen Modellen können wir aus diesem Bereich für unsere Zwecke übernehmen: 1951 und 1952 hat E.B. Newman Untersuchungen veröffentlicht, in denen er sich mit den Mustern beschäftigte, die sich ergeben, wenn man die Rekurrenz bzw. die Alternation zwischen den Merkmalen vokalisch, konsonantisch von Buchstaben in laufenden Texten betrachtet. Hierbei bediente er sich der Methode der Autokorrelation. Dieser Ansatz wurde von Köhler (1983) verallgemeinert und zur Untersuchung von Folgen phonetischer Merkmale verwendet. In dieser Arbeit wird auch der mathematische Zusammenhang zwischen der Autokorrelationsanalyse und dem stochastischen Modell der Markovketten hergeleitet. Dieses Modell erlaubt es, die Abhängigkeit der Auftretenswahrscheinlichkeit für ein phonetisches Merkmal an einer bestimmten Stelle der Äußerung von den Merkmalsausprägungen, die unmittelbar vorhergehen, mathematisch exakt auszudrücken. Die Anzahl der unmittelbar aufeinanderfolgenden Positionen, die man kennen muß, um die nächstfolgende vorherzusagen, nennt man die Ordnung der Markovketten.

Bei der Autokorrelationsanalyse geht man folgendermaßen vor: Eine Position in der Lautkette wird fixiert und deren Merkmalsausprägung wird festgestellt (üblicherweise werden lediglich dichotome Merkmalsausprägungen angesetzt: +/- vokalisch, +/-konsonantisch, +/- sonorant, +/-stimmhaft etc.). Für alle weiteren Stellen der Lautkette - in der Praxis genügen die darauffolgenden zehn Positionen - wird festgestellt, ob die Merkmalsausprägungen für das gleiche Merkmal mit der fixierten Stelle übereinstimmen oder nicht. Der gleiche Vorgang wird für alle Positionen der Lautkette wiederholt, wobei die Ergebnisse jeweils für eine bestimmte Entfernung (Verschiebungsweite) zusammengefaßt werden. Auf diese Weise erhält man ein Bild darüber, wie sich das Vorhandensein oder Nichtvorhandensein eines phonetischen Merkmals an einer bestimmten Stelle darauf auswirkt, ob es an der nächstfolgenden usw. Stelle wahrscheinlich vorkommen wird oder nicht, sowie mit welcher Wahrscheinlichkeit. Zur Veranschaulichung des

Gesagten soll ein Beispiel dienen: gegeben sei ein Text in japanischer Sprache, der so transkribiert wurde, daß für jedes Phonem ein eindeutiges Symbol gewählt wurde. Ein Abschnitt dieses Textes habe dann folgende Form:

/...ʔɔka:saŋga dɛtɛ kimaʃta ʔɔdʒi:sanjɔkɯ ʔiraʃ .../

Mit Hilfe einer Phonem-Merkmals-Matrix, die jedem Phonem des Phonemsystems den Vektor seiner Merkmalsausprägungen zuordnet, kann diese Symbolfolge in eine Folge von Merkmalsausprägungen eines Merkmals übersetzt werden. Eine 1 entspricht der Aussage, daß das gewählte Merkmal zutrifft, eine 0 dem Gegenteil. Für das Merkmal "vokalisch" erhalten wir somit

...010101001 0101 0101001 01010100101 01010...

Der Autorkorrelationskoeffizient ϕ_r, der einen Wert von -1.0 bis +1.0 annehmen kann, ist ein Maß für die Übereinstimmung eines Merkmals in bezug auf aufeinanderfolgende Positionen im Text. Allgemein hat der Pearsonsche Korrelationskoeffizient die Form:

$$\rho = \frac{\mathrm{Cov}\,(x,y)}{V(x)\;V(y)}$$

$$\rho = \frac{E(X_j X_{+r}) - E(X_j)E(X_{j+r})}{\sqrt{\left[E(X_j^2) - E^2(X_j)\right]\left[E(X_{j+r}^2) - E^2(X_{j+r})\right]}}$$

Nehmen wir als Beispiel den Autokorrelationskoeffizienten ϕ_3; er soll uns Auskunft darüber geben, wie stark die Übereinstimmung des Merkmals vokalisch einer jeden Textstelle mit dem jeweils drittnächsten Phonem ist:

```
        . . . . . .
... 0 1 0 1 0 1 0 0 1 0 1 0 1 0 1 0 0 1 0 1 0 1 ...
      . . . . . . .
```


Eine vollständige Übereinstimmung bestünde darin, daß jeder 1 ("vokalisch") an der drittnächsten Stelle ebenfalls eine 1 folgte und außerdem jeder 0 ("nicht vokalisch") wieder eine 0; diese Situation ergäbe einen Koeffizienten $\phi_3 = -1.0$. Völliger Zufälligkeit entspricht ein Koeffizient $\phi = 0$. In unserem Beispiel ist es so, daß die weitaus meisten Vergleiche mit der Schrittweite r=3 eine Nichtübereinstimmung ergeben; einige wenige (siehe zum Beispiel den fetten Pfeil) ergeben Übereinstimmung.

Die exakte mathematische Ermittlung von ϕ_3 für den von uns untersuchten Text führt zu einem Wert von -0.438. Führt man eine solche Analyse für das Merkmal vokalisch für alle Schrittweiten r=0 bis r=9 durch, kommt man bei unserem Text auf die Autokorrelationskoeffizienten:

$\varphi_0 = 1.0$	$\varphi_3 = -0.438$	$\varphi_6 = 0.120$	$\varphi_9 = -0.052$
$\varphi_1 = -0.766$	$\varphi_4 = 0.292$	$\varphi_7 = -0.056$	
$\varphi_2 = 0.598$	$\varphi_5 = -0.195$	$\varphi_8 = -0.002$	

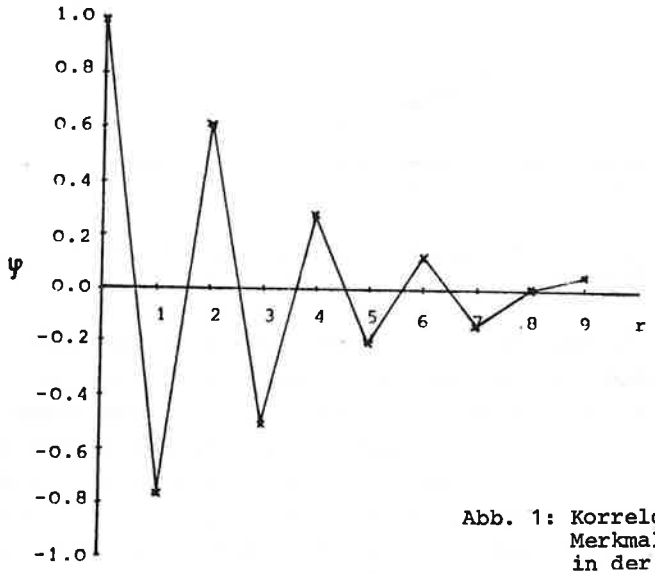


Abb. 1: Korrelogramm des Merkmals 'vokalisch' in der japanischen Sprache

Aus den Ergebnissen der Beispielsanalyse lassen sich exakte quantitative Aussagen über die Merkmalsdisharmonie von aufeinanderfolgenden Phonemen bezüglich der Vokalität entnehmen.

Ein Maß dafür, wie stark sich das Vorhandensein oder Nichtvorhandensein eines Merkmals an einer Stelle überhaupt auf die nächste Stelle auswirkt, läßt sich dadurch gewinnen, daß man die absoluten Werte der Korrelationskoeffizienten für alle folgenden Stellen summiert. Auf diese Weise ergibt sich der Korrelationssummenkoeffizient. Bei der Frage, wie stark sich die Ausprägung eines Merkmals an einer beliebigen Stelle auf seine Umgebung auswirkt, wollen wir ja nicht zwei unterschiedliche Maße für Harmonie bzw. Disharmonie verwenden. Wir benutzen daher den absoluten Wert der Korrelationskoeffizienten (vgl. Abb. 2).

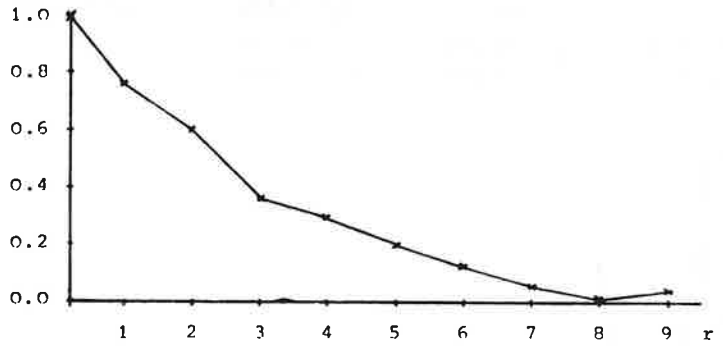


Abb. 2: Absolute Werte für das Merkmal vokalisch in der japanischen Sprache

Die Summierung der einzelnen ϕ-Werte über alle Verschiebungsweiten nehmen wir vor, weil wir eine Gesamtaussage für ein Merkmal erreichen wollen, das von den unterschiedlichen Harmoniestärken (die ja auch je nach der Verschiebungsweite differieren) unabhängig ist. Wir können zu diesem Zweck auch nicht ein ausgewähltes ϕ_r einer bestimmten Schrittweite r als repräsentativ für das Korrellogramm eines Merkmals einer Sprache ansehen, weil der Verlauf der Korrelationswerte bei weitem nicht immer so monoton ist wie im obigen Beispiel. Wir definieren also den Korrelationssummenkoeffizienten:

$$\sum_{\phi} = \sum_{i=o}^{r} |\phi_i| .$$

Insbesondere dieser Korrelationssummenkoeffizient scheint ein geeignetes Maß dafür zu sein, welche Rolle das untersuchte Merkmal bei der akustischen Realisierung der Zeichenfolgen einer Sprache spielt (vgl. Köhler 1983). Universelle Bauprinzipien, die bei der materiellen Ausformung der Zeichen aller Sprachen gleichermaßen Geltung haben, könnten mit Hilfe der Autokorrelationsanalyse nachgewiesen werden, wenn sie sprachunabhängige Korrelationsmuster für die einzelnen phonetischen Merkmale der Sprachen ergeben. Eine vollständige Gleichheit der Korrellogramme bzw. der Korrelationssummenkoeffizienten unter verschiedenen Sprachen darf man freilich nicht erwarten; Fluktuationen müssen sich unter anderem daraus ergeben, daß die Anforderungen, denen die äußere Form von Sprachen Rechnung tragen muß, keine abgeschlossene oder unveränderliche Menge darstellen. Die kanaltechnischen, wahrnehmungspsychologischen und artikulationsphysiologischen Anforderungen dürften zwar als für alle Sprachen gültige Randbedingungen gelten; in einigen Sprachen aber sind weitere Funktionen bekannt, denen Regelmäßigkeiten der Merkmalsausprägungen dienen. So erleichtern in den Finno-Ugrischen und den Türksprachen die Vokalharmonien die Segmentierung der Lautkette in Wörter. Solche und ähnliche Unterschiede in der Verwendung der sprachlichen Mittel, die mit den phonetischen Merkmalen zusammenhängen, können ein Grund zur Variabilität in den Korrelationsergebnissen sein.

Mit Hilfe statistischer Methoden läßt sich jedoch leicht entscheiden, ob die untersuchten Sprachen sich in dieser Hinsicht so sehr voneinander unterscheiden, daß eine Gleichheitshypothese abgelehnt werden muß, oder ob die Unterschiede vernachlässigbar sind. Das hier verwendete Verfahren besteht darin, die phonetischen Merkmale jeder untersuchten Srpache in der Rangreihenfolge der Merkmale auf Übereinstimmung unter allen untersuchten Sprachen gleichzeitig zu testen. Hierzu wird der Konkordanzko-

effizient von Kendall herangezogen, mit dessen Hilfe entschieden werden kann, ob Sprachen die Autokorrelationsstärke ihrer Merkmale auf gleiche Weise organisieren.

Zur Datengewinnung werden Textkorpora verschiedener Sprachen phonetisch transkribiert und Merkmalsmatrizes für die Segmentinventare aufgestellt. Beides wird sodann auf Datenträger zur maschinellen Auswertung kodiert. Die Berechnung der Übergangswahrscheinlichkeitsmatrizen der Markovketten und der Autokorrelationstabellen geschieht mit Hilfe des Programms Automark (vgl. Köhler 1983).

3. Zu Material und Materialaufbereitung

Die Hypothesen zur Sprachunabhängigkeit gewisser prosodischer Konstruktionsprinzipien wurden an Textkorpora zehn verschiedener Sprachen bzw. Sprachstufen überprüft. Das Sprachmaterial wurde folgenden Quellen entnommen, und - soweit keine Transkripte vorlagen - transkribiert:

Niederländisch

Robert E. Howard / L. Sprague de Camp / Lin Carter,
Conan Utrecht / Antwerpen 1976
Kapitel "De toren von de olifant" (S. 24 f)

Dane County Kölsch

McGraw, Peter A., Dane County Kölsch.
(= PHONAI. Lautbibliothek der europäischen
Sprachen und Mundarten. Band 21).
Tübingen 1979, S. 70 ff

Modernes Englisch

Lewis, J. Windsor, People Speaking.Phonetic
Readings in Current English. Oxford 1977. S.33 ff

Alt-, Mittel- und Frühneuenglisch

Finnie, W. Bruce, The Stages of English
Boston 1972, S. 36, 84, 96

Afrikaans

Roux, Justus (University of Stellenbosch)
Spezielles Transkript für diese Untersuchung

Japanisch

Ikuji Ehon, Shitagiri suzume. Tokio 1971

Französisch

aus 'Le sourire qui mord', Bulletin de liasion
No. 4, Paris o. J.

Türkisch

aus Langenscheidt, Praktisches Lehrbuch Türkisch
S. 318, (Ausschnitt aus Y.V. Yolcusen "für Reisende
im Schlafwagen")

Die Auswahl enthält sowohl typologisch sehr verschiedene Sprachen, als auch vier Sprachstufen ein und derselben Sprache. Dies soll es ermöglichen, die Sprachunabhängigkeit der phonetischen Konstruktionsprinzipien zu überprüfen und darüber hinaus möglicherweise Indizien für eine diachronische Gesetzmäßigkeit in dieser Hinsicht zu entdecken. Der letzte Aspekt bleibt einer späteren Studie vorbehalten.

Aus Platzgründen sind die Textdaten und Auswertungen nicht abgedruckt, können aber auf Anfrage von den Autoren zur Verfügung gestellt werden.

Länge der Texte:

- Modern English:
  699 Wörter
  Mittlere Wortlänge 2.9485

- Early Modern English:
  616 Wörter
  Mittlere Wortlänge 2.5065

- Middle English:
  599 Wörter
  Mittlere Wortlänge 2.971

- Old English:
  497 Wörter
  Mittlere Wortlänge 3.5875

- Französisch:
  945 Wörter
  Mittlere Wortlänge 2.7143

- Afrikaans:
  1205 Wörter
  Mittlere Wortlänge 3.0830

- Niederländisch:
  571 Wörter
  Mittlere Wortlänge 3.5954

- Niederländisch 1:
  463 Wörter
  Mittlere Wortlänge 4.3132

- Türkisch
  200 Wörter
  Mittlere Wortlänge 5.3000

- D.C. Kölsch
  959 Wörter
  Mittlere Wortlänge 3.0667

- Japanisch
  649 Wörter
  Mittlere Wortlänge 3.9060

Soweit die Texte nicht in transkribierter Form vorlagen, wurden sie transkribiert. Dabei wurde in allen Texten /tʃ / und /ʤ/ monophonematisch bewertet, sowie das Schwa teilweise als Phonem behandelt. Für Niederländisch wurden die Diphtonge sowohl monophonematisch als auch biphonematisch transkribiert, was sich jedoch in der Auswertung als unnötig herausstellte. Hier ist die monophonematische Version zugrundegelegt. Anschließend wurden die Transkriptionen in das Zeicheninventar des Rechners umkodiert.

Für jede der untersuchten Sprachen wurde eine Phonem-Merkmals-Matrix erstellt, in der die Ausprägungen der konstitutiven Merkmale jedes Phonems angegeben sind. Die Auswahl der verwendeten Merkmale beruht auf der phonetischen Analyse von Jakobson, Fant, Halle und von Chomsky/Halle. Deren Analyse wurde anderen Ansätzen vor allem aus folgenden Gründen vorgezogen. Zunächst versuchen Jakobson/Fant/Halle (1952) mit ihren 13 Merkmalen alle möglichen Kontraste von Sprachen zu erfassen und suchen somit lautliche Universalien. Zweitens versuchen Chomsky/Halle (1968) nicht nur die phonologischen Kontraste zu erfassen, sondern versuchen mit ihrer Revision der distinktiven Merkmale zugleich die phonetische Substanz der durch phonologische Regeln abgeleiteten Segmente zu beschreiben. Sie verbinden also eine funktionale Lautanalyse mit einer materialen. Drittens definieren Chomsky/Halle (1968) die distinktiven Merkmale primär artikulatorisch. Dies ist für die vorliegende Untersuchung insofern von Vorteil, als eventuelle Befunde möglicherweise artikulatorisch erklärt werden können, wie zum Beispiel mit dem Prinzip des geringsten Aufwandes in der Artikulationsbasis des Sprechers. Ein weiterer Vorzug des Ansatzes von Jakobson, Fant, Halle und Chomsky/Halle liegt in der

Tatsache, daß viele Einzelanalysen diesen Ansatz benutzen, was eine eventuelle Erweiterung der Datenbasis erleichtern würde.

Die verwendeten Merkmale sind: vocalic, consonantal, high, low, back, anterior, coronal, tense, voiced, continuant, sonorant, nasal, strident, length, glide (für Diphtonge bei monophonematischer Wertung), stress (zur Identifizierung des stets unbetonten Schwa), sowie nasalized als zusätzliches relevantes Merkmal im Japanischen.

Sodann wurde das Inventar aller in den Transkriptionen vorkommenden Laute in einer Lautinventarstabelle zusammengefaßt. Jedem Laut dieser Tabelle wurde sodann aufgrund der Merkmalsmatrix seine Merkmalskombination in +/- Notation zugeordnet.

4. *Analyse und Resultate*

Anhand der Textkorpora und der Phonem-Merkmals-Matrizen wurden für jede der 10 Sprachen folgende Analysen durchgeführt:

1. Absolute und relative Häufigkeiten der Phoneme
2. Für jedes Merkmal des Phonemystems:

   a) Häufigkeiten und Wahrscheinlichkeiten der Merkmalskombinationen bis zu Clustern 5. Ordnung.
   b) Aus der empirischen Wahrscheinlichkeitsmatrix 1. Ordnung die theoretischen Übergangswahrscheinlichkeiten bis zur Verschiebungsweite 9.
   c) Die Ordnung der Markovkette (d.h. die Reichweite von Merkmalsharmonie oder -disharmonie).
   d) Die Autokorrelationen für die Merkmalsausprägungen bis zur Verschiebungsweite 9.
   e) Der Korrelationssummenkoeffizient.

Tabelle 1a

Old English

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	3.	2.	1,387
cons	1.	3.	1,581
sono	6.	2.	1,355
high	5.	1.	1,372
low	2.	2.	1,410
back	15.	0.	1,163
ante	8.	3.	1,333
coro	7.	2.	1,346
roun	14.	0.	1,198
tens	10.	1.	1,315
voic	13.	3.	1,207
cont	12.	2.	1,208
nasa	4.	1.	1,379
stri	11.	0.	1,221
leng	9.	1.	1,318
stre	16.	0.	1,153

Tabelle 1 b

Middle English

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	2.	2.	1,685
cons	1.	4.+	1,765
cono	3.	2.	1,679
high	14.	0.	1,185
low	6.	2.	1,442
back	7.	0.	1,368
ante	4.	2.	1,574
coro	5.	2.	1,475
roun	16.	0.	1,129
tens	8.	2.	1,314
voic	11.	2.	1,270
cont	13.	2.	1,232
nasa	9.	1.	1,306
stri	12.	0.	1,256
leng	10.	1.	1,304
stre	15.	0.	1,168

Tabelle 1 c

Early Modern English

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	2.	3.	1,597
cons	1.	2.	1,726
sono	4.	2.	1,475
high	8.	3.	1,379
low	10.	1.	1,323
back	12.	0.	1,267
ante	5.	2.	1,453
coro	6.	3.	1,436
roun	13.	1.	1,252
tens	16.	0.	1,183
voic	7.	1.	1,413
cont	3.	1.	1,487
nasa	14.	1.	1,235
stri	9.	2.	1,335
leng	15.	1.	1,194
stre	11.	1.	1,291

Tabelle 1 d

Modern English

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	2.	2.	1,695
cons	1.	3.	1,698
sono	3.	2.	1,417
high	7.	1.	1,315
low	14.	1.	1,166
back	17.	0.	1,139
ante	5.	2.	1,392
coro	4.	2.	1,398
roun	15.	0.	1,150
tens	16.	0.	1,142
voic	9.	0.	1,261
cont	12.	1.	1,182
nasa	12.	1.	1,182
stri	10.	1.	1,248
leng	6.	1.	1,320
stre	8.	1.	1,303
glid	11.	1.	1,223

Tabelle 1 e

Nederlands (1, Diphthonge monophonematisch)

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	2.	3.	1,689
cons	1.	2.	2,364
sono	5.	0.	1,431
high	16.	1.	1,147
low	14.	1.	1,210
ante	5.	2.	1,431
coro	3.	4.+	1,552
roun	7.	2.	1,388
tens	13.	0.	1,237
voic	9.	0.	1,357
cont	11.	0.	1,300
nasa	8.	1.	1,367
stri	12.	2.	1,275
leng	10.	1.	1,308
back	15.	0.	1,185
stre	4.	2.	1,494
glid	17.	0.	1,117

Tabelle 1 f

Nederlands (2, Diphthonge bi-phonematisch)

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	2.	3.	1,669
cons	1.	4.+	2,105
sono	5.	0.	1,428
high	16.	0.	1,136
low	15.	1.	1,184
ante	6.	2.	1,391
coro	3.	4.+	1,485
roun	8.	2.	1,360
tens	13.	0.	1,225
voic	9.	0.	1,358
cont	12.	0.	1,287
nasa	7.	1.	1,375
stri	11.	2.	1,295
leng	10.	2.	1,339
back	14.	0.	1,188
stre	4.	2.	1,465

Tabelle 1 g

Afrikaans

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	2.	4.+	1,545
cons	1.	2.	2,093
sono	5.	4.+	1,407
high	4.	3.	1,243
low	8.	2.	1,297
back	15.	0.	1,166
ante	4.	4.+	1,515
coro	3.	4.+	1,543
roun	12.	2.	1,202
tens	10.	4.+	1,229
voic	6.	0.	1,366
cont	16.	1.	1,150
nasa	14.	2.	1,185
stri	13.	1.	1,194
leng	11.	2.	1,203
stre	7.	1.	1,345
glid	17.	1.	1,141

Tabelle 1 h

Dane County Koelsch

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	2.	3.	1,624
cons	1.	3.	1,835
sono	13.	0.	1,189
high	14.	1.	1,188
low	7.	2.	1,312
back	4.	3.	1,401
ante	6.	2.	1,357
coro	5.	4.	1,362
roun	15.	0.	1,183
tens	17.	0.	1,081
voic	3.	3.	1,483
cont	8.	3.	1,289
nasa	10.	2.	1,262
stri	11.	2.	1,254
leng	12.	2.	1,247
stre	9.	3.	1,272
glid	16.	0.	1,122

Tabelle 1 i

Japanisch

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	1.	4.+	3,518
cons	2.	4.+	3,072
sono	3.	4.+	2,271
high	13.	0.	1,285
low	4.	4.+	2,090
back	12.	4.+	1,314
ante	6.	3.	1,794
coro	5.	4.+	1,985
roun	7.	2.	1,603
tens	15.	0.	1,164
voic	8.	4.+	1,495
cont	9.	3.	1,482
nasa	11.	1.	1,350
nasl	16.	0.	1,130
stri	10.	2.	1,460
leng	14.	0.	1,201

Tabelle 1 j

Französisch

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	4.	1.	1,521
cons	1.	2.	2,301
sono	6.	1.	1,355
high	16.	0.	1,096
low	5.	1.	1,378
back	7.	1.	1,273
ante	3.	2.	1,591
coro	2.	3.	1,792
roun	9.	2.	1,225
tens	11.	1.	1,212
voic	8.	1.	1,237
cont	12.	1.	1,201
nasa	13.	0.	1,188
stri	10.	1.	1,216
leng	15.	1.	1,179
stre	14.	1.	1,182

Tabelle 1 k

Türkisch

Merkmal	Rang	Ordnung	Summenkoeffiz.
voca	2.	2.	1,802
cons	1.	2.	2,033
high	11.	0.	1,329
low	12.	1.	1,320
back	3.	3.	1,567
ante	6.	2.	1,443
coro	5.	4.+	1,528
roun	4.	2.	1,559
tens	14.	0.	1,214
voic	8.	2.	1,403
cont	10.	2.	1,393
sono	9.	2.	1,396
nasa	12.	1.	1,320
stri	15.	0.	1,172
leng	7.	0.	1,422
stre	16.	0.	1,057

Die Ordnung der Kette ist ein Maß für die Reichweite der Harmonie bzw. der Disharmonie zwischen Merkmalen. Dabei bedeutet der Wert O die Unabhängigkeit der Ausprägung der Merkmale in allen Positionen, d.h. die Zufälligkeit in allen Positionen. Der Wert 1 bedeutet, daß das Zutreffen eines Merkmals Einfluß auf die nächstfolgende bzw. die vorhergehende Position hat. Beispiel: Ordnung der Kette für das Merkmal nasal 1 bedeutet, daß die Übergangswahrscheinlichkeit zu einer anderen Position der Kette aufgrund von Merkmalsharmonie nur auf die nächste Stelle signifikante Auswirkungen hat. Wert 2 bedeutet die Reichweite beträgt 2 Stellen. Um vorherzubestimmen, ob an einer Stelle ein Merkmal stehen bzw. nicht stehen wird, muß man zwei Positionen vor bzw. nach der jeweiligen Position kennen.

Der Summenkoeffizient ist ein Maß für die Stärke der Harmonie bzw. Disharmonie. Dabei ist der Wert 1 ein minimaler Wert, da eine Position immer mit sich selbst übereinstimmen muß. Der theoretisch maximale Wert liegt bei n Schritten bei n. Interpretation: Der Summenkoeffizient für Französisch von 1,521 und Japanisch 3,518 besagt: Im Japanischen ist die sequentielle Merkmalsabhängigkeit für "vokalisch" sehr viel höher als im Französischen. Dies

bedeutet, daß das Vorhandensein bzw. Nichtvorhandensein von "vokalisch" an einer Stelle auf andere Stellen einen großen Einfluß hat. Die Ordnung der Kette (4) gibt die Reichweite der sequentiellen Abhängigkeit an. Im Französischen genügt eine Stelle, um die erreichbare Voraussagepräzision zu erlangen; die Berücksichtigung weiterer Stellen erhöht die Vorhersagbarkeit nicht. Im Japanischen dagegen erreicht man durch Hinzunahme von weiteren Stellen eine größere Präzision der Voraussage, d.h. es ergibt sich eine Zunahme der Regelhaftigkeit der feature-Abfolge vokalisch im Japanischen.

Die Abbildungen 3a - 3c zeigen die zeitliche Entwicklung der Korrelationssummenkoeffizienten der Merkmale der vier untersuchten Sprachstufen des Englischen.

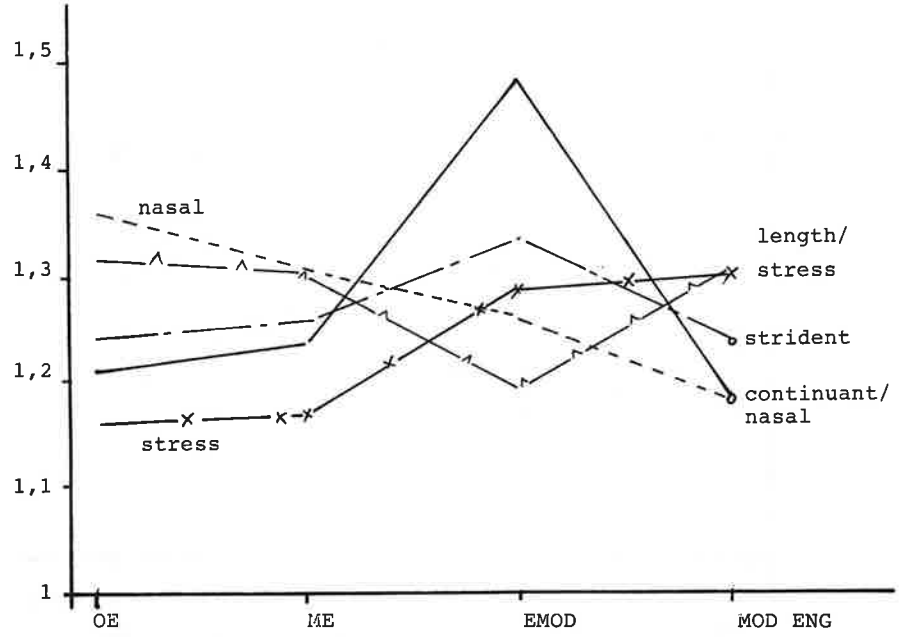


Abb. 3a: Σ_ϕ-Koeffizient der vier Sprachstufen des Englischen

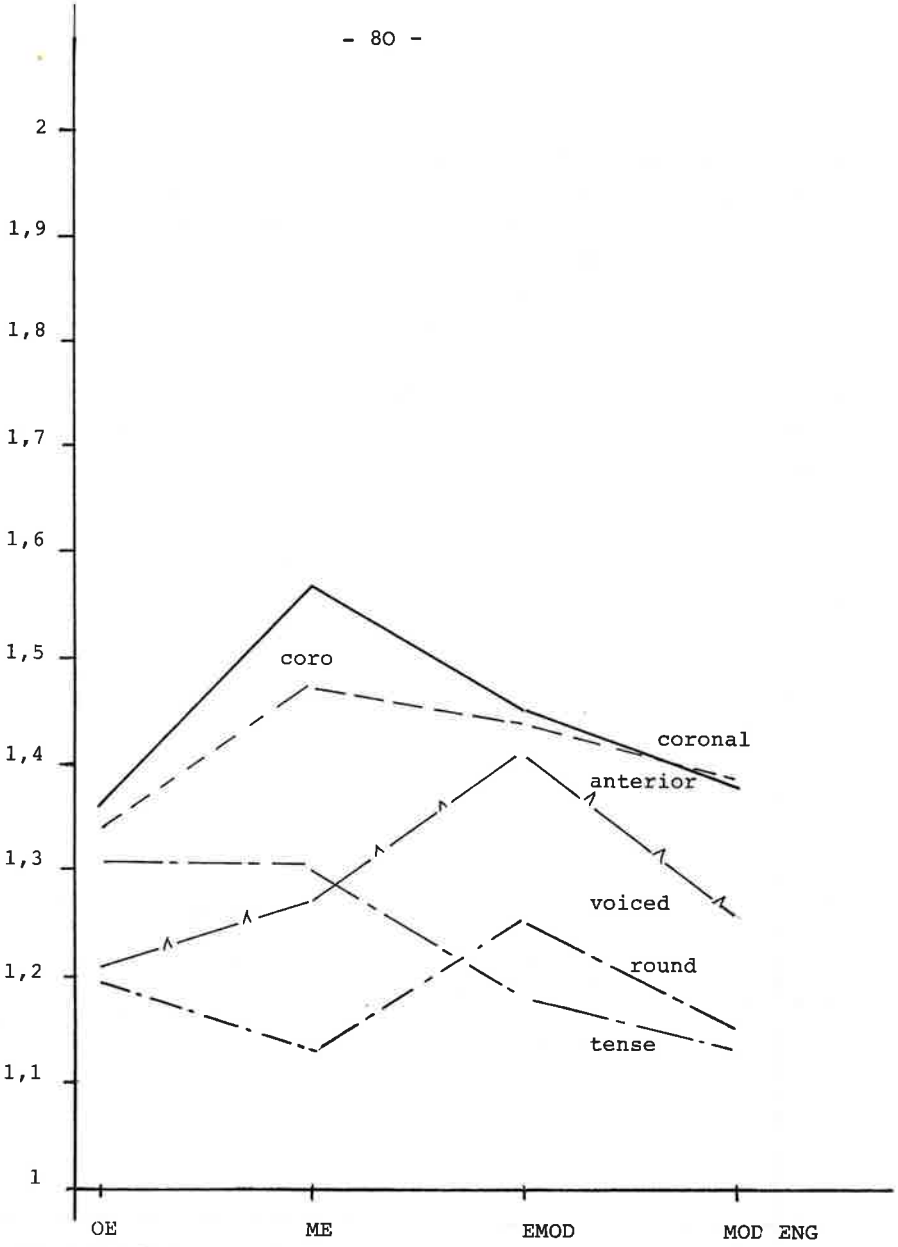


Abb. 3b Σ_ϕ-Koeffizient der vier Sprachstufen des Englischen

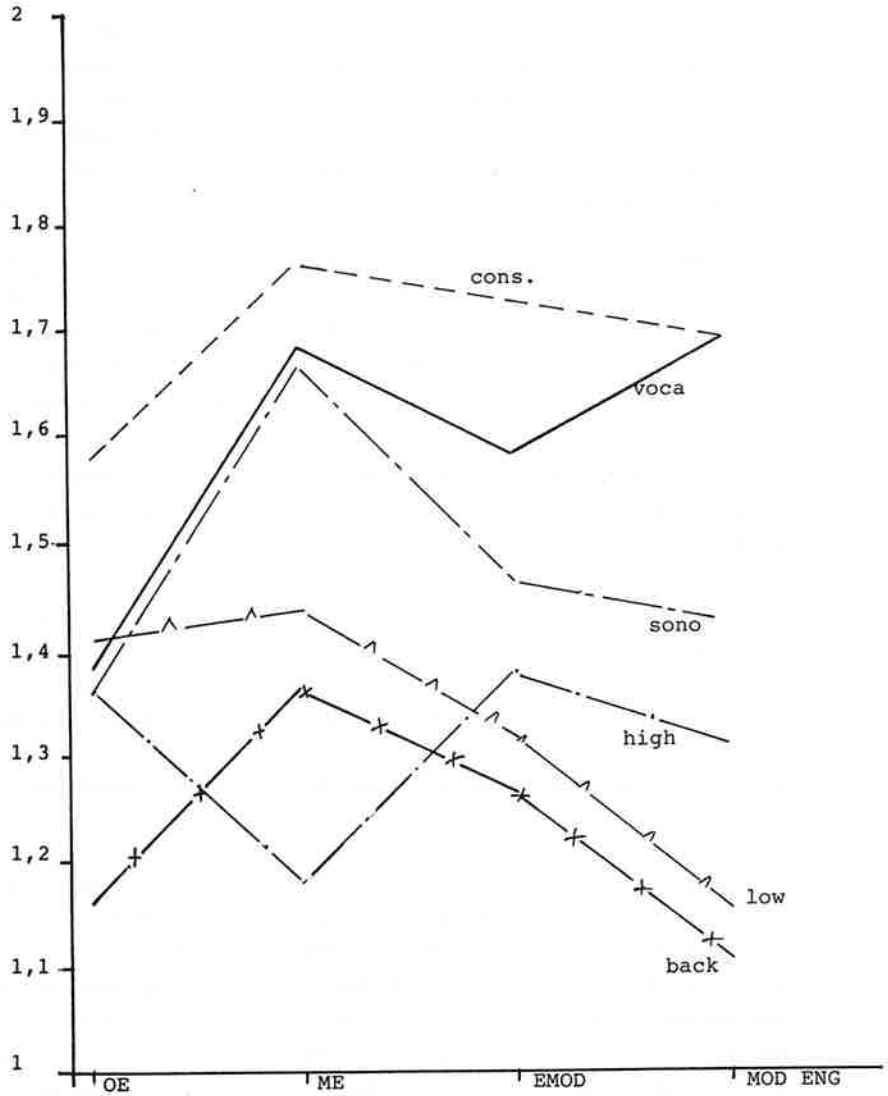


Abb. 3c Σ_ϕ-Koeffizient der vier Sprachstufen des Englischen

In den Abb. 4a-4k (s. Anhang) ist der Summenkoeffizient graphisch dargestellt, und zwar in der Rangreihenfolge ihrer Ausprägungsstärke. Dies dient zur Vorbereitung der Untersuchung von Übereinstimmungen unter den verschiedenen untersuchten Sprachen. Da der Summenkoeffizient relativ und vom Zahlenwert her nicht ohne weiteres in Intervallen angehbar ist, wird die Rangreihenfolge der relativen Stärke ermittelt. Die features-Reihenfolge in den verschiedenen Sprachen ist unterschiedlich. Vergleichen wir das Japanische mit dem modernen Englisch, so stellen wir fest, daß im Japanischen offensichtlich sehr viel stärkere Merkmalsharmonie bzw. Disharmonie herrscht, während im modernen Englischen die Aufeinanderfolge freier ist. Das heißt, im Japanischen kann man weniger Silben- und Morphemtypen erwarten. Selbst wenn man über die Sprache bzw. Sprachstruktur nichts oder nicht sehr viel weiß, kann man einige Kenntnisse aus dem Säulendiagramm ableiten und zwar,

- daß aufgrund der höheren Ausprägung der Koeffizienten im Japanischen eine größere Restriktivität der Phonotaktik vorhanden sein muß,

- daß man eine größere Einheitlichkeit von Silben und Morphemstruktur erwarten kann.

Zur Reihenfolge der Summenkoeffizienten ist folgendes zu sagen:

- In einigen Sprachen unterscheidet sich die Ausprägung der Merkmale sehr wenig.

- Die Reihenfolge der Stärken der Merkmale unter den Sprachen ist sehr ähnlich. Die Tatsache, daß einige Ausprägungen sich kaum unterscheiden, führt dazu, daß Meßfehler einen relativ großen Einfluß haben können. Nimmt man an, daß es ein natürliches Bauprinzip für Sprachen gibt, so könnte es sein, daß durch andere Einflüsse (wie eben genannte Meßfehler) dieses Bauprinzip verdeckt wird. Es erhebt sich darum die Frage, ob die Unterschiede so klein sind, daß sie vernachlässigt werden können, so daß die Unterschiede der sequentiellen Abhängigkeit nicht signifikant sind.

Die Hypothese, daß die verschiedenen phonetischen Merkmale bei der Konstruktion der realen sprachlichen Zeichen eine sprachunabhängige Rolle spielen, wollen wir auf folgende Weise testen. Wir ordnen die Merkmale jeder Sprache nach der Reihenfolge der Stärke ihres jeweiligen Korrelationssummenkoeffizienten an. Die Stelle eines Merkmals in dieser Liste kennzeichnet den Rang des jeweiligen Merkmals in der betroffenen Sprache. Die nachstehende Tabelle zeigt die Rangreihenfolge der Summenkoeffzienten aller Merkmale der 10 untersuchten Sprachen.

Tabelle 2: Rangreihenfolge der Summenkoeffzienten incl. gebundene Ränge

	1	2	3	4	5	6	7	8	9	10
Merkmal	OE	ME	EM	MoE	Fr	Af	Ja	Ne	Tü	D.C.K.
voca	3	2	2	2	4	2	1	2	2	2
cons	1	1	1	1	1	1	2	1	1	1
sono	6	3	4	3	6	5	3	5	9	13
high	5	14	8	7	16	9	13	16	11	14
low	2	6	10	14	5	8	4	14	12,5	7
back	15	7	12	17	7	15	12	15	3	4
ante	8	4	5	5	3	4	6	5	6	6
coro	7	5	6	4	2	3	5	3	5	5
roun	14	16	13	15	9	12	7	7	4	15
tens	10	8	16	16	11	10	15	13	14	17
voic	13	11	7	9	8	6	8	9	8	3
cont	12	13	3	12,5	12	16	9	11	10	8
nasa	4	9	14	12,5	13	14	11	8	12,5	10
stri	11	12	9	10	10	13	10	12	15	11
leng	9	10	15	6	15	11	14	10	7	12
stre	16	15	11	8	14	7	17,5	4	16	9
glid	17,5	17,5	17,5	11	17,5	17	17,5	17	17,5	16
nasl	17,5	17,5	17,5	18	17,5	18	16	18	17,5	18

In die Sprache der Statistik übersetzt lautet unsere Hypothese nun folgendermaßen: Die Rangreihenfolgen der Korrelationssummenkoeffizienten der verschiedenen Sprachen unterscheiden sich nicht signifikant. Um dies zu testen, werden die Rangfolgen aller Sprachpaare miteinander korreliert. Die daraus resultierenden Korrelationskoeffizienten zeigt die Tabelle 3.

Jede Sprache wird mit jeder anderen Sprache in bezug auf ihre Rangziffernstärke korreliert. Diese liegen theoretisch zwischen -1 und +1. In unserer Tabelle 3 liegen alle Werte relativ hoch gegen +1. Sämtliche Werte in der Tabelle sind auf dem 0.05-Niveau signifikant;alle bis auf 5 sogar auf dem Niveau 0.01. Dies bedeutet, daß die Anordnung nicht als zufällig angesehen werden kann. Um eine Aussage darüber zu machen, ob dieses Ergebnis auch gültig ist, wenn man alle Sprachen zusammen betrachtet, wird der Konkordanzkoeffizient verwendet.

Unsere obige Hypothese darf also, aufgrund unserer Untersuchungen, nicht abgelehnt werden.

Um den Kendallschen Konkordanzkoeffizienten (vgl. z.B. Siegel 1956) zu bestimmen, berechnen wir zunächst den Mittelwert der Korrelationskoeffzienten aller Sprachpaare:

$$\bar{r} = \frac{LW-1}{L-1}$$

Hieraus ergibt sich der Konkordanzkoeffizient W mit L = 10 (Anzahl der Sprachen) als

$$W = 68.51.$$

Die Variable $W \cdot L \cdot (k - 1)$ hat eine asymptomische χ^2 -Verteilung mit k - 1 Freiheitsgraden, wobei k die Anzahl der Merkmale in der Rangtabelle ist. Wir erhalten somit:

$$\chi^2 = 116.46$$

Da der beobachtete Wert von χ^2 größer ist als der theoretische ($\chi^2_{17;0.05} = 3.72$), ist die Nullhypothese, nach der die zehn Sprachen verschiedene Rangfolgen haben,also abzulehnen. Die zehn Variablen stammen aus derselben Grundgesamtheit, und wir sehen die Unterschiede in den Rangplätzen der Merkmale als zufällig an. In bezug auf das vorliegende Datenmaterial sind wir berechtigt, von einem gemeinsamen Bauprinzip der Sprachen auszugehen. Die beobachtbaren Unterschiede können auf Meßfehler etc. zurückgeführt werden, d.h. es bestätigt sich die Annahme, daß die konstitutiven Merkmale der Phonemsysteme in bezug auf ihre syntagmatischen Eigenschaften in den jeweiligen Sprachen gleich stark ausgeprägt sind. Dies legt die Vermutung nahe, daß den Merkmalen bestimmte sprachunabhängige Funktionen zukommen.

Tabelle 3: Rangkorrelationsanalyse

R	OE	ME	EM	MoE	Fr	Afr	Jap	Ned	Tü	DoK
OE	-	0,703*	0,608*	0,605*	0,607*	0,645*	0,675*	0,555	0,541	0,540
ME	-	-	0,645*	0,612*	0,750*	0,695*	0,707*	0,633*	0,649*	0,653*
EM	-	-	-	0,678*	0,680*	0,678*	0,722*	0,657*	0,624*	0,681*
MoE	-	-	-	-	0,562	0,725*	0,592*	0,695*	0,556	0,567
Fr	-	-	-	-	-	0,702*	0,761*	0,663*	0,689*	0,695*
Afr	-	-	-	-	-	-	0,669*	0,725*	0,608*	0,629*
Jap	-	-	-	-	-	-	-	0,646*	0,675*	0,640*
Ned	-	-	-	-	-	-	-	-	0,619*	0,610*
Tü	-	-	-	-	-	-	-	-	-	0,606*
DoK	-	-	-	-	-	-	-	-	-	-

5. Ausblick

Die Untersuchung der phonischen Realisation der Phonemsysteme verschiedener Sprachen ergab zwar für jede dieser Sprachen individuelle Charakteristika; die sich daraus ergebenden Unterschiede erweisen sich als nicht signifikant. Dieses Ergebnis stützt die Vermutung, daß die phonische Realisierung beliebiger Phonemsysteme außersprachlichen Einflüssen unterliegt. Weiter ist zu vermuten, daß die beobachtbaren Gemeinsamkeiten durch Gesetze formuliert werden könnten. Diese außersprachlichen Einflüsse müßten in einem umfassenden Modell abgebildet werden, das als Komponenten mindestens die betrachteten sprachlichen Teilsysteme (hier vor allem die phonischen Teilsysteme), sowie die Anforderungen von Sprecher, Kanal und Hörer an die phonetische Substanz umfaßt. Aufgrund der Vorüberlegungen in Kapitel 1 liegt es nahe, dieses auf dem Wege der funktionalanalytischen Modellierung zu versuchen (vgl. Köhler/ Altmann 1983).

Unser Kenntnisstand in bezug auf die zu berücksichtigenden Systemgrößen reicht zwar zu einer Durchführung einer entsprechenden Modellkonstruktion nicht aus, aber wir können bereits eine Reihe außersystemischer Anforderungen (Bedürfnisse) formulieren:

- Minimaler Produktionsaufwand
- Minimierung des Gedächtnisaufwandes
- Minimierung des Wahrnehmungsaufwandes
- Ausreichende Redundanz zur Übertragungssicherung
- Ausreichendes unausgeschöpftes Potential zur Anpassung an neue Anforderungen

Diesen und weiteren Bedürfnissen entsprechen im Modell Prozesse (vgl. die Zipfschen Unifikations- und Diversifikationskräfte). Ein gegebener Systemzustand ist als Resultat der Auswirkungen dieser Prozesse aufzufassen. Diese Bedürfnisse sind teils komplementär, teils antagonistisch: das Bedürfnis nach Übertragungssicherung (Redundanz) und das nach unausgeschöpftem Reservepotential wirken gleichsinnig, während das Bedürfnis nach minimalem Produktionsaufwand und minimalem Diskriminationsaufwand antagonistisch wirken. Im letzteren Fall kommt die Ausprägung einer Systemgröße durch Ausbildung eines optimalen Kompromisses zwischen widerläufigen Anforderungen zustande.

Sobald die wesentlichen Bedürfnisse und Systemeigenschaften bekannt sind, läßt sich ein funktionalanalytisches Modell für die phonische Ebene natürlicher Sprachen in Analogie zu Köhler/ Altmann (1983) erstellen, aus dem dann Sprachgesetze für die phonische Ebene und ihre Randbedingugnen als Konsequenz ableitbar sind.

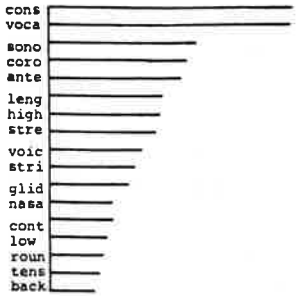


Abb. 4a: Modern English
Rangreihenfolge der Summen-
koeffizienten, 17 features

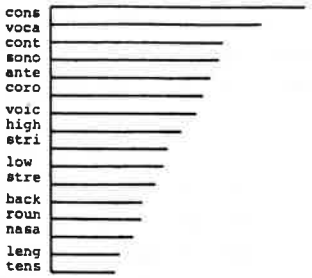


Abb. 4b: Early Modern English
Rangreihenfolge der Summenko-
effizienten, 16 features

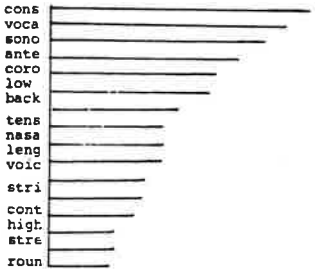


Abb. 4c: Middle English
Rangreihenfolge der Summen-
koeffizienten, 16 features

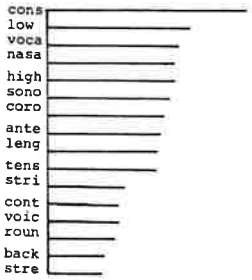


Abb. 4d: Old English
Rangreihenfolge der Summenko-
effizienten, 16 features

Abb. 4e: Japanisch
Rangreihenfolge der Summenkoeffizeinten,
16 features

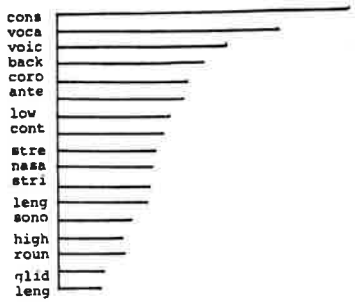


Abb. 4f: Dane County Kölsch
Rangreihenfolge der Summen-
koeffzienten, 17 features

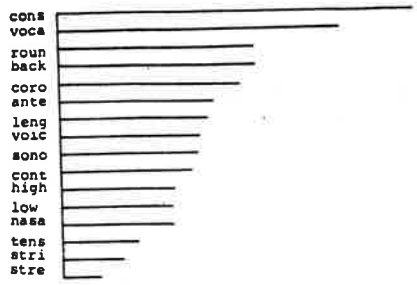


Abb. 4g: Türkisch
Rangreihenfolge der Summenko-
effizienten, 16 features

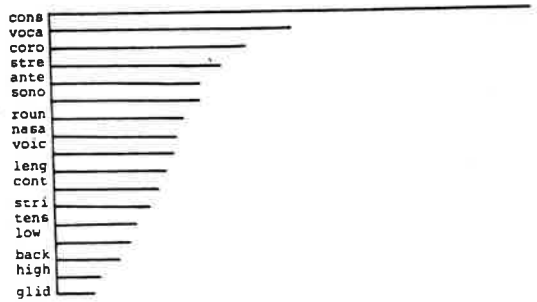


Abb. 4h: Nederlands 1
(monophonematische) Rangreihenfolge
der Summenkoeffizienten, 17 features

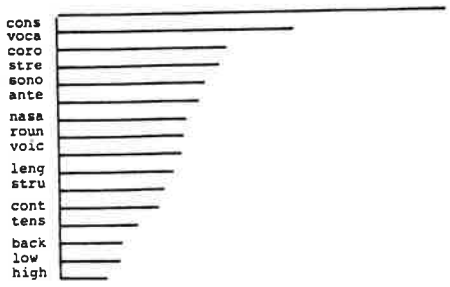


Abb. 4i: Nederlands 2
Rangreihenfolge der Summenkoeffizienten
16 features

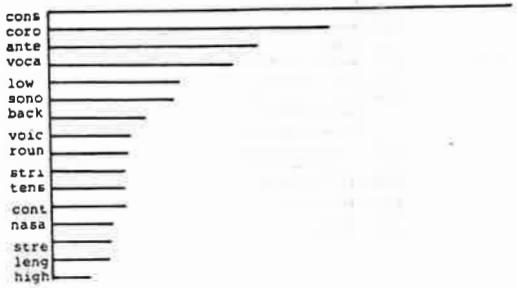


Abb. 4j: Französisch
Rangreihenfolge der Summenkoeffizienten, 16 features

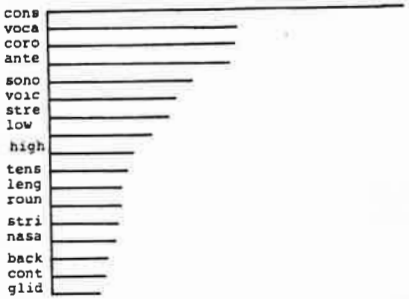


Abb. 4k: Afrikaans
Rangreihenfolge der Summenkoeffizienten, 17 features

Bibliographie

Altmann, G., Zur Funktionalanalyse in der Linguistik. In: Esser, J. & Hübler, A. (Eds.), Forms and Functions: Papers in General, English, and Applied Linguistics: Presented to Villiam Fried on the Occasion of his 65. Birthday. Tübingen: Narr. 1981 (Tübinger Beiträge zur Linguistik, Bd. 149), 25-32

Altmann, G. & Lehfeldt, W., Einführung in die Quantitative Phonologie. Bochum: Brockmeyer, 1980, (Quantitative Linguistik 7)

Chomsky, N. & Halle, M., The Sound Patterns of English. New York: Harper & Row, 1968

Crystal, D., Prosodic Systems and Intonation in English. Cambridge: Cambridge University Press, 1976, 61-62

Firth, J.R., Sounds and Prosodies. In: Transactions of the Philosophical Society. London, 1949, 127-152

Hockett, C.F., Logical Considerations in the Study of Animal Communication. In: Lanyon, W.E. & Tavolga, W.N. (Eds.) Sounds and Communication. Washington, DC.: American Institute of Biological Sciences, 1960a, 392-430

Hockett, C.F., The Origin of Speech. In: Scientific American, 1960b, 203, 89-96

Hockett, C.F. & Altmann, S., A Note on Design Features. In: Sebeok, T.A. (Ed.) Animal Communication. Bloomington, Ind.: University of Indiana Press, 1968, 61-72

Jakobson, R., Fant, C.G.M. & Halle, M., Preliminaries to Speech Analysis. The Distinctive Features and their Correlates. Cambridge/Mass.: MIT Press, 1952

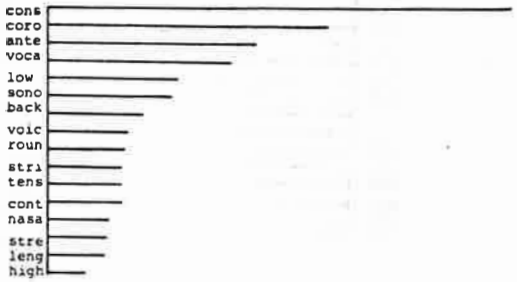


Abb. 4j: Französisch
Rangreihenfolge der Summenkoeffizienten, 16 features

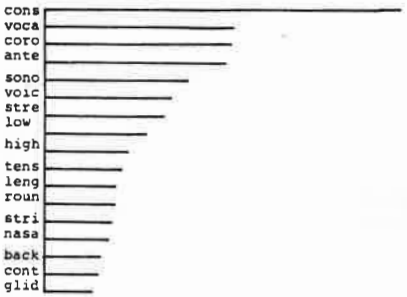


Abb. 4k: Afrikaans
Rangreihenfolge der Summenkoeffizienten, 17 features

Bibliographie

Altmann, G., Zur Funktionalanalyse in der Linguistik. In: Esser, J. & Hübler, A. (Eds.), Forms and Functions: Papers in General, English, and Applied Linguistics: Presented to Villiam Fried on the Occasion of his 65. Birthday. Tübingen: Narr. 1981 (Tübinger Beiträge zur Linguistik, Bd. 149), 25-32

Altmann, G. & Lehfeldt, W., Einführung in die Quantitative Phonologie. Bochum: Brockmeyer, 1980, (Quantitative Linguistik 7)

Chomsky, N. & Halle, M., The Sound Patterns of English. New York: Harper & Row, 1968

Crystal, D., Prosodic Systems and Intonation in English. Cambridge: Cambridge University Press, 1976, 61-62

Firth, J.R., Sounds and Prosodies. In: Transactions of the Philosophical Society. London, 1949, 127-152

Hockett, C.F., Logical Considerations in the Study of Animal Communication. In: Lanyon, W.E. & Tavolga, W.N. (Eds.) Sounds and Communication. Washington, DC.: American Institute of Biological Sciences, 1960a, 392-430

Hockett, C.F., The Origin of Speech. In: Scientific American, 1960b, 203, 89-96

Hockett, C.F. & Altmann, S., A Note on Design Features. In: Sebeok, T.A. (Ed.) Animal Communication. Bloomington, Ind.: University of Indiana Press, 1968, 61-72

Jakobson, R., Fant, C.G.M. & Halle, M., Preliminaries to Speech Analysis. The Distinctive Features and their Correlates. Cambridge/Mass.: MIT Press, 1952

Köhler, R., Markov-Ketten und Autokorrelation in der Sprach- und Textanalyse. In: Köhler, R. & Boy, J. (Eds.) Glottometrika. Bochum: Brockmeyer, 1983, 5, 137-167

Köhler, R. & Altmann, G., Systemtheorie und Semiotik. In: Zeitschrift für Semiotik, 1983, 5/4, 424-431

Krampen, M., Ferdinand de Saussure und die Entwicklung der Semiologie. In: Krampen, M., Oehler, K., Posner, R., & von Uexküll, T., (Eds.) Die Welt als Zeichen. Berlin: Severin und Siedler, 1981, 99-142

Lyons, J., Semantics I, Cambridge: Cambridge University Press, 1977, 70-85

Newman, E.B., Computational Methods Useful in Analyzing Series of Binary Data. In: American Journal of Psychology, 1951a, 64, 252-262

Newman, E.B., The Pattern of Vowels and Consonants in Various Languages. In: American Journal of Psychology, 1951b, 64, 369-379

Newman, E.B., A New Methods for Analyzing Printed English. In: Journal of Experimental Psychology, 1952, 44, 114-125

Poyatos, F., Silence and Stillness: Toward a New Status of Non-Activity. In: Kodikas/Code, 1981, 3/1, 3-26

Siegel, S., Nonparametric Statistics for the Behavioral Sciences. New York: Mc Graw-Hill, 1956

Thorpe, W.E., The Comparison of Vocal Communication in Animals and Man. In: Hinde, R.A. (Ed.) Nonverbal Communication. Cambridge: Cambridge University Press, 1972, 27-47

Old English

	voca	cons	sono	high	low	back	ante	coro	roun	tens	voic	cont	nasa	stri	leng	stre
i:	+	-	+	+	-	-	-	-	-	+	+	+	-	-	+	+
ɪ	+	-	+	+	-	-	-	-	-	-	+	+	-	-	-	+
y:	+	-	+	+	-	-	-	-	+	+	+	+	-	-	+	+
y	+	-	+	+	-	-	-	-	+	-	+	+	-	-	-	+
u:	+	-	+	+	-	+	-	-	+	+	+	+	-	-	+	+
ʊ	+	-	+	+	-	+	-	-	+	-	+	+	-	-	-	+
e:	+	-	+	-	-	-	-	-	-	+	+	+	-	-	+	+
ɛ	-	-	+	-	-	-	-	-	-	-	+	+	-	-	-	+
o:	+	-	+	-	-	+	-	-	+	+	+	+	-	-	+	+
ɔ	+	-	+	-	+	+	-	-	+	-	-	-	-	-	-	+
æ:	+	-	+	-	+	-	-	-	-	+	+	+	-	-	+	+
ɑ:	+	-	+	-	+	+	-	-	-	+	+	+	-	-	+	+
ɑ	+	-	+	-	+	+	-	-	-	-	+	+	-	-	-	+
æ	+	-	+	-	+	-	-	-	-	-	+	+	-	-	-	+
ə	+	-	+	-	-	-	-	-	-	-	+	+	-	-	-	-
j	-	+	-	+	-	-	-	-	-	-	-	+	-	-	-	+
p	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-	+
b	-	+	-	-	-	-	+	-	-	-	+	-	-	-	-	+
t	-	+	-	-	-	-	+	+	-	+	-	-	-	-	-	+
d	-	+	-	-	-	-	+	+	-	-	+	-	-	-	-	+
g	-	+	-	+	-	+	-	-	-	-	+	-	-	-	-	+
k	-	+	-	+	-	+	-	-	-	+	-	-	-	-	-	+
f	-	+	-	-	-	-	+	-	-	+	-	+	-	+	-	+
v	-	+	-	-	-	-	+	-	-	-	+	+	-	+	-	+
θ	-	+	-	-	-	-	+	+	-	-	-	+	-	-	-	+
ð	-	+	-	-	-	-	+	+	-	-	+	+	-	-	-	+
s	-	+	-	-	-	-	+	+	-	+	-	+	-	+	-	+
z	-	+	-	-	-	-	+	+	-	-	+	+	-	+	-	+
ʃ	-	+	-	+	-	-	-	+	-	-	-	+	-	+	-	+
ʒ	-	+	-	+	-	-	-	+	-	-	+	+	-	+	-	+
h	-	+	+	-	-	-	-	-	-	+	-	+	-	-	-	+
m	-	+	+	-	-	-	+	-	-	-	+	+	+	-	-	+
n	-	+	+	-	-	-	+	+	-	-	+	+	+	-	-	+
ŋ	-	+	+	+	-	+	-	-	-	-	+	-	+	-	-	+
l	+	+	+	-	-	-	-	+	+	-	+	+	-	-	-	+
r	+	+	+	-	-	-	-	+	-	-	+	+	-	-	-	+
w	-	-	+	+	-	+	-	-	+	-	-	-	-	-	-	+
ʧ	-	+	-	+	-	-	-	+	-	-	-	-	-	+	-	+
ʤ	-	+	-	+	-	-	-	+	-	-	+	-	-	+	-	+
x	-	+	-	+	-	+	-	-	-	+	-	+	-	-	-	+
o	+	-	+	-	-	-	-	-	+	-	+	+	-	-	-	+

Middle English

	i:	ɪ	ɛ:	ɔ:	æ	ʊ	e:	ɛ	o:	ɔ	ə	a:	a	j	p	b	t	d	g	k	f	v	θ	ð	s	z	ʃ	ʒ	h	m	n	ŋ	l	r	w	ʧ	ʤ	e	u:	x
voca	+	+	+	+	+	+	+	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	+	-
cons	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	-	+
sono	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	-	-	+	+	-
high	+	+	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	+	+	-	-	-	-	-	+	+	-	-	-	+	-	-	+	+	+	-	+	+
low.	-	-	-	+	+	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
back	-	-	-	+	-	+	-	-	+	+	-	+	+	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	+	-	-	+	-	-	+	-	-	-	+	+
ante	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	+	+	+	+	+	+	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
coro	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	+	+	+	+	+	-	-	+	-	+	+	-	+	+	-	-	-	-	-	-
roun	-	-	-	+	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	+	-	-	-	-	-
tens	+	-	-	+	-	-	+	-	+	-	-	+	-	-	+	-	-	+	+	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	+
voic	+	+	+	+	+	+	+	+	+	+	-	+	+	-	-	+	+	-	-	+	-	+	-	+	-	+	-	+	+	+	+	+	-	-	+	+	+	-	-	+
cont	+	+	+	-	-	+	+	+	+	-	-	+	+	+	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	-	-	+	+	+
nasa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-
stri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	+	-	-	-	-	-	-	-	-	+	+	-	-	-
leng	+	-	+	+	-	-	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
stre	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Early Modern English

	i:	ɪ	ɜ	ɜ:	u:	ʊ	ɛ	o:	ɔ	ɛ:	a	æ	ɐ	j	p	b	t	d	g	k	f	v	θ	ð	s	z	ʃ	ʒ	h	m	n	ŋ	l	r	w	ʧ	ʤ	ɜ:	a:	æ:
voca	+	+	+	+	+	+	-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	+	+
cons	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	-	-
sono	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	-	-	+	+	+	+
high	+	+	-	-	+	+	-	-	-	-	-	-	-	+	-	-	-	-	+	+	-	-	-	-	-	+	+	-	-	-	+	-	-	+	+	+	-	-	-	-
low.	-	-	+	+	-	-	-	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
back	-	-	-	-	+	+	-	+	+	-	+	-	-	-	+	+	-	-	-	+	+	-	-	-	-	-	-	-	+	-	-	+	-	-	+	+	-	+	+	-
ante	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	+	+	+	+	+	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
coro	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	+	+	+	+	+	-	-	+	-	+	+	-	+	+	-	+	+	-	-	-
roun	-	-	+	+	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-
tens	+	-	+	+	+	-	-	+	-	-	-	-	-	+	-	+	-	-	+	+	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	+	+	-
voic	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	+	+	-	-	+	-	+	-	+	-	+	-	+	+	+	+	+	-	-	+	-	+	+	+	+
cont	+	+	-	-	+	+	+	+	-	+	+	-	-	+	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	-	-	-	+	+
nasa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-
stri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	+	-	-	-	-	-	-	-	-	+	+	-	-	-
leng	+	-	-	+	+	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
stre	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Modern English

	a	ʌ	e	ɜ	ə	æ	o	ɔ	ʊ	ɪ	ɨ	u:	j	eɪ	əʊ	ɔɪ	aʊ	aɪ	p	b	t	d	g	k	f	v	θ	ð	s	z	ʃ	ʒ	h	m	n	ŋ	l	r	w	ʧ	ʤ
voca	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
cons	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+
sono	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	-	-
high	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	+	+	-	-	-	+	-	-	+	+	+
low.	+	-	-	-	-	+	-	+	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
back	+	+	-	-	-	-	+	+	-	-	+	-	-	+	+	+	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-
ante	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	+	+	+	+	+	-	-	-	+	+	-	-	-	-	-	-	-
coro	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	+	+	+	+	-	-	+	-	+	+	-	+	+	-	+	+
roun	-	-	-	+	-	-	+	+	+	-	-	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-
tens	-	-	-	+	-	-	-	-	-	-	+	+	-	+	+	+	+	+	+	-	+	-	-	+	+	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-
voic	+	+	+	+	-	+	+	+	+	+	+	-	+	+	+	+	+	-	+	-	+	+	-	-	+	-	+	-	+	-	+	+	-	+	+	+	+	-	-	+
cont	+	-	+	+	-	-	+	-	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	-	+	+	-	-	-	-	-
nasa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-
stri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	-	-	-	-	-	-	-	+	+
leng	-	-	-	+	-	-	-	-	-	-	+	+	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
stre	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
glid	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Niederländisch (monophonematisch)

	aː	a	eː	e	ɛ	iː	ɪ	oː	ɔ	øː	uː	ʊ	y	yː	x	ʒ	ʃ	ŋ	z	ə	m	n	l	r	p	b	f	v	t	d	s	j	k	g	h	o	au	ɛɪ	œʏ
voca	+	+	+	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+	-	+
cons	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
sono	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
high	-	-	-	-	-	+	+	-	-	-	+	+	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-
low.	+	+	+	-	+	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
ante	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	+	-	-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
coro	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	+	-	-	+	+	+	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-
roun	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-
tens	+	-	+	-	-	+	-	+	-	+	+	-	-	+	+	-	-	-	-	-	-	-	-	+	-	+	-	+	-	+	-	+	-	+	-	-	-	-	+	-
voic	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	-	+	-	+	+	-	+	-	+	-	+	-	+	-	+	-	+	+	+	+	+	+
cont	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+	-	+	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+	+	+	+
nasa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
stri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	+	+	-	-	+	-	-	-	+	-	-	-	-	-	-	-
leng	+	-	+	-	-	+	-	+	-	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
back	+	+	-	-	-	-	-	+	+	-	+	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-
stre	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
glid	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+

Niederländisch (biphonematisch)

	aː	a	eː	e	ɛ	i	ɪ	oː	ɔ	øː	uː	ʊ	y	yː	x	ʒ	ʃ	ŋ	z	ə	m	n	l	r	p	b	f	v	t	d	s	j	k	g	h	o
voca	+	+	+	+	−	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	+	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	+
cons	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−
sono	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	+	−	−	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	+	+
high	−	−	−	−	−	+	+	−	−	−	+	+	+	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−
low.	+	+	+	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
ante	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	−	−	+	+	+	+	+	+	+	−	−	−	−	−	−
coro	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	−	−	+	+	+	−	−	−	−	+	+	+	−	−	−	−	−
roun	−	−	−	−	−	−	−	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	+
tens	+	−	+	−	−	+	−	+	−	+	+	−	−	+	+	−	−	−	−	−	−	−	−	+	−	+	−	+	−	+	−	+	−	+	−	−	
voic	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	−	+	+	−	+	+	+	−	+	−	+	−	+	−	−	−	+	−	+	−	+
cont	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	−	+	−	+	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	+
nasa	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
stri	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	−	−	−	−	−	−	+	+	−	−	−	+	−	−	−	−	−	−
leng	+	−	+	−	−	+	−	+	−	+	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
back	+	+	−	−	−	−	−	−	+	+	−	+	+	−	−	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+
stre	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Afrikaans

	aː	a	eː	e	iː	ɪ	uː	oː	o	ɛ	ɛː	œ	ɔ	ɔː	ɔ̃	ɔ̃ː	ə	ɛ̃	ɛ̃ː	əɪ	œy	æ	œ	yː	ʊ	ʃ	ŋ	ʔ	ãː	øː	ø	ç	ʤ	ʧ	s	z	m	n	l	r	p	b	f	v	t	d	j	k	g	h	y
voca	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	−	−	−	+	+	+	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−	+
cons	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	+	−
sono	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	−	+	−	+	+	+	−	−	−	−	−	+	+	+	+	−	−	−	−	−	−	−	−	−	−	+
high	−	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−
low.	+	+	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	+	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
back	+	+	−	−	−	−	+	+	+	−	−	+	+	+	−	−	−	−	−	−	−	+	−	+	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−
ante	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	+	+	+	−	−	−	−	+	+	−	−	−	−	−	−
coro	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	+	+	+	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−
roun	−	−	−	−	−	−	+	+	+	−	−	+	+	+	+	+	−	−	−	−	+	−	+	+	+	−	−	−	−	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+
tens	+	−	+	−	+	−	+	+	−	−	−	−	+	−	+	−	−	−	−	−	+	−	−	+	−	−	−	+	+	+	+	−	−	+	−	−	−	−	+	−	+	−	+	−	−	+	−	+	−
voic	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	−	+	−	−	+	+	+	+	−	+	−	+	−	−	+	+	−	−	+	−	−	−	+
cont	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	+	+	−	+	−	+	+	+	+	−	−	+	+	+	−	−	+	+	−	−	+	−	−	+	+	−	−	+	+	−	−	−	+
nasa	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	+	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−
stri	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	+	+	+	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−
leng	+	−	+	−	+	−	+	+	−	−	+	−	−	+	−	+	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
stre	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
glid	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−

Dane County Kölsch

	iː	ɪ	eː	e	ə	œː	ɛː	ɒ	ɒː	aː	a	uː	oː	ʌː	əj	ej	œj	aw	ow	ʌ	œ	yː	ʊ	o	ʃ	ʒ	ð	ŋ	p	b	t	d	g	k	f	v	s	z	h	m	n	l	r	w	y	j	x
voca	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	−	−
cons	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	+	+
high	+	+	−	−	−	−	−	−	−	−	+	−	−	+	−	−	−	−	−	−	−	+	+	−	+	+	−	+	−	−	−	−	−	+	+	−	−	−	−	−	−	−	+	+	+	+
low.	−	−	−	−	−	−	−	+	+	+	−	−	−	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
back	−	−	−	−	−	−	−	+	+	+	+	+	+	+	−	−	+	+	+	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+	+	−	−	−	−	−	−	−	+	−	−	+
ante	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	+	−	−	+	+	+	−	+	+	+	−	+	+	−	−	−	−	−	−
coro	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	+	+	−	−	−	+	+	−	−	−	+	+	−	−	+	+	−	−
roun	−	−	−	−	−	+	−	+	+	−	−	+	+	−	−	−	+	−	+	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	−	−
tens	+	−	+	−	−	+	−	−	+	+	−	+	+	+	+	+	+	+	+	−	+	−	−	−	−	−	−	−	+	−	+	−	−	+	+	−	+	−	+	−	−	−	−	−	−	+
voic	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	−	+	−	+	+	−	−	+	−	+	−	+	+	+	+	−	+	+	−	−
cont	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+
sono	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	−	−
nasa	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−
stri	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	+	+	+	−	+	−	−	−	−	−	−	−	−
leng	+	−	+	−	−	+	−	+	+	−	+	+	+	+	+	+	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
glid	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
stre	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Französisch

	a	aː	ã	e	eː	ɛ	ɛ̃	ɪ	iː	o	õ	oː	œ̃	ə̃	y	yː	uː	ʊ	ɵ	œ	ɲ	ŋ	ʃ	ʒ	w	j	œː	z	s	m	n	l	r	p	b	f	v	t	d̊	k	g	h	ɔ	ɛː	ɔ̃	ɔː	ɔ̃ː	æ
voca	+	+	+	+	+	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	+	−	−	−	−	+	+	−	−	−	−	−	−	−	−	+	−	+	+	+	+
cons	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−
sono	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	+	−	−	+	−	+	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+
high	−	−	−	−	−	−	−	+	+	−	−	−	−	−	+	+	+	+	−	−	+	+	−	−	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−
low.	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	+
ante	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−
coro	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	+	+	−	+	+	+	−	−	−	+	+	−	−	−	−	−	−	−
roun	−	−	−	−	−	−	−	−	−	+	+	+	+	−	+	+	+	−	+	−	−	−	−	−	+	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	+	−
tens	−	+	−	−	+	−	−	−	+	−	−	+	−	−	+	−	−	−	−	−	−	−	−	−	+	−	+	−	−	−	−	+	−	+	−	+	−	+	−	+	−	+	−	−	−	−	−	−
voic	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	+	−	−	+	−	+	+	−	+	−	−	+	+	−	+	+	+	−	+	−	+	−	−	−	+	+	+	+	+
cont	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	−	−	+	+	−	+	+	−	+	+	+	+	+	+	+	+	−	−	+	+	−	−	−	+	−	+	+	+	−
nasa	−	−	+	−	−	−	+	−	−	−	+	−	+	+	−	−	−	−	−	−	+	+	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	−
stri	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	+	+	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−
leng	−	+	−	−	+	−	−	−	+	−	−	+	−	−	−	+	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	−
back	+	+	+	−	−	−	−	−	−	+	+	+	−	−	−	−	−	+	−	−	+	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	−	+	+	+	−
stre	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Japanisch

	ɪ	iː	ɛ	ɛː	a	aː	ɔ	ɔː	ɯ	ɯː	ɾ	p	b	m	t	tː	d	n	nː	ŋ	ɲ	ɲ̃	ʒ	ʃ	ʧ	ʤ	s	z	ç	k	kː	h	g	g̃	w	j	ʔ
voca	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cons	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-
sono	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-
high	+	+	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	+	+	-	+	+	+	-	-	+	+	+	-	+	+	+	+	+	-
low.	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
back	-	-	-	-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	+	+	+	-	+	+	+	-	+	+
ante	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-	-	+	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-
coro	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	+	-	-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
roun	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
tens	-	+	-	-	-	+	-	+	-	+	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	+
voic	+	+	+	+	+	+	+	+	+	+	-	+	+	-	-	+	+	+	+	+	+	+	-	-	+	-	+	+	-	-	-	+	+	-	-	-	+
cont	+	+	+	+	+	+	+	-	-	+	+	-	-	-	+	-	-	-	+	+	-	-	+	+	+	-	-	+	+	-	-	+	-	-	-	+	-
nasa	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
nasl	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-
stri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-
leng	-	+	-	+	-	+	-	+	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-

Türkisch

	ɯ	o	oː	a	aː	e	eː	ɾ	iː	u	uː	y	ə	b	r	j	l	g	n	ʃ	d	h	k	v	t	s	z	p	ʤ	ʧ	œ	m	f
voca	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
cons	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+
high	+	-	-	-	-	-	-	+	+	+	+	-	-	-	+	-	+	-	-	-	-	+	-	-	-	-	-	-	+	+	-	-	-	-
low.	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
back	+	+	+	+	+	-	-	-	-	+	+	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
ante	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	+	-	-	+	+	+	+	-	-	-	+	+	-	+	+
coro	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	+	+	+	-	-	+	+	+	-	+	+	-	-	-	-	-
roun	-	+	+	-	-	-	-	-	-	+	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
tens	-	-	+	-	+	-	+	-	+	-	+	-	-	-	-	-	-	-	-	-	+	+	-	+	+	-	+	-	-	+	-	-	-	-
voic	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	+	+	-	+	-	-	+	-	-	+	-	+	-	+	+	+	-
cont	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	+	-	+	-	-	+	-	+	-	+	+	-	-	-	+	+	+	+
sono	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	-	+	+	-	+	-	-	-	-	-	-	-	-	+	+	+	-
nasa	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	+	+
stri	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	-	-	+	+	-	-	-	+
leng	-	-	+	-	+	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
stre	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

TENDENZIELLE VOKALHARMONIE

G. Altmann, Bochum

1. DAS PROBLEM

In mehreren Sprachen gibt es eine kategorische Vokalharmonie, die in Grammatiken als Regel angegeben wird. Weniger bekannt ist die nichtkategorische Vokalharmonie, d.h. eine Tendenz, die Vokale der zweiten, dritten usw. Silbe des Wortes nach dem Vokal der vorangegangenen Silbe zu richten, bzw. die Vokale des Wortes irgendwie stochastisch interdependent zu plazieren. Um eine derartige Tendenz in einer Sprache zu entdecken und ihre Stärke zu charakterisieren, braucht man irgendeine statistische Methode.

Betrachten wir als Beispiel 2767 zweisilbige Grundwörter der Hanunoo-Sprache auf den Philippinen aus dem Wörterbuch von CONKLIN (1953), die in der Tabelle 1 nach den Vokalkombinationen ausgezählt wurden. Die Zahlen sind folgendermaßen zu lesen: Es gibt 698 Grundwörter, die in beiden Silben /a/ haben; es gibt 224 Grundwörter, die in der ersten Silbe /a/, in der zweiten /i/ haben; usw.

Tab. 1: Vokalkombinationen in zweisilbigen Grundwörtern des Hanunoo

		Vokal der zweiten Silbe			
		a	i	u	
Vokal der ersten Silbe	a	698	224	425	1347
	i	221	143	151	515
	u	324	96	485	905
		1243	463	1061	2767

Üblicherweise überzeugt man sich von der gegenseitigen Abhängigkeit der beiden Vokale mit einem Chiquadrat-Test, der hier einen sehr hoch signifikaten Wert liefert, aber dies ist nur ein Teil unserer Untersuchung. Es ergeben sich noch weitere Wortbildungshypothesen wie

(1) Welche Kombinationen von Vokalen werden bevorzugt, welche gemieden?

(2) Sind die Tendenzen symmetrisch, d.h. werden z.B. "a-u" und "u-a" mit gleicher Vorliebe benutzt oder gibt es Unterschiede (vgl. BOWKER 1948, KULLBACK 1968: 178)?

(3) Besteht eine Tendenz, in ein Grundwort die gleichen Vokale zu setzen? Das heißt weist die Diagonale der obigen Kontingenztabelle als ganze signifikant große Häufigkeiten auf?

Da die statistischen Tests für die anderen Probleme bekannt sind, werden wir uns hier nur mit dem letzten Problem beschäftigen und einen Test ableiten.

2. ABLEITUNG

Betrachten wir eine quadratische (k x k) Kontingenztabelle wie in Tabelle 2 dargestellt. Die inneren Zahlen, n_{ij}, bezeichnen die Zahl der Wörter mit den Vokalen i in der ersten und j in der zweiten Silbe. Weiter bedeuten

$n_{i.}$ - die Summe der Zeile i

$n_{.j}$ - die Summe der Spalte j

n - die Summe aller Häufigkeiten, d.h.

$$n = \sum_i \sum_j nij = \sum_i n_{i.} = \sum_j n_{.j}$$

Tab. 2: Quadratische (kxk) Kontingenztabelle

	1	2	....	k	
1	n_{11}	n_{12}	....	n_{1k}	$n_{1.}$
2	n_{21}	n_{22}	....	n_{2k}	$n_{2.}$
⋮	⋮	⋮		⋮	⋮
k	n_{k1}	n_{k2}	....	n_{kk}	$n_{k.}$
	$n_{.1}$	$n_{.2}$		$n_{.k}$	n

Die Wahrscheinlichkeitsverteilung der inneren Zahlen $f(n_{11},n_{12},\ldots, n_{kk})$ die wir als $f(n_{ij})$ bezeichnen, kann durch eine mehrfache Multinominalverteilung

$$f(n_{ij}) = \frac{n!}{\prod\limits_{ij} n_{ij}!} \prod\limits_{ij} p_{ij}^{n_{ij}} \tag{1}$$

dargestellt werden.

Will man die Ausprägung der Diagonale als ganze testen, so kann man auf zwei Arten verfahren.

(i) Wir stellen die Summe der Diagonalfelder als eine Variable

$$S = n_{11} + n_{22} + \ldots + n_{kk} \tag{2}$$

dar und transformieren sie auf eine normierte Normalvariable durch

$$\frac{S - E(S)}{\sqrt{V(S)}} = z, \tag{3}$$

wo E(S) die mathematische Erwartung und V(S) die Varianz von S ist. Um diese zwei Funktionen von S zu finden, gehen wir von der Annahme der Unabhängigkeit aus, d.h. wir nehmen an, daß

$$p_{ij} = p_{i.}p_{.j}.$$

Da die Randverteilungen fixiert sind, finden wir die bedingte Verteilung der n_{ij} als

$$f(n_{ij}|n_{i.},n_{.j}) = \frac{\prod_i n_{i.}! \prod_j n_{.j}!}{n! \prod_i \prod_j n_{ij}!}. \tag{4}$$

Aus (4) lassen sich die für (3) notwendigen Größen leicht berechnen. Es ist

$$E(n_{ij}) = \frac{n_{i.}n_{.i}}{n} \tag{5}$$

$$V(n_{ii}) = \frac{n_{i.}n_{.i}(n-n_{i.})(n-n_{.i})}{n^2(n-1)}. \tag{6}$$

$$\mathrm{Cov}(n_{ii},n_{i'i'}) = \frac{n_{i.}n_{.i}n_{i'.}n_{.i'}}{n^2(n-1)}. \tag{7}$$

Die Kovarianzen einzelner Felder auf der Diagonale brauchen wir, weil diese nicht unabhängig voneinander sind. Die Funktionen E(S) und V(S) ergeben sich als Summen von (5)-(7), so daß wir nach Einsetzen in (3) erhalten:

$$\frac{S - \sum_i \frac{n_{i.}n_{.i}}{n}}{\sqrt{\frac{1}{n^2(n-1)}\left[\sum_i n_{i.}n_{.i}(n-n_{i.})(n-n_{.i}) + 2\sum_{i<i'} n_{i.}n_{.i}n_{i'.}n_{.i'}\right]}} \tag{8}$$

Ein Beispiel anhand des Hanunoo wird die Einfachheit dieser Rechnung veranschaulichen. Aus Tabelle 1 bekommen wir

$$S = 698 + 143 + 485 = 1326$$

$$E(S) = \frac{1}{n}\Sigma n_{i.}n_{.i} = \frac{1}{2767}\left[1347(1243) + 515(463) + 905(1061)\right]$$

$$= 1038.2982$$

$$\frac{1}{n^2(n-1)}\Sigma n_{i.}n_{i.}(n-n_{i.})(n-n_{.i})$$

$$= \frac{1}{2767^2(2766)}\left[1347(1243)(2767-1347)(2767-1243)\right.$$

$$\left. + 515(463)2252(2304) + 905(1061)1852(1706)\right]$$

$$= 372.7738$$

$$\frac{2}{n^2(n-1)}\sum_{i<i'} n_{i.}n_{.i}n_{i'.}n_{.i'} = \frac{2}{2767^2(2766)}\left[1347(1243)515(463) + \right.$$

$$\left. + 1347(1243)905(1061) + 515(463)905(1061)\right] = 214.2017.$$

Setzt man diese Zahlen in (3) ein, so erhält man

$$z = \frac{1326 - 1038.2982}{\sqrt{372.7738 + 214.2017}} = 11.87.$$

Die Interpretation dieser Zahl findet man im §3.

(ii) Eine andere Ableitung erhält man, wenn man das likelihood-ratio benutzt. Wir interessieren uns nur für die "Randverteilung" der Diagonale, die wir als eine Variable S nehmen, und den Rest der Kontingenztabelle, n-S. Dann bekommen wir statt (1):

$$f(S,n-S) = \frac{n!}{S!(n-S)!}(\Sigma p_{ii})^S(1-\Sigma p_{ii})^{n-S}. \qquad (9)$$

Unsere Hypothesen lauten

$$H_o: \sum_i p_{ii} = \sum_i p_{i.}p_{.i}$$

$$H_1: \sum_i p_{ii} \neq \sum_i p_{i.}p_{.i}, \qquad (10)$$

so daß die Schätzungen von p_{ii} sich als

$\hat{p}_{ii} = \frac{n_{i.}n_{.i}}{n^2}$ unter der Nullhypothese und

$\hat{p}_{ii} = \frac{n_{ii}}{n}$ unter der Alternativhypothese

ergeben. Daraus folgt das likelihood-ratio

$$\lambda = \frac{\left(\sum_i \frac{n_{i.}n_{.i}}{n^2}\right)^S\left(1 - \sum_i \frac{n_{i.}n_{.i}}{n^2}\right)^{n-S}}{\left(\frac{S}{n}\right)^S\left(\frac{n-S}{n}\right)^{n-S}}. \qquad (11)$$

Die Größe - 2 ln λ ist in diesem Fall ungefähr wie ein χ^2 mit 1 Freiheitsgrad verteilt. Führen wir diese Transformation durch, so erhalten wir aus (11)

$$-2 \ln \lambda = 2S \ln\left(\frac{nS}{\Sigma n_{i.}n_{.i}}\right) + 2(n-S)\ln\left(\frac{n(n-S)}{n^2 - \Sigma n_{i.}n_{.i}}\right). \qquad (12)$$

Diese Formel ist für das Testen schon geeignet, aber man kann sie noch weiter vereinfachen, indem man die Logarithmen in eine Taylorreihe entwickelt und nur die ersten zwei Glieder in Betracht zieht.

Der Rest ist eine vernachlässigend kleine Größe der Ordnung O(1/n). So bekommen wir

$$\chi^2 = \frac{n(nS - \Sigma n_{i.}n_{.i})^2}{\Sigma n_{i.}n_{.i}(n^2 - \Sigma n_{i.}n_{.i})}, \qquad (13)$$

auch eine ungefähr wie ein Chiquadrat mit 1 Freiheitsgrad verteilte Größe.

Das Testen der Diagonale als ganze, d.h. das Testen der Existenz einer "Identitätsvokalharmonie" kann dann alternativ mit der Formel (8) oder Formel (13) durchgeführt werden.

Als Beispiel benutzen wir wieder Hanunoo und berechnen (13), wofür wir bereits alle Werte haben:

$$S = 1326$$

$$\Sigma n_{i.}n_{.i} = 2872971$$

$$n = 2767$$

so daß

$$\chi^2 = \frac{2767[2767(1326) - 2872971]^2}{2872971(2767^2 - 2872971)} = 127.60$$

Da $\chi^2_1 = z^2$, soll der Wert aus (8) ungefähr der Wurzel aus (13) entsprechen. In unserem Beispiel ist $\sqrt{127.60} = 11.30$, während in (8) z = 11.87 war. Da diese beiden Größen Approximationen sind, kann man das Resultat als identisch betrachten.

3. ANWENDUNG

Die obige Methode haben wir bei der Analyse einiger indonesischer Sprachen angewendet. Die Zählungen sind in den Tabellen 3 bis 7 dargestellt.

Tab. 3: Vokalkombinationen des Uraustronesischen

	a	ə	i	u	Σ
a	428	56	152	216	852
ə	127	95	44	92	358
i	155	24	99	77	355
u	173	16	70	239	498
Σ	883	191	365	624	2063

Tab. 4: Vokalkombinationen des Indonesischen

	a	e	i	o	u	Σ
a	806	0	268	0	400	1474
e	97	100	0	50	0	247
i	294	0	133	0	111	538
o	130	73	1	168	0	372
u	374	0	131	0	294	799
ĕ	648	4	198	1	267	1118
Σ	2349	177	731	219	1072	4548

Tab. 5: Vokalkombinationen der Bare´e-Sprache

	a	e	i	o	u	Σ
a	401	143	197	180	220	1141
e	166	220	74	152	65	677
i	128	44	133	109	91	505
o	216	171	128	342	166	1023
u	182	88	100	73	238	681
Σ	1093	666	632	856	780	4027

Tab. 6: Vokalkombinationen des Angkola-Batakischen

	a	e	i	o	u	Σ
a	1148	195	502	404	525	2774
e	185	337	8	136	78	744
i	445	48	266	208	180	1147
o	682	99	273	664	282	2000
u	599	100	197	154	549	1599
Σ	3059	779	1246	1566	1614	8264

Tab. 7: Vokalkombinationen des Sundanesischen

	a	e	ĕ	ə̆	i	o	u	Σ
a	1261	254	261	116	569	324	593	3379
e	239	352	16	0	15	220	15	857
ĕ	387	230	336	9	225	225	265	1677
ə̆	98	0	0	222	56	0	1	377
i	365	13	128	68	311	37	215	1137
o	272	202	12	0	34	505	3	1028
u	438	20	147	23	205	9	585	1427
Σ	3060	1071	901	438	1415	1320	1677	9882

Als Quellen wurden folgende Wörterbücher verwendet: Für Hanunoo CONKLIN (1953), für Malayisch/Indonesisch direkt die Daten von EMEIS (1955), für Angkola-Batakisch (Summatra) EGGINK (1936), für Bare'e (Celebes) ADRIANI (1928), für Sundanesisch (Jawa) SATJADIBRATA (1950) und für das Uraustronesisch DEMPWOLFF (1938).

In der Tabelle 8 findet man die Resultate des Testens. In der vorletzten Spalte ist die Wahrscheinlichkeit berechnet, mit der der beobachtete oder noch extremere χ^2-Wert zu erwarten wäre. Alle untersuchten Sprachen haben eine starke Identitätsvokalharmonie, die in dem Maße anwächst, wie groß die erhobene Stichprobe (oder das ganze Vokabular) ist. Die Anzahl der Vokale scheint hier keine Rolle zu spielen, weil wir die Diagonale als eine einzige Größe betrachtet haben. Es ist jedoch in der Hinsicht wichtig, daß sich mit der wachsenden Zahl der Vokale immer mehrere (starke) Einzeltendenzen entwickeln können (s. die Zahlen mit $n_{ij} = 0$, bzw. 1 im Indonesischen und Sundanesischen), weil man nicht alle Kombinationen ausnutzen kann.

Im Indonesischen fehlt die Spalte von /ĕ/, weil hier in der zweiten Silbe kein /ĕ/ stehen darf. Man kann sich dort überall eine 0 denken. In den Klammern stehen die Tests ohne die Zeile von /ĕ/.

Um den Einfluß der Stichprobengröße, die beim χ^2 arithmetisch wächst zu eliminieren, relativieren wir (13) auf

$$C = \sqrt{\chi^2/n}. \qquad (14)$$

Die Werte von C findet man in der letzten Spalte der Tabelle 8.

Tab. 8: Tests für die Diagonale

Sprache	nach (8)	nach (13)	p	C
Hanunoo	11.87	127.60		0.2147
Uraustronesisch	12.93	145.00	alle	0.2651
Indonesisch	20.22 (19.62)	234.74 (295.84)	kleiner als 2×10^{-10}	0.2272
Bare'e	18.99	348.87		0.2943
Angkola	27.26	672.33		0.2852
Sundanesisch	47.62	1968.62		0.4463

Das vorgeschlagene Verfahren ermöglicht es, die Stärke der Vokalharmonie in einer Sprache oder Sprachfamilie historisch zu verfolgen, typologische Vergleiche durchzuführen, aber vor allem regt es an, nach den Ursachen dieser Erscheinung zu suchen und ihren Zusammenhang mit anderen Eigenschaften der Sprache zu untersuchen.

LITERATUR

ADRIANI, N., Bare'e-Nederlandsch Woordenboek. Leiden 1928

BOWKER, A.H., A test for symmetry in contingency tables. Journal of the American Statistical Association 43, 1948, 572-574

CONKLIN, H.C., Hanunóo-Englisch Dictionary. Berkeley and Los Angeles 1953

DEMPWOLFF, O., Vergleichende Lautlehre des austronesischen Wortschatzes III. Berlin 1938

EGGINK, H.J., Angkola- en Mandailing-Bataksch-Nederlandsch Woordenboek. Bandoeng 1936

EMEIS, M.G., Vocaal-harmonie in de Maleise kern van de Indonesische vocabulaire. Bijdragen tot de Taal-, Land- en Volkenkunde 111, 1955, 191-201

KULLBACK, S., Information theory and statistics. New York 1968.

SATJADIBRATA, R., Kamoes Soenda-Indonesia. Djakarta 1950

STUART, A., A test for the homogeneity of the marginal distributions in a two way classification. Biometrika 42, 1955, 412-416

UNTERSUCHUNGEN ZUR MATHEMATISCHEN BESCHREIBUNG DES MARTINGESETZES DER ABSTRAKTIONSEBENEN

R. Hammerl, Kielce

1. VORBEMERKUNGEN

Den Gegenstand unserer Untersuchungen stellt das Martingesetz der Abstraktionsebenen dar (ALTMANN, KIND 1983), das die Abhängigkeit zwischen der Zahl der durch Substantive repräsentierten Begriffe auf einer bestimmten Begriffsebene (Symbol: y_x) und dem Index dieser Begriffsebene (Symbol: x) untersucht.

Als Datenbasis dienen Untersuchungsergebnisse zu diesem Gesetz für die polnische Sprache (SAMBOR 1982) und für die französische Sprache (MARTIN 1974). In beiden Untersuchungen wurde aus einem einsprachigen Erklärungswörterbuch eine bestimmte Zahl von Substantiven ausgelost. Untersucht wurde die Bedeutungsbeschreibung dieser Substantive im benutzten Wörterbuch, die - von einigen Ausnahmen und Problemen abgesehen (ALTMANN, KIND 1983, 2ff.) - durch genus proximum und differentia specifica gegeben wurde.

Der Ausgangsbegriff war von 1. Ordnung (x = 1), der durch das genus proximum repräsentierte Begriff von 2. Ordnung (x = 2); dieser Begriff 2. Ordnung wurde - soweit möglich - analog unter einem Begriff 3., 4. usw. Ordnung subsumiert. Auf diese Weise fand man Begriffsfolgen, wobei die zu einer bestimmten Folge gehörigen Begriffe auf verschiedenen Begriffsebenen liegen. Für die polnische Sprache wurden z.B. folgende Begriffsfolgen gefunden (SAMBOR 1982; Tabelle 1).

An den wenigen Beispielen in Tabelle 1 erkennt man,

- daß die Länge solcher Begriffsketten unterschiedlich ist,
- daß sich dieselben Begriffe auf denselben Begriffsebenen wiederholen können (z.B. *gaz* bei x = 2 und *substancja* bei x = 3), was ganz einfach davon zeugt, daß z.B. die relativ speziellen Begriffe *amoniak* und *powietrze* unter ein und demselben Oberbegriff *gaz* subsumiert wurden; daraus folgt, daß die Zahl (y_x) unterschiedlicher Begriffe auf höheren Begriffsebenen (mit größer werdendem x) kleiner wird;

- daß sich dieselben Begriffe auch auf verschiedenen Begriffsebenen wiederholen können (z.B. *część* bei x = 4 und x = 7),
- daß in jeder untersuchten Stichprobe jeder Begriffsebene (x) eine bestimmte Zahl (y_x) von unterschiedlichen Begriffen zugeordnet werden kann.

Tab.1: Begriffsfolgen für die polnische Sprache

Begriffsebene x	Beispiele für Begriffsfolgen			
1	*amoniak* (Ammoniak)	*alkowa* (Alkoven)	*abstrakt* (Abstraktum)	*powietrze* (Luft)
2	*gaz* (Gas)	*pokój* (Zimmer)	*pojęcie* (Begriff)	*gaz* (Gas)
3	*substancja* (Substanz)	*pomieszczenie* (Raum, Wohnstätte)	*składnik* (Bestandteil)	*substancja* (Substanz)
4	*materia* (Materie)	*miejsce* (Platz,Raum)	*część* (Teil)	*materia* (Materie)
5		*przestrzeń* (Raum,Gebiet)		
6		*obszar* (Raum, Gebiet, Territorium)		
7		*część* (Teil)		

MARTIN (1974), der nach dem oben geschilderten allgemeinen methodischen Vorgehen die französische Sprache untersuchte, kam zu folgenden Ergebnissen (Tabelle 2)[1]:

Tab. 2: Zahl der Wörter auf einzelnen Abstraktionsebenen für die französische Sprache (mit und ohne Begriffswiederholungen)

Ebene x	Zahl der Wörter y_x	
	mit Begriffswiederholungen	ohne Begriffswiederholungen
1	1723	1375
2	348	240
3	108	69
4	39	26
5	13	10
6	3	3

Die entsprechenden Daten für die Untersuchungen zur polnischen Sprache werden in Tabelle 3 angeführt.

Um nun einen exakten Vergleich der Zahlenwerte der Tabellen 2 und 3 durchführen zu können, müßten zumindest drei Voraussetzungen erfüllt sein:

1) das methodische Vorgehen für die Suche der Begriffsketten müßte in beiden Fällen (Untersuchungen zur polnischen und französischen Sprache) gleich sein; inwieweit diese Bedingung erfüllt ist, kann aufgrund der Literaturangaben (MARTIN 1974; SAMBOR 1986) nicht eindeutig entschieden werden;

2) die Methodik der Erarbeitung der in beiden Untersuchungen ausgenutzten Wörterbücher müßte gleich sein; die Überprüfung dieser Bedingung geht weit über den Rahmen dieses Artikels hinaus;

3) der Zusammenhang zwischen y_x und x muß mathematisch möglichst exakt beschrieben werden, d.h., es muß ein möglichst allgemeines mathematisches Modell für die Beschreibung der Abhängigkeit zwischen y_x und x gefunden werden. Im folgenden wollen wir und besonders der Suche dieses Modells zuwenden.

Tab. 3: Zahl der Wörter auf einzelnen Abstraktionsebenen für die polnische Sprache (mit und ohne Begriffswiederholungen)

Ebene x	Zahl der Wörter y_x	
	mit Begriffswiederholungen	ohne Begriffswiederholungen
1	1000	382
2	618	347
3	271	161
4	110	66
5	44	28
6	16	7
7	9	6
8	3	2
9	1	1

Bei der Suche allgemeiner mathematischer Modelle geht man meist von bestimmten theoretischen Überlegungen aus (z.B. ALTMANN, KIND 1983; ORLOV 1982; ALTMANN 1980), die möglichst allgemein und relativ einfach mathematisch beschreibbar sind. Die Überprüfung des mathematischen Modells an empirischen Daten (aus repräsentativen

Stichproben) zeigt, inwieweit die allgemeinen theoretischen Grundlagen zutreffend oder modifizierungsbedürftig sind.

Ein mathematisches Modell für die Beschreibung des Zusammenhangs zwischen y_x und x haben ALTMANN und KIND (1983) an den Daten zur französischen Sprache überprüft. Dieses Modell beschreibt die Daten für die französische Sprache relativ gut. Wie im folgenden gezeigt wird, ist es jedoch für die Beschreibung der Daten der polnischen Sprache nicht besonders geeignet.

Ziel unserer Untersuchungen ist demzufolge die Suche eines mathematischen Modells zur Beschreibung des Zusammenhangs zwischen y_x und x für die polnische und französische Sprache (bisher liegen keine weiteren Ergebnisse aus entsprechenden empirischen Untersuchungen anderer Sprachen vor).

2. ABLEITUNG EINES MATHEMATISCHEN MODELLS

Ausgangspunkt unserer Untersuchungen ist das Modell von ALTMANN, KIND (1983). Mit den allgemeinen Überlegungen, daß die Zahl unterschiedlicher Begriffe y_{x+1} auf der Ebene x+1

- von der Zahl y_x der Begriffe auf der jeweils niedrigeren Ebene x und
- von der Ordnung x+1 der Ebene selbst abhängt, wurde aus

$$y_{x+1} \sim (x+1)y_x \qquad (1)$$

und

$$y_{x+1} = c(x+1)y_x \qquad (2)$$

die allgemeine Formel

$$y_x = y_1\, x!\, c^{x-1} \qquad (3)$$

gefunden, wo c eine Konstante darstellt, die nach

$$c = \frac{y_2}{2y_1} \qquad (4)$$

abgeschätzt werden kann. Die Genauigkeit der c-Schätzung kann durch Anwendung der Methode der kleinsten Quadrate oder der von ALTMANN, KIND (1983, 7ff.) beschriebenen iterativen Methode verbessert werden.

Die Tabellen 4 und 5 führen die nach Gleichung (3) von ALTMANN und KIND berechneten $\hat{y}_x$-Zahlenwerte für die französische Sprache auf.

Unter Anwendung des F-Tests kann man sich schnell davon überzeugen, daß die Differenzen zwischen $\hat{y}_x$ und y_x als zufällig angesehen werden können, d.h., daß Formel (3) die Abhängigkeit zwischen y_x und x hier relativ gut beschreibt.

Untersucht man jedoch Gleichung (2); so stellt man mit $c > 0$ und somit $c^{x-1} > 0$ leicht fest, daß bei einem bestimmten x der Ausdruck $c(x+1) > 1$ und somit $y_{x+1} > y_x$ wird, was den empirischen Daten widerspricht (siehe Tabelle 6).

Tab. 4: nach Gleichung (3) berechnete Zahlenwerte $\hat{y}_x$ und empirische Zahlenwerte y_x für die französische Sprache (mit Begriffswiederholungen)

x	y_x	$\hat{y}_x$
1	1723	1723.00
2	348	344.60
3	108	103.38
4	39	41.35
5	13	20.68
6	3	12.40
		c = 0.1

Tab. 5: nach Gleichung (3) berechnete Zahlenwerte $\hat{y}_x$ und empirische Zahlenwerte y_x für die französische Sprache (ohne Begriffswiederholungen)

x	y_x	$\hat{y}_x$
1	1375	1375.00
2	240	240.00
3	69	62.84
4	26	21.94
5	10	9.57
6	3	5.01
		c = 0.087273

Durch eine geschickte c-Schätzung kann man bei Anwendung von Gleichung (2) oder (3) wohl der widersprüchlichen Tatsache

$$y_{x+1} > y_x$$

entgehen (was Tabelle 7 zeigt), die Anpassung des Modells an die empirischen Daten aber ist auch hier nicht besonders gut. Aus Tabelle 4 und 5 kann weiterhin abgelesen werden, daß das Modell von ALTMANN, KIND für kleine x fast regulär kleinere und für größere x größere $\hat{y}_x$-Werte liefert (im Vergleich mit den entsprechenden y_x-Werten). Dies überzeugte uns davon, daß man durch eine geeignete Modifizierung des Modells von ALTMANN und KIND über dessen Verallgemeinerung eine genauere Beschreibung der empirischen Daten erreichen kann.

Tab. 6: y_x- und $\hat{y}_x$-Werte (berechnet nach Gleichung (2)) für die polnische Sprache

x	mit Begriffswiederholungen		ohne Begriffswiederholungen	
	y_x	$\hat{y}_x$	y_x	$\hat{y}_x$
1	1000	1000.00	382	382.00
2	618	618.00	347	347.00
3	271	572.89	161	472.80
4	110	708.09	66	858.95
5	44	1094.00	28	1950.58
6	16	2028.28	7	5135.50
7	9	4387.17	6	16899.35
8	3	10845.08	2	61402.78
9	1	30160.17	1	250991.23
	c = 0.309		c = 0.45418	

Tab. 7: y_x- und $\hat{y}_x$-Werte (berechnet nach Gleichung (2)) für die polniche Sprache

x	mit Begriffswiederholungen		ohne Begriffswiederholungen	
	y_x	$\hat{y}_x$	y_x	$\hat{y}_x$
1	1000	1000.00	382	382.00
2	618	222.96	347	76.40
3	271	74.57	161	22.92
4	110	33.25	66	9.17
5	44	18.53	28	4.58
6	16	12.93	7	2.75
7	9	9.67	6	1.93
8	3	8.62	2	1.54
9	1	8.65	1	1.39
	c = 0.11148		c = 0.100	

Wir gingen von der Grundannahme aus, daß (durch eine Verallgemeinerung von (2)) direkte Proportionalität zwischen

$$y_{x+1} \sim \frac{1}{(x+1)^a} \cdot y_x^b \qquad (5)$$

vorliegt, wobei a und b Konstanten sind. Wir sind demzufolge der Meinung, daß die Exponenten a und b die fast regulären Abweichungen zwischen y_x und $\hat{y}_x$ (Tabellen 4 und 5) zum großen Teil erfassen und erklären können. Unsere Grundformel lautet demzufolge:

$$y_{x+1} = \frac{c}{(x+1)^a} \cdot y_x^b, \qquad (6)$$

wobei c den Proportionalitätsfaktor darstellt.

Es ist leicht einzusehen, daß mit a = -1 und b = 1 Gleichung (6) in Gleichung (2) übergeht, so daß das Modell von ALTMANN, KIND demzufolge ein Spezialfall unseres Modells ist.

Tab. 8: y_x- und $\hat{y}_x$-Werte (berechnet nach Gleichung (6)) für die polnische Sprache

x	mit Begriffswiederholungen		ohne Begriffswiederholungen	
	y_x	$\hat{y}_x$	y_x	$\hat{y}_x$
1	1000	1000.00	382	382.00
2	618	622.6	347	347.01
3	271	274.2	161	161.00
4	110	111.4	66	66.00
5	44	46.6	28	28.53
6	16	20.92	7	13.62
7	9	10.19	6	7.18
8	3	5.37	2	4.12
9	1	3.04	1	2.54
	a = 1.25528 b = 0.65691 c =15.9		a = 1.7764783 b = 0.49578 c =62.37	

Die unter Anwendung der Methode der kleinsten Quadrate und der iterativen Methode für 2 Parameter (a und b) nach Gleichung (6) berechneten $\hat{y}_x$-Werte für die polnische und französische Sprache führen die Tabellen 8 und 9 auf.

Tab. 9: y_x- und $\hat{y}_x$-Werte (berechnet nach Gleichung (6)) für die französische Sprache

x	mit Begriffswiederholungen		ohne Begriffswiederholungen	
	y_x	$\hat{y}_x$	y_x	$\hat{y}_x$
1	1723	1723.00	1375	1375.00
2	348	347.36	240	243.40
3	108	106.78	69	67.90
4	39	36.67	26	26.96
5	13	7.01	10	13.50
6	3	0.03 (1.00)	3	7.86
		$a = -6.9529841$ $b = 2.5445722$ $c = 1.6353376 \cdot 10^{-8}$		$a = 1.53314623$ $b = 0.3782292$ $c = 45.7975$

Im folgenden soll durch die Anwendung des F-Tests überprüft werden, inwieweit die Abweichungen zwischen den Werten von y_x und $\hat{y}_x$ für die 4 unterschiedenen Fälle der Tabellen 8 und 9 statistisch signifikant sind. Dazu berechnen wir die Testgröße F nach

$$F = \frac{SSR/(k-1)}{SSE/(n-k)} \qquad (7)$$

mit

$$SSR = \sum_{x=1}^{n} (\ln \hat{y}_x - \ln \bar{y})^2 \qquad (8)$$

und

$$SSE = \sum_{x=1}^{n} (\ln y_x - \ln \hat{y}_x)^2, \qquad (9)$$

wobei k die Zahl der Parameter der Regressionsfunktion (bei uns: k = 3) und n die Zahl der Meßwerte (bei uns: n = 9 für die polnische Sprache und n = 6 für die französische Sprache) ist. Unter Ausnutzung der Gleichungen (7) - (9) gilt:

a) polnische Sprache, mit Begriffswiederholungen
   $F(2,6) = 79.68$ $P = 0.00005$

b) polnische Sprache, ohne Begriffswiederholungen
   $F(2,6) = 59.95$ $P = 0.0001$

c) französische Sprache, mit Begriffswiederholungen
   $F = (2,3) = 23.10$ $P = 0.015$

d) französische Sprache, ohne Begriffswiederholungen
   $F = (2,3) = 34.51$ $P = 0.0005$

Da für alle vier Fälle die Testgröße F größer als der kritische $F_{0.05}(\nu_1,\nu_2)$-Wert (ν_1 und ν_2 stellen die Zahl der Freiheitsgrade dar) ist, kann geschlußfolgert werden, daß die Abweichungen zwischen den Werten von y_x und $\hat{y}_x$ bei einem Signifikanzniveau von $\alpha = 0.05$ als zufällig angesehen werden können, was die gute Anpassung unseres Modells an die empirischen Zahlenwerte bestätigt.

Bei einer Anwendung unseres Modells auf entsprechende Untersuchungsergebnisse aus anderen Sprachen wird sich zeigen, inwieweit von einer "Universalität" dieses Modells gesprochen werden kann (empirische Validierung).

Die theoretische Validierung beruht auf der Einordnung unseres Modells in ein System anderer (Sprach)gesetze, in der erschöpfenden sprachpsychologischen und linguistischen Interpretation der ausgenutzten Parameter. Bisher stehen uns jedoch zu wenig empirische Daten zur Verfügung, und auch unser Wissen über allgemeine Sprachgesetze bzw. Möglichkeiten der Parameterinterpretation ist noch nicht groß. Trotzdem soll versucht werden, eine einführende Diskussion des vorgestellten Modells vorzunehmen und einige Konsequenzen aufzuzeigen.

3. DISKUSSION

Zunächst sei darauf verwiesen, daß das Martingesetz der Abstraktionsebenen im Grunde genommen sprachpsychologische Probleme beschreibt und somit auch Wissen um psycholinguistische, psychologische und kybernetische Probleme mit einschließt. Es ist in diesem Zusammenhang auch kein Zufall, daß in unser mathematisches Modell Potenzfunktionen eingegangen sind, da ja gerade solche Funktionen in der psychologischen Forschung reiche Anwendung finden (z.B. die Potenzfunktionen von Stevens; siehe SYDOW, PETZOLD 1981, 155ff.).

Unser allgemeines Modell beinhaltet mehrere interessante Spezialfälle, die nun kurz genannt werden sollen. Dadurch soll lediglich angedeutet werden, daß aus einem Vergleich der empirischen Daten (Tabellen 2 und 3) mit den aus den Spezialfällen unseres Modells berechneten Daten wertvolle Hinweise über Spezifika der Begriffshierarchie in verschiedenen Sprachen und über die Interpretation der Parameter a und b gewonnen werden können.

Da jedoch die empirischen Daten für die polnische und französische Sprache keinen exakten Vergleich zulassen (weil MARTIN keine ausreichenden Hinweise über die Datenerfassung angibt), soll die Möglichkeit eines solchen Vergleichs nur an einem theoretisch interessanten Spezialfall angedeutet werden.

1. Spezialfall: $a = -1$, $b = 1$

Setzt man diese Parameterwerte in Gleichung (6) ein, so erhält man:

$$y_{x+1} = c \cdot (x+1) \cdot y_x$$

oder allgemein: $$y_x = y_1 \cdot x! \cdot c^{x-1}.$$

Dieser Spezialfall betrifft das von ALTMANN und KIND (1983) vorgeschlagene Modell. Um einen Vergleich der nach der angegebenen Formel berechenbaren $\hat{y}_{x+1}$-Zahlenwerte mit den entsprechenden empirischen Zahlenwerten für Daten verschiedener Sprachen durchführen zu können, führen wir zunächst die Größe

$$f_x = \frac{y_x}{y_{x-1}} \qquad \text{für } x \geq 2 \tag{10}$$

ein, wobei gilt:

$$f_x = \frac{c \cdot x \cdot y_{x-1}}{y_{x-1}} = c \cdot x. \tag{11}$$

Da auch f_x von c abhängig ist und da wir für unseren Vergleich eine von c unabhängige Größe benötigen, definieren wir eine Funktion g_x, die nach

$$g_x = \frac{f_x}{f_{x-1}} = \frac{y_{x-2} \cdot y_x}{(y_{x-1})^2} \qquad \text{für } x \geq 3 \tag{12}$$

oder bei Ausnutzung von Gleichung (10) nach

$$g_x = \frac{c \cdot x}{c(x-1)} = \frac{x}{x-1} \qquad \text{für } x \geq 3 \tag{13}$$

berechnet werden kann.

Tabelle 10 führt die nach Gleichung (12) unter Ausnutzung der Zahlenwerte der Tabellen 2 und 3 berechneten empirischen Werte für g_x auf und die nach Gleichung (13) berechneten theoretischen Werte für $\hat{g}_x$.

Die Zahlenwerte der Tabelle 10 deuten auf die Brauchbarkeit des hier untersuchten Spezialfalles bei der Beschreibung der empirischen Zahlenwerte der französischen Sprache hin, was natürlich exakt überprüft werden muß.

Bei der Prüfung, ob die Differenzen zwischen den Werten von g_x und $\hat{g}_x$ signifikant sind, müssen wir einen verteilungsfreien Test anwenden. Da wir es hier nicht mit Meßwerten im eigentlichen Sinne, sondern mit Quotienten aus Häufigkeiten zu tun haben, haben wir uns für die Anwendung des X-Testes von VAN DER WAERDEN entschieden (STORM 1974, 258). Da die Gütefunktion dieses Tests (d.h. die Wahrscheinlichkeit, die Nullhypothese abzulehnen) kleiner als bei Parametertests ist, kann im Falle der Ablehnung der Nullhypothese jeder andere eventuell einsetzbare Parametertest kein anderes Ergebnis liefern (STORM 1974, 249). Bei der Anwendung dieses Tests gehen wir von der Nullhypothese H_o aus, daß beide Stichproben (1. Stichprobe: alle g_x-Werte; 2. Stichprobe: alle $\hat{g}_x$-Werte) aus Grundgesamtheiten mit der gleichen diskreten Verteilungsfunktion stammen. Die Testgröße X berechnet man dann nach

$$X_{test} = \left| \sum_r \varphi\left(\frac{r}{n_1+n_2+1}\right) \right|, \tag{14}$$

wo n_1 (bzw. n_2) die Anzahl der Zahlenwerte in Stichprobe 1 (bzw. in Stichprobe 2) darstellt, r die Rangzahlen der Meßwerte der ersten Stichprobe, falls alle Werte (g_x und $\hat{g}_x$) der Größe nach geordnet und mit Rangzahlen (bei 1 beginnend) versehen wurden.

Ersetzen wir in Gleichung (14) die Größe $\frac{r}{n_1+n_2+1}$ durch y', so gilt:

$$\varphi\left(\frac{r}{n_1+n_2+1}\right) = \varphi(y'). \tag{15}$$

Tab. 10: Zahlenwerte für g_x und $\hat{g}_x$ für die polnische und französische Sprache

	polnische Sprache				französische Sprache			
	mit Begriffswiederholungen		ohne Begriffswiederholungen		mit Begriffswiederholungen		ohne Begriffswiederholungen	
x	g_x	$\hat{g}_x$	g_x	$\hat{g}_x$	g_x	$\hat{g}_x$	g_x	$\hat{g}_x$
3	0.71	1.50	0.51	1.50	1.54	1.50	1.65	1.50
4	0.93	1.33	0.88	1.33	1.16	1.33	1.31	1.33
5	0.99	1.25	1.03	1.25	0.92	1.25	1.02	1.25
6	0.91	1.20	0.59	1.20	0.69	1.20	0.78	1.20
7	1.55	1.17	3.43	1.17				
8	0.59	1.14	0.39	1.14				
9	1.00	1.13	1.50	1.13				

$\varphi(y')$ stellt die Umkehrfunktion für die Verteilungsfunktion der standardisierten Normalverteilung

$$y = \Phi(x) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \qquad (16)$$

dar.

Demonstrieren wir nun die Berechnung von X_{test} anhand der Zahlenwerte für die polnische Sprache (mit Begriffswiederholungen).

a) Zunächst werden die Zahlenwerte von g_x und $\hat{g}_x$ der Größe nach geordnet und mit Rangzahlen r versehen.

g_x bzw. $\hat{g}_x$	r
1.55	1
1.50	2
1.33	3
1.25	4
1.20	5
1.17	6
1.14	7
1.13	8
1.00	9
0.99	10
0.93	11
0.91	12
0.71	13
0.59	14

b) Da uns lediglich die Rangzahlen der g_x-Werte interessieren, ist es ratsam, diese Werte zu unterstreichen und die entsprechenden r-Werte gesondert zu notieren.

$$r = 2,\ r = 3,\ r = 4,\ r = 5,\ r = 6,\ r = 7,\ r = 8$$

c) Nun berechnen wir für jedes r (bei $n_1=n_2=7$) nach

$$y' = \frac{r}{n_1+n_2+1}$$

den entsprechenden Zahlenwert y'. Für unser Beispiel gilt:

$$y' = \frac{2}{7+7+1} = 0.13333;\ y' = 0.20000;\ y' = 0.26667;$$

$$y' = 0.40000;\ y' = 0.46667;\ y' = 0.53333.$$

d) Unter Ausnutzung von entsprechenden Tabellen liest man nun für jedes y' den entsprechenden Wert von $\varphi(y')$ ab und addiert die derhaltenen Werte. Der Betrag der erhaltenen Summe stellt die Testgröße X_{test} dar.

$$\varphi(0.13333) = -1.11$$
$$\varphi(0.20000) = -0.84$$
$$\varphi(0.26667) = -0.62$$
$$\varphi(0.33333) = -0.43$$
$$\varphi(0.40000) = -0.25$$
$$\varphi(0.46667) = -0.08$$
$$\varphi(0.53333) = +0.08$$
$$X_{test} = 3.25$$

Wendet man das beschriebene Verfahren auf alle vier in Tab. 8 und 9 beschriebenen Fälle an, so erhält man folgende Ergebnisse:

a) polnische Sprache (mit Begriffswiederholungen)
$X_{test} = +3.25$ $\quad X_{\alpha=0.05,7,7} = 3.11$

b) polnische Sprache (ohne Begriffswiederholungen)
$X_{test} = +2.16$ $\quad X_{\alpha=0.05,7,7} = 3.11$

c) französische Sprache (mit Begriffswiederholungen)
$X_{test} = +1.17$ $\quad X_{\alpha=0.05,7,7} = 2.40$

d) französische Sprache (ohne Begriffswiederholungen)

$x_{test} = +0.25$ $\quad x_{\alpha=0.05,7,7} = 2.40$

Bei zweiseitiger Fragestellung bei einem Signifikanzniveau von $\alpha = 0.05$ kann für den ersten Fall (polnische Sprache - mit Begriffswiederholungen) die Nullhypothese abgelehnt werden, d.h., beide Stichproben wurden aus Populationen mit unterschiedlichen Verteilungsfunktionen entnommen. Das zur Berechnung der $\hat{g}_x$-Werte angewandte Modell von ALTMANN, KIND sollte in diesem Falle nicht zur Beschreibung der empirischen Daten (g_x-Werte) herangezogen werden. Besser geeignet scheint dieses Modell für die Beschreibung der Daten der französischen Sprache zu sein.

Nun noch einige Bemerkungen zu den Zahlenwerten der Parameter a, b und c. Zwischen diesen Zahlenwerten besteht eine funktionale Abhängigkeit, so daß die Veränderung des Zahlenwertes eines Parameters gleichzeitig Veränderungen der Zahlenwerte der anderen Parameter nach sich zieht.

Gehen wir von der Gleichung

$$y_x = \frac{c}{x^a} \cdot y_{x-1}^b \tag{17}$$

aus und beachten, daß für die zu beschreibenden empirischen Daten $y_x > y_{x-1} > y_{x-2}$ usw. gilt. Große und positive b-Werte ($b \geq 1$) ziehen dann

- relativ große positive a-Werte und relativ kleine c-Werte

nach sich oder

- negative a-Werte und sehr kleine c-Werte (Tabelle 9).

a und b beeinflussen sich gegenseitig bei der Annäherung der Kurvenkrümmung der theoretischen Kurve (definiert durch die Menge der Punkte $(\hat{y}_x, x)$) an die empirische Kurve (definiert durch die Menge der Punkte (y_x, x)); c stellt lediglich einen Proportionalitätsfaktor dar. Für den hier untersuchten Spezialfall (b = 1, a = -1) gilt, daß ein relativ großes b (mit b = 1) und ein relativ kleines a (mit a = -1) nur durch ein relativ kleines c ausgeglichen werden kann, um die Bedingung $y_x > y_{x-1} > y_{x-2}$ usw. zu erfüllen. Der Proportionalitätsfaktor in dem von ALTMANN, KIND vorgeschlagenen Modell muß demzufolge einen relativ kleinen Zahlenwert annehmen. Zur Zeit ist es uns noch nicht möglich, eine genauere Interpretation der Parameter a und b zu geben.

2. Spezialfall: a = 1, b = 0, c > 1

$$y_x = \frac{c}{x^a} \cdot y_{x-1}^b = \frac{c}{x} \tag{18}$$

Mit c = 1000 erhält man die in Tabelle 11 angegebenen Zahlenwerte.

3. Spezialfall: a = 0, b = 0, $c \neq 0$

$$y_x = \frac{c}{x^a} \cdot y_{x-1}^b = c \tag{19}$$

4. Spezialfall: a = 0, b = 1, $c \neq 0$

$$y_x = \frac{c}{x^a} \cdot y_{x-1}^b = c\, y_{x-1} \tag{20}$$

Die für c = 0,5 und $y_1 = 1000$ errechneten Zahlenwerte führt Tabelle 11 auf.

5. Spezialfall: a = 1, b = 1, c = 1

$$y_x = \frac{c}{x^a} \cdot y_{x-1}^b = \frac{1}{x}\, y_{x-1} \tag{21}$$

oder allgemein

$$y_x = \frac{y_1}{x!} \tag{22}$$

Die für $y_1 = 1000$ errechneten Zahlenwerte führt Tabelle 11 auf.

6. Spezialfall: $a \neq 0$, b = 0, $c \neq 0$

$$y_x = \frac{c}{x^a} \cdot y_{x-1}^b = \frac{c}{x^a} \tag{23}$$

7. Spezialfall: $a = 0$, $b \neq 0$, $c \neq 0$

$$y_x = \frac{c}{x^a} \cdot y_{x-1}^b = c \cdot y_{x-1}^b \tag{24}$$

Tab. 11: nach den Gleichungen (18), (20) und (22) berechnete $\hat{y}_x$-Zahlenwerte

x	$\hat{y}_x$ (nach Gleichung 18)	$\hat{y}_x$ (nach Gleichung 22)	$\hat{y}_x$ (nach Gleichung 20)
1	1000.00	1000.00	1000.00
2	500.00	500.00	500.00
3	333.33	166.67	250.00
4	250.00	41.67	125.00
5	200.00	8.33	62.50
6	166.67	1.39	31.25

4. ZUSAMMENFASSUNG

Das von uns vorgeschlagene Modell zur Beschreibung des Martingesetzes der Abstraktionsebenen wurde an empirischen Daten zur französischen und polnischen Sprache getestet, wobei gute Resultate erzielt wurden. Die Brauchbarkeit dieses mathematischen Modells kann sich jedoch erst bei der Anwendung auf Material aus anderen Sprachen endgültig zeigen.

Es wird vermutet, daß eine völlig andere Anordnung des empirischen Materials und dessen Zuordnung zu einzelnen Abstraktionsebenen (z.B. erhält die höchste Abstraktionsebene den kleinsten x-Wert) völlig andere Einsichten in die untersuchte Problematik erlaubt. Dies muß jedoch Gegenstand weiterer intensiver Untersuchungen sein. Dabei muß aber auch die Frage nach objektiven Kriterien bei der Datenerfassung gelöst werden.

Der von uns untersuchte Zusammenhang zwischen dem Niveau der Abstraktionsebenen (x) und der Zahl der ihnen zugeordneten Begriffe (y_x) ist auch aus psychologischer, psycholinguistischer und kybernetischer Hinsicht höchst interessant. Hier kann z.B. Fragen der Effektivität in der Kodierung sprachgebundener Informationen, der psychologischen Realität der Hierarchie von sprachlichen Begriffen (siehe: KLIX, KUKLA, KÜHN 1979), Fragen der Informationsspeicherung usw. nachgegangen werden.

Für Sprachstatistiker erscheint uns der Zusammenhang zwischen der mittleren Länge der Substantive (als Repräsentanten von Begriffen) auf bestimmten Begriffsebenen und den Veränderlichen x bzw. y_x von Bedeutung.

ANMERKUNG

1) Die Daten für die französische Sprache ohne Begriffswiederholungen wurden von ALTMANN, KIND (1983) errechnet.

LITERATUR

ALTMANN, G., Prolegomena to Menzerath's law. In: Grotjahn, R. (ed.), Glottometrika 2, Bochum, Brockmeyer, 1980, 1-10

ALTMANN, G., KIND, B., Ein semantisches Gesetz. In: Köhler, R., Boy, J. (Hrsg.), Glottometrika 5, Bochum, Brockmeyer, 1983, 1-13

KLIX, F., KUKLA, F., KÜHN, R., Zur Frage der Unterscheidbarkeit von Klassen semantischer Relationen im menschlichen Gedächtnis. In: Bierwisch, M. (Hrsg.), Psychologische Effekte sprachlicher Strukturkomponenten, Berlin, Akademie-Verlag, 1979, 131-144

MARTIN, R., Syntaxe de la définition lexicographique: étude quantitative des définissants dans le "Dictionnaire fondamental de la langue française", In: David, J., Martin, R. (Hrsg.), Statistique et linguistique. Paris, Klincksieck 1974, 61-71

ORLOV, Ju.K., Dynamik der Häufigkeitsstrukturen. In: Orlov, Ju.K., Boroda, M.G., Nadarejšvili, I.Š., Sprache, Kunst, Text, Quantitative Analysen. Bochum, Brockmeyer, 1982, 82-117

SAMBOR, J., Lexikographische Definitionen. Bochum 1982 (unveröffentlichte Sammlung von 1000 Begriffsketten für die polnische Sprache unter Ausnutzung folgenden Wörterbuches: Skorupka, S., Auderska, H., Łempicka, Z., Mały słownik języka polskiego, Warszawa, Państwowe Wydawnictwo Naukowe, 1968)

SAMBOR, J., O budowie tzw. ciągów definicyjnych (na materiale definicji leksykalnych) (Über den Bau der sogenannten Definitionsfolgen - am Material lexikalischer Definitionen). In: Księga ku czci prof. Adama Heinza (1986)

STORM, R., Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle. Leipzig, VEB Fachbuchverlag, 1974

SYDOW, H., PETZOLD, P., Mathematische Psychologie. Berlin, VEB Deutscher Verlag der Wissenschaften, 1981

EINE VERALLGEMEINERUNG DES GESETZES DER SEMANTISCHEN DIVERSIFIKATION

G. Altmann, K.-H. Best, B. Kind

1. In einigen Untersuchungen haben BEÖTHY und ALTMANN (1984, 85,86) gezeigt, daß die nach ihrer Größe geordneten Vorkommenshäufigkeiten der verschiedenen Bedeutungen von Wörtern und Morphemen der negativen Binomialverteilung folgen. Die Autoren gingen dabei von der inzwischen schon geläufigen Annahme aus, daß Häufigkeiten sprachlicher Entitäten sich durch stochastische Gesetze beschreiben lassen. Zur Modellierung des genannten Phänomens verwendeten die Autoren einen Poissonschen Prozeß und kamen zu dem Resultat, daß die Rang-Häufigkeitsverteilung einzelner Bedeutungen einer linguistischen Einheit der verschobenen negativen Binomialverteilung:

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

wobei $r > 0$, $0 < p < 1$, $q = 1 - p$, folgt. Eine Korrektur der Ableitung hat SCHWEIGER (1986) gebracht.

2. Die Überprüfung der Hypothese wurde an fünf ungarischen Verbalpräfixen (BEÖTHY/ALTMANN 1984, 1985, 1986) und an der französischen Konjunktion *et* (ROTHE 1985) durchgeführt. Alle Resultate haben die Verteilung (1) überzeugend bestätigt.

Die vorliegende Untersuchung deutscher Wortbildungsaffixe zeigte zwar in vielen Fällen ebenfalls eine Übereinstimmung mit dem Modell; es gab aber auch etliche Fälle, die es falsifizierten.

Das Material wurde aus DW I-III (1973, 1975, 1978) entnommen. Aus mathematischen Gründen konnten nur solche Wortbildungselemente berücksichtigt werden, für die mindestens vier verschiedene Bedeutungen angegeben waren. Dabei wurde angenommen, daß immer dann, wenn in DW I-III durch hochgestellten Index eine besondere "Funktionsgruppe" (DW I, 144ff.) bzw. ein besonderer "transformationeller Wert" (DW II, 50ff.; DW III, 75ff.) angegeben ist, auch eine eigene Bedeutung des entsprechenden Affixes vorliegt. Die Parameter p und r aus (1) wurden zuerst mit klassischen Methoden geschätzt (Momentenmethode, maximum-likelihood), dann wurde die

Schätzung iterativ verbessert, um das kleinste Chiquadrat zu bekommen. Als Beispiel findet man in Tab. 1 die Anpassung an die Diversifikation des Affixes *-os/-ös* (DW III, 116).

Tab. 1: Häufigkeiten einzelner Bedeutungen von *-os/-ös*

Rang x	Leitformen	transformationeller Wert	Basiswortart	Frequenz f_x	berechnete Frequenz NP_x
1	ambitiöser Sportler	Sportler, der Ambitionen hat	S	59	58.92
2	gelatinöses Nebenprodukt	Nebenprodukt, das wie Gelatine ist	S	31	30.10
3	inzestuöse Verbindung	Verbindung, die ein Inzest ist	S	20	17.05
4	ruinöser Wettbewerb der Konzerne	Wettbew., der den Ruin bewirkt	S	6	9.97
5	tuberkulöse Erkrankung	Erkrankung, die die Tuberkel(bakterie)n hervorrufen	S	5	5.93
6	seröse Flüssigkeit	Flüssigkeit, die aus Serum besteht	S	4	3.55
7	religiöser Zweifel	Zweifel, der die Religion betrifft	S	2	2.14
8	venöses Blut	Blut, das sich in der Vene befindet	S	2	1.30
≥ 9	medikamentöse Behandlung	Behandlung, die durch Medikamente geschieht	S	2	2.03
$p = 0.3781$; $r = 0.8217$; $X^2_6 = 2.7082$; $P = 0.84$					

X^2 stellt hier den kleinsten Wert des Suchalgorithmus dar; P ist die Wahrscheinlichkeit des gewonnenen oder eines noch extremeren X^2-Wertes.

Die Werte der sechsten Spalte ergeben sich aus der Formel (1) mit den Parametern wie in der Legende angegeben. Der Chiquadrattest zeigt eine gute Übereinstimmung der Daten mit dem Modell.

In Tab. 2 sind alle Affixe aufgeführt, deren Verteilungen dem Modell folgen. Eine Verteilung entspricht dem Modell umso mehr, je höher ihr P-Wert ist, wobei $P \geq 0.05$ sein muß.

Tab. 2: Diversifikation deutscher Wortbildungsaffixe nach den Daten von DW I-III

Affix	Parameter p	Parameter r	Chiquadrat	FG	P	Anmerkung
-al/-ell	.2375	1.0490	10.57	13	0.65	1,2
-ar/-är	.2447	1.3177	2.87	12	0.996	
-bar	.1476	0.0486	3.45	3	0.33	2a
-(at)iv	.2185	0.8216	14.11	10	0.17	
-mäßig	.6596	2.3237	3.35	3	0.34	
-os/-ös	.3782	0.8217	2.71	6	0.84	
-sam	.3461	0.5167	1.55	2	0.46	
-fertig	.4614	0.5402	0.57	1	0.45	
-fest	.4923	0.8929	0.63	1	0.43	
-frei	.3533	0.6700	4.29	4	0.37	
-freundlich	.4117	0.5680	1.76	2	0.41	
-(is)ier-(en)	.2007	0.9425	15.26	14	0.36	
-(e)l-(n)	.2041	0.6833	1.43	6	0.96	1
auf-	.3656	1.1145	3.61	4	0.46	2b
über-	.4319	0.7941	3.27	1	0.07	
ein-	.4212	0.1514	4.34	2	0.11	
er-	.3406	0.8137	13.55	7	0.06	
ge--ø/-e	.3314	0.3252	5.04	3	0.17	
-(at)or	.5516	1.4322	2.98	1	0.08	
-enz/-anz	.3820	0.3779	1.14	2	0.57	
-(er)ei	.2643	0.4835	8.59	4	0.07	2b
-ler	.5385	1.0142	1.10	2	0.58	
-ling	.4138	1.3297	7.73	3	0.05	2a
-lich	.1656	1.0389	24.21	16	0.08	2b
-ik	.5706	1.7954	4.48	3	0.21	
-ismus	.1250	0.1227	12.84	9	0.17	1,2c
-nis	.3697	1.0663	6.28	3	0.10	
-tum	.3991	0.3306	0.83	2	0.66	1,2c

Anmerkungen:

1) Die "Restklasse" wurde in so viele Klassen der Häufigkeit 1 zerlegt, wie sie Fälle umfaßte.
2) Die letzten a) zwei, b) drei Klassen wurden zusammengefaßt; c) es erfolgten andere Zusammenfassungen. In Tab. 2 bedeutet FG: Freiheitsgrad und P die Wahrscheinlichkeit, mit der man einen noch extremeren als den beobachteten Wert erwarten konnte.

3. Wie man sieht, folgen 28 Affixe (Tab. 2) dem Modell, das Suffix *-ist* wurde nicht berücksichtigt, da für dieses u.a. ein ungefährer Wert angegeben ist (DW II, 82-83). Die Affixe *vor-*, *-chen*, *-ent/-ant* (Nomen), *-in* enthalten nur zwei Klassen und sind dadurch nicht prüfbar.

Die restlichen Verteilungen konnten mit der Formel (1) nicht erfaßt werden und werden unten untersucht. Ein solches Ergebnis ist ein Zeichen dafür, daß die Daten eine Eigenart haben, der im Modell nicht Rechnung getragen wurde. In einem solchen Fall kann man entweder das Modell verwerfen und nach einem neuen suchen, oder man versucht, es der Eigenart der Daten anzupassen. Wir haben in diesem Fall den zweiten Weg eingeschlagen.

Man kann unter diesen Umständen auf verschiedene Weise vorgehen.

(1) Wie man aus den Daten in DW I-III ersieht, dienen etliche Affixe dazu, um Wörter aus verschiedenen - statt sonst nur aus einer - Wortklasse(n) abzuleiten. So werden mit *-ant/-ent* Ableitungen von Wörtern gebildet, deren Grundwort der Klasse der Verben oder der Substantive angehört; mit *-ig* werden Wörter abgeleitet, deren Grundwörter aus der Klasse der Verben, Substantive, Adverbien, Adjektive oder Pronomina stammen; etc. Dies kann man als ein Anzeichen dafür werten, daß in solchen Fällen gleichzeitig mehrere Diversifikationsprozesse ablaufen und ihre Häufigkeitsränge sich daher überlagern, so daß man sie nicht mehr mit einem einfachen Modell erfassen kann. In einigen Fällen lassen sich gute Anpassungen dadurch erreichen, daß man die Verteilung nach der Basiswortform aufteilt. So ergibt sich z.B. bei *-ig* für Substantive als Basisform eine Verteilung mit $p = 0.1456$, $r = 0.1884$, $X^2 = 12.41$, $FG = 8$, $P = 0.13$, und für Verben als Basisform: $p = 0.4117$, $r = 0.2034$, $X^2 = 2.31$, $FG = 2$, $P = 0.31$; dies ist jedoch nicht in allen Fällen durchführbar und nicht überall sinnvoll.

Die einfachste, plausible Adaptation des Modells wäre nun eine Mischung von zwei oder mehr negativen Binomialverteilungen. Wir werden im Folgenden nur einige der vielen Möglichkeiten als Anregung für die weitere Forschung präsentieren, und zwar jeweils die nichtverschobene Form.

(a) Mischung zweier negativer Binomialverteilungen mit einem gleichen Parameter:

r gleichbleibend:

$$P_x = \alpha\binom{r+x-1}{x}p_1^r q_1^x + (1-\alpha)\binom{r+x-1}{x}p_2^r q_2^x$$

p gleichbleibend:

$$P_x = \alpha\binom{r_1+x-1}{x}p^{r_1}q^x + (1-\alpha)\binom{r_2+x-1}{x}p^{r_2}q^x.$$

(b) Beide Parameter ungleich:

$$P_x = \alpha\binom{r_1+x-1}{x}p_1^{r_1}q_1^x + (1-\alpha)\binom{r_2+x-1}{x}p_2^{r_2}q_2^x.$$

(c) Mischung von negativen Binomialverteilungen des Decay-Typs mit einfacher Verschiebung:

$$P_x = \alpha\binom{r+x-1}{x}p^rq^x + (1-\alpha)\binom{r+x-2}{x-1}p^rq^{x-1}.$$

Kombiniert man mehrere Glieder, so kann man sehr gute Approximationen erreichen. Man nimmt dabei jedoch in Kauf, daß mehr Parameter auftreten und dadurch Freiheitsgrade verloren gehen.

(2) Da eine Verteilung vollständig von den Parametern bestimmt wird, kann man davon ausgehen, daß die Abweichung von der theoretischen Kurve dadurch entsteht, daß einer oder beide Parameter der Verteilung nicht konstant sind, sondern als Variablen mit einer eigenen Verteilung zu betrachten sind. In diesem Fall erhält man:

(a) bei variablem r:

$$P_x = \sum_y\binom{ry+x-1}{x}p^{ry}q^xf(y) \quad \text{oder} \quad \int\binom{ry+x-1}{x}p^{ry}q^xf(y)\,dy$$

wobei f(y) die Wahrscheinlichkeitsfunktion von y ist;

(b) bei variablem p:

$$P_x = \int\binom{r+x-1}{x}(py)^r(1-py)^xf(y)\,dy,$$

wobei y üblicherweise eine stetige Variable mit $0 < yp < 1$ ist;

(c) bei variablem r und p z.B.:

$$P_x = \int\sum_y\binom{ry+x-1}{x}(pz)^{ry}(1-pz)^xf(y)f(z)\,dz.$$

(3) Eine dritte Möglichkeit ergibt sich dadurch, daß man die "anomalen" Umstände direkt in die Formeln des stochastischen Prozesses einsetzt und die so modifizierten Gleichungen löst.

4. Im Folgenden werden wir nur den Fall (1a) untersuchen. Wie oben gezeigt, sind in

$$P_x = \alpha\binom{r+x-1}{x}p_1^rq_1^x + (1-\alpha)\binom{r+x-1}{x}p_2^rq_2^x \qquad (7)$$

vier Parameter zu schätzen, was erhebliche Schwierigkeiten mit sich bringt (vgl. EVERITT & HAND, 1981). Wir benutzen die maximum-likelihood-Methode nach HASSELBLAD (1969) für p_1, p_2 und α; r ermitteln wir mit dem Davies-Swann-Campey-Algorithmus (vgl. JACOBY/KOWALIK/PIZZO 1972: 63). Notiert man:

$$P_{1x} = \binom{r+x-1}{x}p_1^rq_1^x \quad \text{und}$$
$$P_{2x} = \binom{r+x-1}{x}p_2^rq_2^x, \qquad (8)$$

so erhält man drei Gleichungen, die man zusammen mit r iterativ lösen kann, und zwar:

$$\hat{p}_1 = \frac{r\Sigma f_xP_{1x}/P_x}{\Sigma f_x(r+x)P_{1x}/P_x}$$

$$\hat{p}_2 = \frac{r\Sigma f_xP_{2x}/P_x}{\Sigma f_x(r+x)P_{2x}/P_x}$$

$$\hat{\alpha} = \frac{\alpha}{N}\Sigma f_xP_{1x}/P_x.$$

Die iterative Lösung mit sich anpassendem r ist wegen ständig neuer Berechnung der Wahrscheinlichkeiten recht langwierig, bietet aber - wenn die Verteilung geeignet ist - meistens bessere Resultate als die einfache negative Binomialverteilung. Einfachheitshalber haben wir hier die nichtverschobene Verteilung präsentiert (d.h. für x = 0, 1, ...), die man auch als eine einfache lineare Transformation der Variablen (z.B. y = 1, 2, ...) betrachten kann, und zwar x = y-1. Als Beispiel für eine Anpassung vgl. die beobachte-

ten und die berechneten Häufigkeiten für die Diversifikation von *-al/-ell* in Tab. 3.

Die gemischte negative Binomialverteilung hat den Nachteil, daß man vier Parameter hat und dadurch fünf Freiheitsgrade verliert, so daß man eine Anpassung nur bei solchen Verteilungen überprüfen kann, die mindestens sechs Häufigkeitsklassen aufweisen. Im obigen Beispiel (Tab. 3) sieht man, daß die Anpassung - wie zu erwarten - besser ist als bei der einfachen negativen Binomialverteilung.

Tab. 3: Anpassung der gemischten negativen Binomialverteilung an die Häufigkeiten von *-al/-ell*

x	f_x	NP_x einfache NBV	NP_x gemischte NBV
1	93	89.21	95.59
2	68	71.35	72.92
3	56	55.74	55.62
4	46	43.19	42.43
5	32	33.34	32.36
6	23	25.67	24.69
7	20	19.73	18.83
8	19	15.15	14.36
9	11	11.62	10.96
10	7	8.91	8.36
11	7	6.83	6.38
12	6	5.23	4.86
13	6	4.00	3.71
14	4	3.06	2.83
15	4	2.34	2.16
16≥	1	7.61	1.65
		p = 0.2375 r = 1.0490 X^2_{13} = 10.57 P = 0.65	p_1 = 0.2376 p_2 = 0.2369 r = 1 α = 0.3813 X^2_{11} = 6.66 P = 0.83

In Tab. 4 findet man diejenigen Affixe, deren Verteilungen nur mit der gemischten negativen Binomialverteilung erfaßt werden konnten.

Tab. 4: Anpassung der gemischten negativen Binomialverteilung an die Diversifikation deutscher Affixe (Daten nach DW I-III)

Affix	p_1	p_2	r	α	X^2	FG	P
-ant/-ent	.9757	.8332	20.70	.9373	3.72	1	0.05
-haft	.9993	.6587	3.50	.6295	3.04	3	0.39
-isch	.9613	.7690	14.00	.4445	14.80	8	0.06
-ø-(en)	.1432	.1432	1.12	.3000	26.72	22	0.22
-e	.9776	.7830	10.45	.4607	6.04	3	0.11
-schaft	.9357	.5122	8.83	.9125	8.26	8	0.41

Bei den restlichen 15 Verteilungen in DW I-III konnte eine Übereinstimmung mit dem Modell nicht erreicht werden. Das hat folgende Gründe:

(1) Beim Affix *-ig-(en)* gibt es zwar vier Häufigkeitsklassen, aber bei der Anpassung ergibt sich, daß die theoretische Häufigkeit der letzten Klasse für den Test zu klein ist (s. Tab. 5).

Tab. 5: Anpassung der negativen Binomialverteilung für *-ig-(en)*

x	1	2	3	4
f_x	6	3	1	1
NP_x	6.00	3.06	1.24	0.70

p = 0.7019

r = 1.7124

Die Anpassung ist zwar sehr gut; aber wenn man die Klassen x=3 und x=4 zusammenfaßt, bleiben keine Freiheitsgrade übrig. Ohne Zusammenfassung ergibt sich ein X^2_1 = 0.003, was einer Wahrscheinlichkeit von P = 0.96 entspricht. Diesen Fall kann man daher als mit dem Modell übereinstimmend betrachten.

(2) Von den übrigen 14 Verteilungen haben neun (*ab-, an-, be-, durch-, um-, -ie, -ist*) lediglich vier oder fünf Häufigkeitsklassen, so daß man die gemischte Binomialverteilung, die ja vier Parameter hat, nicht anpassen kann. Zum Suffix *-ist* s.o.

(3) Lediglich sieben Verteilungen bleiben noch als Problemfälle:

-ig (26 Häufigkeitsklassen; Ableitungen aus 5 Wortklassen)

auf- (9/3)

ver- (11/3)
er- (19/3)
-heit (10/3)
-ung (16/2)
-Ø (7/1).

Hier müßte man bei *-ig* eine fünffach gemischte, bei *auf-*, *ver-*, *-er* und *-heit* eine dreifach gemischte Verteilung anwenden. Das *-Ø*-Affix erscheint in Kombination mit vielen verschiedenen Präfixen (*Ab*stoß, *Auf*bruch, *Ein*stieg, *Ent*gelt, *Be*leg, *An*hang, *Nach*wuchs, usw.), was bei der Zusammenstellung der Daten nicht in Betracht gezogen wurde. Eine entsprechende Aufteilung der angeführten Daten könnte evtl. zu einem positiven Resultat führen. Diese Überlegung läßt sich auf das Suffix *-ung* bzw. auf alle Suffixe ausdehnen. Bevor nun eine entsprechende Aufteilung der Daten durchgeführt ist, lohnt sich die theoretische Weiterentwicklung des Modells nicht; wahrscheinlich erübrigt sich auch die Einsetzung von mehrfach gemischten negativen Binomialverteilungen.

5. Abschließend läßt sich folgendes sagen: Die "klassische" linguistische Arbeit, die Festsetzung der Kategorien, die Begriffsbildung, die Feststellung der Bedeutungen aufgrund von Transformationen oder auf anderen Wegen usw. ist nur ein Teil der Sprachanalyse, und zwar der statische, eher für die Praxis geeignete. Der theoretische Teil der Arbeit geht von der Annahme einer Dynamik in der Sprache aus, von den Kräften nämlich, die innerhalb der Sprache wirken, und sagt die Resultate ihres Wirkens voraus. Eine Nichtübereinstimmung mit dem Modell kann verschiedene Ursachen haben: Unüberprüfbarkeit der Daten - das ist keine Nichtübereinstimmung im eigentlichen Sinne -, Fehler im Modell, oder, was im hier behandelten Fall besonders markant ist, eine z.T. falsche Datengewinnung. Diese Daten sind ja nicht etwa vorgegeben, sondern man macht Beobachtungen sprachlicher Phänomene zu Daten, und zwar im Licht seiner eigenen Begriffe oder im Licht einer Theorie, sobald man eine entwickelt hat. Falls diese Begriffe nicht korrekt gewählt wurden, kann eine Übereinstimmung der Daten mit dem Modell nur ein Glücksfall sein. Im Fall der deutschen Wortbildungsaffixe wäre es vermutlich besser, jede Kombination eines Präfixes mit einem Suffix als eigene Wortbildungskategorie aufzufassen und danach die entsprechenden Häufigkeiten zu ermitteln. Auf diese Weise kann ein mathematisches Modell auch als Kontrollinstanz für die Adäquatheit einer linguistischen Analyse dienen.

LITERATUR

ALTMANN, G., "Semantische Diversifikation." Folia Linguistica 19, 1985, 177-200

BEÖTHY, E./ ALTMANN, G., "The diversification of meaning of Hungarian verbal prefixes. I. "meg-"."Nyelvtudományi Közlemények 1986 (erscheint)

BEÖTHY, E./ ALTMANN, G., "The diversification of meaning of Hungarian verbal prefixes. II. "ki-"." Finnisch- Ungarische Mitteilungen 8, 1984, 29-37

BEÖTHY, E./ ALTMANN, G., "The diversification of meaning of Hungarian verbal prefixes. III. "el-", "föl-", "be-"." Glottometrika 7. Ed. by U. Rothe. Bochum 1985, 46-56

DW: Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. 3 Bde. Düsseldorf: 1973, 1975, 1978 (= Sprache der Gegenwart 29, 32, 43)

EVERITT, B.S. / HAND, D.J., Finite mexture distributions. London/ New York, 1981

JACOBY, S.L.S. / KOWALIK, J.S. / PIZZO, J.T., Iterative methods for nonlinear optimization problems. Englewood Cliffs: Prentice-Hall, 1972

ROTHE, U., Die Semantik des textuellen *et*. Bochum, Diss. 1985

SCHWEIGER, E., persönliche Mitteilung (1986)

THE RELATIONSHIPS BETWEEN DEGREES OF CONTRAST IN RHYTHMIC STRUCTURES

M.A. Krasnoperova

I. STATEMENT OF PURPOSE

The problem dealt with in the present paper arises out of a more general one; namely, how to construct a model explaining the process by which rhythmic structures in a poetic text are perceived. An important step in constructing such a model can be considered the description of the rhythmic interaction between adjacent lines of verse, especially the description of the interactional process between the rhythmic structures beginning each line.[1] This also represents the basic approach used in this paper.

It is a reasonable assumption that there are pairs of initial rhythmic structures at the input end of the process which we wish to describe, and specific perceptions at the output end of it. Such perceptions result from the interaction of the elements within the pairs (rhythmic effects):

Sbežali | mutnymi ruč'jami
U ⏑ U
Na potoplennye | luga.
U – U ⏑ U –

If it were possible to describe properly the input and output ends of this process, the whole problem could be reduced to an explanation of how phenomena corresponding to the descriptions of the effects of the rhythmic structures arise on the basis of phenomena corresponding to the description of each pair. This explanation would in fact represent the sought-after model.

It is in fact reasonable to assume that an adequate description of rhythmic structures is feasible. In regard to the effects of such structures, however, the problem becomes more complicated. One can assume that the description of the process generating rhythmic effects will ultimately also represent the description of the selfsame effects.[2] In this fashion a vicious circle arises: in

constructing such a model it is necessary to describe rhythmic effects, but the description is itself based on the original model. As a result of this circumstance, it is necessary to seek intermediate means of description. We wish to carry out this intermediate description by using certain formal characteristics which have the ability to reproduce some of the impressions that rhythmic effects make. Finally, each rhythmic structure representing an initial pair will be provided with a system displaying the relation of its effect with the effects of other structures. This model should then serve to explain how these relationships between rhythmic effects can be obtained under given relationships between the rhythmic structures.

Since the comparison of rhythmic structures on the basis of selected characteristics must be based on relatively specialized research, it is necessary further to narrow the problem area before us. Thus, in this paper we have limited our research object to the comparison of the contrastive degrees of the effects of rhythmic structures beginning paired iambic lines.[3]

II. THE RHYTHMIC STRUCTURES

The basic units of our comparative approach are the various concrete manifestations of the following rhythmic structures: those with the first metrical stress in the first rhythmic unit and without a metrical stress in the second unit: ($M\bar{M}$: ⏑́ ⏑ ⏑́ etc.), those with identical pre-stress lengths of both units and a greater post-stress length in the first unit ($M_1 > M_2$: ⏑́ ⏑ ⏑ etc.) and those having an arrangement opposite to the ones already mentioned ($\bar{M}M$, $M_1 < M_2$). The structures listed above will be considered "basic" ones. As auxiliary points of reference we have employed a structure with an additional stress on the first rhythmic unit and with the first metrical stress on the second unit (DMM: ⏑́ ⏑́ etc.), a structure with the opposite distribution of stress (MDM), and a structure with identical units (OD ⏑́ ⏑ etc.). $M\bar{M}$, $M_1 > M_2$ and DMM have been designated as structures with a first-position more weighted beginning and the opposite cases as structures with a second-position more weighted beginning.

The concrete form of any designated rhythmic structure is called an elementary variant. For further brevity of description

two numerals will be said to correspond to each unit of an elementary variant: the first numeral denotes the number of pre-stress syllables, the second, the number of post-stress syllables. A completely elementary variant will be described by using four numerals with a period between them: ∪́∪ = 10.11; ∪́∪́ = 30.11. Any set of the elementary variants of one and the same structure will be designated as the structure's variant. To describe non-elementary variants, a letter will be places in the variable part of each unit: 10.1x, 3x.11, etc. In such a case the numerals denote the common units of the elementary variants making up the given non-elementary variant, and the letter symbolizes the part differing from the rest: {10.11, 10.12} -10.1x etc. If the meanings of the unknown quantity are not given special mention, it is assumed that the set includes all elementary variants of the given type.

Further on in this study two sets of structures will also be dealt with: ODČ (in which the constituent parts of each variant have an identical pre-stress length) and RDČ, (in which the constituent parts have a different pre-stress length).

III. THE "RHYTHMIC CHARACTERISTICS" UNDER STUDY

The most adequate definition of the degree of contrast possessed by the effect of a given rhythmic structure would be one based on a model which describes the interaction of a particular structure's units. We will first seek to convey the substance of this characteristic in both explanatory and axiomatic fashion. When referring to it we will use abbreviated phrases: "degree of contrast of a structure", "contrast between structures" etc.

Explanation

A structure's degree of contrast is characterized by our ability to perceive a difference between the two units, felt at the second of them:

10.32	Odnim	na vremja očarovan
	Razočarovannyj	drugim.

In order to express the fact that one structure shows a greater degree of contrast than another, we have placed the symbol for such structures in quotation marks and separated them with the sign <or> according to the direction of the relationship: thus, "10.30">"10.11" means that 10.30 has a greater degree of contrast than 10.11; "11.10"<"10.11" that 11.10 has a lesser degree of contrast than 10.11. If the symbol for a non-elementary variant appears in quotation marks, it means that any given elementary variant corresponding to the non-elementary variant is in the given relation with the indicated structure: "10.3x"> "10.11". The same applies for entries in which the symbol for a basic structure appears in quotation marks.

Axioms

1) The structure OD has a zero degree of contrast:

10.10	Vnušat'	ljubov' dlja nix beda
	Pugat'	ljudej dlja nix otrada

All remaining axioms have been formulated for the structures designated "basic".

2) If the rhythmic structures happen to coincide with one of the rhythmic units, then the structure with the greatest degree of contrast will be the one containing a strong non-accented position in another rhythmic unit (this axiom does not include the relations between $M_1 > M_2$ and $M_1 < M_2$ and does not pertain to the structures without an initial metrical stress in both rhythmic units).

Example: "11.30">"11.10"

11.30	Bezmolvno I o bylom	budu ja zevat' vospominat'
11.10	Derevnja, Byla	gde skučal Evgenij, prelestnyj ugolok.

3) In all variants of one and the same structure having a common rhythmic unit, the degree of contrast relationships are defined in identical fashion according to the unit differing from the others.

Example: If in the case of 3x.10 it turns out that "32.10"> "31.10">"30.10", then in the case of 3x.11 there will be an analogous relationship: "32.11">"31.11">"30.11".

4) If the variants of MM̄ have an identical initial rhythmic unit, the variant with the highest degree of contrast is the one having the longer post-stress part in the second unit:

Example: "10.32">"10.30".

10.32	{ Odnim { Razočarovannyj	na vremja očarovan drugim
10.30	{ S bol'nym { Ne otxodja	sidet' i den' i noč' ni šagu proč'.

5) Of two variants of one and the same rhythmic structure with an identical pre-stress length of rhythmic units, the variant considered to have a higher degree of contrast is the one in which the difference in the post-stress length is greater.

Examples: "12.10">"12.11"

12.10	{ Blistatel'na, { Smyčku	poluvozdušna, volšebnomu poslušna
12.11	{ Nepravil'nyj, { Netočnyj	nebrežnyj lepet vygovor rečej.

6) "11.12">"10.11"

11.12	{ Javljatsja { Vnimatel'nym	gordym i poslušnym, il' ravnodušnym.
10.11	{ Pugat' { Prijatnoj	otčajan'em gotovym, lest'ju zabavljat'.

7) A rhythmic structure with a relatively high degree of contrast is likely to give rise to a rhythmic effect with a relatively high degree of contrast.

V. METHODOLOGY, CONDITIONS, FAVOURABLE FOR JUDGING THE CONTRAST

We compared Tomaševskij's well-known data on the strength of interlinear pauses in a stanza of Eugene Onegin with our own research regarding the use of the rhythmic structures OD, ODČ, and

RDČ in the same stanza. The results show that the least commonly occurring pause falls on areas with the greatest concentration of ODČ. This fact can be explained in the following manner: ODČ, as opposed to OD, has a non-zero contrast and in comparison to RDČ is for the most part characterized by variants with relatively weak contrast (10.11, 11.10). As a result, the poet increases the contrast of the generated effect in ODČ more often than in other structures. This results in a decrease of average interlinear pauses.[4]

These results permit us to conclude that the parts of the text differing in the concentration of certain interlinear pauses can be regarded as a favorable basis for evaluating contrasts.[5] Our further research is thus based on this material. In this case, the text analyzed is "Eugene Onegin".[6]

Mathematical Procedure

Following Tomaševskij, we have differentiated pauses according to the presence and/or type of punctuation marks occurring in the text: 0 = absence of a punctuation mark, 1 = comma, 2 = semicolon, 3 = period. The mathematical analysis is based on Spearman's rank correlation coefficient (R).[7]

We have applied this coefficient as follows: the text is divided into parts containing distichs formed by even and uneven lines of a stanza: 1-2, 3-4, 5-6, 7-8, etc. In each section we have calculated the frequencies of the analyzed structures and pauses.[8] Both these frequencies were then ranked according to their degree of diminution. The rankings obtained for the same parts were then taken in pairs. After this, R is determined. If R's absolute value does not exceed the critical value, it will be assumed that there is no correlation between the given structure and the given pause. In the opposite case, the correlation is positive or negative according to the symbol R. The absolute value of the critical value is equal to 0.75 at the designated level of permissible error (0.066) and with a given number of selected parts (seven). R is defined using a special table which supplies the information needed for small samples.[9]

Further Elaboration of the Method

Not only individual rhythmic structures, but also some of their sets were tested for their correlation with pauses. If one of these sets happened to show a dependency relationship not apparent in its individual elements, it was assumed that every rhythmic structure included in the set had a tendency to "connect with" or to "avoid" a given pause in accordance with the aforementioned dependency relationship.

VI. BASIC DATA

The study was carried out in the following fashion: first of all, ℝ was defined for the basic structures and the pauses taken both singly and combined $\{0,1\}$, $\{2,3\}$. Those rhythmic structures that did not show some sort of correlation were grouped in sets together with those basic and auxiliary rhythmic structures which did not demonstrate a similar dependency either singly or in combination. We continued to build sets until the appropriate correlation appeared or until all instances which could lead to its appearance were exhausted.

The following meaning of ℝ was obtained for the rhythmic structures and sets showing correlation with pauses:

Table 1:

Pauses \ Structures	$M\bar{M}$	$M_1<M_2$	$\{\bar{M}M, DMM\}$	$\{\bar{M}M, DMM, MDM, M_1>M_2\}$
0	+0,89	+0,64	-0,82	-0,93
1	-0,86	-0,29	+0,39	+0,75
$\{0,1\}$	+0,82	+0,78	-0,93	-0,86

The following table describes the relationships between basic structures and pauses.

Table 2:

Pauses \ Structures	$M\bar{M}$	$\bar{M}M$	$M_1 > M_2$	$M_1 < M_2$
0	+	$\exists^-$	~	~
1	-	$\exists^+$	$\exists^+$	~
$\{0,1\}$	+	$\exists^-$	~	+

Symbols used: + = positive correlation, - = negative correlation, $\exists^+$ = connective tendency, $\exists^-$ = avoidance tendency, ~ = lack of any tendency.[10] From now on we will designate a relationship with a given pause with the appropriate symbol, which is palced in front of the numeral indicating the nature of the pause: + 0, + 1, $\exists^+0$ etc. Analogous to that relationship: $+\{0,1\}$, $\exists^+\{0,1\}$ etc. Since $+\{0,1\}$ is equivalent to $-\{2,3\}$ and $\exists^+\{0,1\}$ is equivalent to $\exists^-\{2,3\}$ etc., only the symbols with $\{0,1\}$ will be used.

We would like to set forth the following hypothesis: the analytically revealed relationships with pauses determine the relationships between the degrees of contrast. The tendency to form a connective relationship and positive correlation with short pauses shows an inclination to heighten the contrast within the rhythmic structure, whereas the opposite tendencies tend to weaken it.

We have tested this hypothesis by comparing it respectively with four alternative ones. If in the course of the analytical process a given alternative proved to be correct, the following alternative was compared with it, and not with the original hypothesis.

VII. TESTING ALTERNATIVE 1

Statement: the analytically revealed relationships with pauses have resulted from the general norms of the language (Russian).

To test this alternative, the prose of Puškin was used. The empirical frequencies found in the texts were compared with the apriori frequencies whose calculation was based on the assumption that there is no correlation between rhythmic structures and pauses

(Table 3). As we have already ascertained in an earlier study, the distribution of structures beginning adjacent segments of four-foot iambic meter in Puškin's prose depends on whether the narrator is the author or the hero.[11] Therefore both aggregate and disaggregated data on these groups of works was examined. No correlation was found in any of the cases. On the basis of this data, the alternative can be considered false.

Table 3: Relative Distribution of Structures and Pauses in Puškin's Prose (Joint Results)

Pauses \ Structures	$M_1 < M_2$		$M\bar{M}$		$\bar{M}M, DMM$		$\bar{M}M, DMM, MDM, M_1>M_2$	
	e	a	e	a	e	a	e	a
0	51	48,9	29	30,7	44	41,9	111	108,6
1	9	9,8	7	6,5	8	8,6	15	21,7
{0,1}	60	58,7	36	37,2	52	50,5	126	130,3

VIII. TESTING ALTERNATIVE 2

Statement: the tendency to combine with short pauses is caused by the desire to weaken the overly strong contrast of the structure's second unit in regard to the broader rhythmic context.

A complete set of analyzed distichs was designated "first-group distichs". The portion containing two lines without a punctuation mark in the initial lines was designated "second-group distichs". In the case of an identical interlinear pause the interval between the two rhythmic units of our structures turned out, on the average, to be larger in the first group. On the basis of this observation one can assume that the second group presents a more favorable basis for weakening contrast between the second rhythmic units and the broader rhythmic context. This means, that in the case of a correctly postulated alternative, the second-group distichs do not show a tendency to intensify connections with shorter pauses. The testing of the second-group distichs was carried out in the same manner as that of the first group. None of the structures, when

taken separately, showed any sort of correlation. For the sets showing correlation with pauses, the following values of R were obtained.

Table 4:

Pauses \ Structures	$\{M\bar{M}, M_1<M_2\}$	$\{M_1>M_2, M_1<M_2\}$	$\{DMM, \bar{M}M\}$	$\{DMM, \bar{M}M, M_1>M_2\}$
0	+0,82	+0,71	-0,79	-0,61
1	-0,86	-0,25	+0,71	+0,79
{0,1}	+0,86	+0,82	-0,82	-0,57

In Table 5, the data obtained for various groups of distichs are compared.

Table 5:

Pauses \ Structures	0		1		{0,1}	
	1 gr.	2 gr.	1 gr.	2 gr.	1 gr.	2 gr.
$M\bar{M}$	+	$\exists^+$	-	$\exists^-$	+	$\exists^+$
$\bar{M}M$	$\exists^-$	$\exists^-$	$\exists^+$	$\exists^+$	$\exists$-	$\exists^-$
$M_1>M_2$	~	~	$\exists^+$	$\exists^+$	~	$\exists^+$
$M_1<M_2$	~	$\exists^+$	~	$\exists^-$	+	$\exists^+$

The results of the comparison reveal a tendency towards a strengthening of connections with short pauses in the second group. In the case of $M_1<M_2$ and $M_1>M_2$ the following correlations were obtained: for $M_1<M_2$ we get $\exists^+0$, and for $M_1>M_2$ we get $\exists^+\{0,1\}$. In the variants of these structures the tendency towards strengthening the connections with short pauses is even more apparent: in the case of 10.11, the relationship +0 results, in the case of {12.10, 12.11} we have +{0,1} (Table 6). This allows us to regard this alternative as false.

However, such an alternative must still be considered unrefuted if its sphere of influence is limited only to structures without an initial metrical stress in one of their rhythmic units ($M\bar{M}$, $\bar{M}M$).

We wish to test this alternative by applying it individually to each of the rhythmic structures we have already set up.

$M\bar{M}$. If our initial hypothesis is true, then on the basis of axiom 4 variants with a shorter post-stress part in the second rhythmic unit should correlate with short pauses at least as strongly as any others. In the opposite case, the same holds true for the variants with a longer post-stress part in the second rhythmic unit.

The results of our analysis allow us to conclude that a tendency to weaken the interactional effects with a broader context has at least been demonstrated for 10.3x, and a tendency to heighten the contrast within the structure in the case of 11.3x.

In regard to 10.3x, we based our conclusion on the fact that in testing the ability of every combination of its elementary variants to connect with pauses 10.30 was the only one to reveal no positive tendency for 0 or for $\{0,1\}$.

In the case of 11.3x our conclusion was based on the observation that in analogous tests of its elementary variants and their sets containing 10.3x only 11.30 showed both such tendencies, whereas the other variants showed, respectively, one of each (see Table 6).[12]

$\bar{M}M$. Taken as a whole, $\bar{M}M$ showed both $\exists^{+}1$ and $\exists^{+}\{2,3\}(\exists^{-}\{0,1\})$. If we substitute this structure by its variants in the set where it reveals the relationship $\exists^{+}\{2,3\}$, this tendency becomes apparent with any type of the second rhythmic units (see Table 6). It thus follows that $\bar{M}M$ at least does not avoid connections of its second rhythmic unit with a broader context. We can thus regard this alternative as false.

IX. TESTING ALTERNATIVE 3

Statement: the tendency to combine with long pauses indicates weakness and not strength of the respective contrasts.

Of all the basic structures, only $\bar{M}M$ tends toward combination with long pauses. As was already mentioned (VIII, $\bar{M}M$), this tendency can be observed in this structure regardless of the type of the second rhythmic unit. In the case of 3x.11 and 3x.12 the alternative

is invalid. Indeed, it follows from axiom 2 that "10.11" < $\{3x.11, 3x.12\}$ (see Table 6). 10.11 shows $+\{0,1\}$. If 10.11 has a degree of contrast strong enough not to be "dislodged" from connections with short pauses, the same applies all the more in the case of 3x.11 and 3x.12. In the case of 3x.10 the alternative was confirmed. The basis for this conclusion are the results of the following analysis (see footnote 14).

X. TESTING ALTERNATIVE 4

Statement: The analytically revealed relationships of rhythmic structures to pauses demonstrate only a general tendency to make prominent various kinds of effects, but does not tell us anything about their degree of contrast.

We began testing Alternative 4 by studying the relationships of all other remaining rhythmic structures with structures showing $+\{0,1\}$. Except for some insignificant corrections, the method used is analogous to the previous one. The results revealed only negative correlations. $M_1<M_2$ correlated exclusively with the set of structures that also had a more-weighted beginning in the second position ($\{\bar{M}M, MDM\}$, R = -0.79). $M\bar{M}$ correlated only with the sets having rhythmic structures characterized by the same position of the more-weighted beginnings ($\{\bar{M}M, MDM, DMM\}$, $\{\bar{M}M, MDM, DMM, M_1>M_2\}$, R = -0.75 and -0.86 respectively).

Each rhythmic structure tending to combine with $\{0,1\}$ was compared for degree of contrast with the structures having an analogous beginning arrangement. The results of the each comparison show that the initial structure mostly has a non-identical degree of contrast with the structures it has been compared to. Thus 1) 10.11 makes up 85% of all realizations of $M_1<M_2$. "10.11" < $\{3x.11, 3x.12\}$ in accordance with Axiom 2. "10.11" > $\{30.10, 31.10\}$ on the basis of the fact that 10.11 shows $+\{0,1\}$ in the first-group distichs while in the case of 30.10 and 31.10 a positive correlation with a short pause occurs only in the second-group distichs (see Table 6). $\{30.10, 31.10, 3x.11, 3x.12\}$ make up 56% of the rhythmic structures compared with $M_1<M_2$. 2) $\{11.3x, 12.3x\}$ > "11.10" as a result of Axiom 2. $\{11.3x, 12.3x\}$ make up 56% of all realizations

of $M\bar{M}$; 11.10 makes up 53% of all structures compared with $M\bar{M}$.

It can thus be seen, that the rhythmic structures not tending to combine with $\{0,1\}$ get forced out of that set of pauses primarily by the rhythmic structure having an identical more-weighted beginning and, a non-identical degree of contrast. On the basis of this information one can conclude that the "forcing-out" structure is used in this case as the optimal expression (on the basis of the degree of contrast) of that general quality contained in all structures having a similar beginning arrangement. This leads us to conclude that the proposed alternative is invalid.

XI. INTERPRETATION OF RESULTS

DEFINITION OF AN OPTIMAL PAUSE

In the second-group distichs $M_1<M_2$ weakens the connections with 1 and strengthens the connections with 0. This rhythmic structure uses pauses to regulate the contrasts between its units. The contrast in this particular instance is not too weak even in the first-group distichs. Therefore, under more favorable conditions, the weakening of the connection should above all result through the pause which contributes more than all others to the heightening of the degree of contrast. It thus follows that the pause having this quality is 1. This means that in case the initial hypothesis is valid, the rhythmic structure tending to connect with 1 should have a lower degree of contrast than any given structure tending to connect with other pauses.

RELATIONSHIPS BETWEEN DEGREES OF CONTRAST

The relationships between the degrees of contrast were defined on the basis of results obtained in the course of analysis, as well as on the basis of the axioms and some additional information obtained for the variants (see Table 6). Data on the variants was set up either on the basis of the calculations of ℝ for the variants themselves or for their sets, or on the basis of the calculations of ℝ for the sets formed from these sets in which the basic rhythmic structures (which include the given variants) showed some kind of tendency in relation to pauses. The new sets were thus formed by substituting variants for the basic structures.

Table 6:

Structures \ Pauses	group number	0	1	{0,1}
{10.30, 10.31}	1	+0,36	-0,72	+0,28
{10.30, 10.32}	1	-0,25	0,00	-0,18
{10.31, 10.32}	1	+0,78	-0,73	+0,86
{10.3x, 11.30, 11.31}	1	+0,71	-0,75	+0,78
{10.3x, 11.30, 11.32}	1	+0,78	-0,64	+0,68
{10.3x, 11.31, 11.32}	1	+0,49	-0,61	+0,49
10.11	1	+0,65	-0,40	+0,76
	2	+0,79	-0,44	+0,88
{12.10, 12.11}	1	+0,50	+0,03	+0,53
	2	+0,68	-0,72	+0,75
{11.10, 3x.3y}	1	-0,03	+0,036	+0,07
	2	-0,68	+0,75	-0,53
{30.10, 31.10}	1	-0,03	+0,00	+0,07
	2	-0,71	+0,75	-0,53
{$M_1>M_2$, DMM, MDM, 32.10, 33.10}	1	-0,75	+0,75	-0,61
3x.10	1	-0,71	+0,68	-0,75
{DMM, 3x.11, 3x.12}	1	-0,61	+0,21	-0,75

Explanation of Table 6

1) Auxiliary results indicating the absence of independent correlations of variants constituting parts of those given in the table have been omitted.

2) Of the combinations including the variants not belonging to $M\bar{M}$, only those have been listed which are dependent on pauses.

3) It is assumed that in the variant 3x.3y x is greater than y.

4) The variant 33.10 appears only three times in the texts analyzed. The data on this variant has everywhere been combined

with that on 32.10, and all conclusions have been made only in regard to the latter.

In the following section we have listed the inequalities characterizing the relationships between contrasts and the reasons for assuming their validity.

$M\bar{M}$. 1) "1y.3x" > "1y.3z", where x is greater than z. Reason: Axiom 4.

2) "12.3x" > "11.3x" > "10.3x". Reason: 12.3x > 11.3x" according to Axiom 2. Since in 10.3x the variant with the least degree of contrast weakens the tendency towards connection between the elements of this structure, and strengthens it in 11.3x (see VIII, $M\bar{M}$), we may assert, in accordance with Axiom 3, that in $M\bar{M}$ 11 contributes to a stronger degree of contrast than 10. Hence, "11.3x > "10.3x".

3) "10.30" = "min $M\bar{M}$".[13] Reason: "1y.30" = "min 1y.3x" according to inequality 1. "10.30" < "11.30" according to inequality 2. "10.30"< "12.30" according to Axiom 2. Hence, "10.30" = "min $M\bar{M}$".

4) "$M\bar{M}$" > "10.11". Reason: "10.11" < "10.30" according to Axiom 2.

$\bar{M}M$. 5) "32.1x" > "$\{31.1x, 30.1x\}$". Reason: both 30.10 and 31.10 display no positive tendency to short pauses in the first-group distichs and display $\exists^{+}1$ in the second group. Since 32.10 already dispalys the same tendency in the first group (see Table 6), we may consider "32.10" >"$\{31.10, 30.10\}$". Thus, according to Axiom 3, the initial inequality is proved.

6) "$\{3x.12, 3x.11\}$" > "10.11". Reason: Axiom 2.

7) "3x.10" < "10.11". Reason: of the positive tendencies to short pauses, 32.10 shows $\exists^{+}1$ in the first-group distichs, 10.11 shows $+\{0,1\}$. Hence, "32.10" < "10.11". The rest follows from inequality 5.[14]

8) "3x.12" > "3x.11" > "3x.10". Reason: Axiom 2, inequalities 6 and 7.

9) "3x.10" < "$M\bar{M}$". Reason: inequalities 4 and 7.

10) "$\{32.12, 32.11\}$" > "$\{11.30, 10.30\}$". Reason: "3x.11" > "11.30" because of the fact that 3x.11 displays $\exists^{-}\{0,1\}$ with the purpose of weakening the contrast between the structures themselves (see IX), 11.30 shows $\exists^{+}\{0,1\}$ with the purpose of intensifying the contrast (see VIII, $M\bar{M}$). The rest follows from inequalities 5, 8 and 3.

$M_1<M_2$. 11) "10.12" > "11.12" > "10.11". Reason: Axioms 5 and 6.

12) "$M_1 < M_2$" > "12.10". Reason: "10.11" > "12.10" because of the fact that 10.11 displays +0 in the second-group distichs, whereas 12.10 shows only $\exists^{+}\{0,1\}$. The rest follows from inequality 11.

13) "$M_1 < M_2$" > "3x.10". Reason: inequalities 7 and 11.

14) "12.10" > "12.11" > "11.10". Reason: "12.11" > "11.10" owing to the fact that in the second-group distichs 12.11 displays $\exists^{+}\{0,1\}$, whereas 11.10 shows only $\exists^{+}$ 1. Hence, owing to Axiom 5, the initial inequality is valid.

15) "$M_1 > M_2$" < "$M_1 < M_2$". Reason: inequalities 12 and 14.

16) "11.10" < "$\bar{M}M$". Reason: "11.10" < "3x.10" according to Axiom 2. The rest follows from inequality 8.

17) "$\{3x.11, 3x.12\}$" > "$M_1 > M_2$". Reason: inequalities 6 and 15.

18) "$\{12.10, 12.11\}$" > "3x.10". Reason: "12.11"> "32.10" because of the fact that 12.11 displays $^{+}\{0,1\}$ in the second-group distichs, whereas 32.12 shows no positive tendencies to short pauses except $\exists^{+}1$. The rest follows from inequalities 5 and 14.

19) "$M_1 > M_2$" < "$M\bar{M}$". Reason: inequalities 4 and 15.

The inequalities found in the course of our analysis allow us to conclude that the variant with the least degree of contrast is 11.10. The variant with the highest degree of contrast belongs to the set $\{12.32, 32.12\}$.

The relationships between degrees of contrast not established in this paper may be found by taking into account additional material.

REMARKS

1 The initial concepts can be found in my article "Ob odnom podxode k opredeleniju effektov ritmičeskix struktur" ("On an Approach to Defining the Effects of Rhythmic Structures") in Lingvističeskie problemy funkcional'nogo modelirovanija rečevoj dejatel'nosti, 2nd edition, Leningrad, 1974, pp. 66 and 69.

2 Ibid, p. 68

3 We have examined only the structures occuring with sufficient frequency in classical iambic meter. The pre-stress length of any unit of such a structure does not exceed three, the post-stress length does not exceed two or three for the pre-stress part having respectively one or more syllables.

4 For the necessary data see Boris Tomaševskij's "Strofika Puškina" ("Puškin's Strophics") in Stix i jazyk, Moskva-Leningrad 1959, as well as my article "Zamečanija o ritme strofy 'Evgenija Onegina'" ("Observations on the Rhythm of a Stanza in Eugene Onegin") in Materialy XXV naučnoj studenčeskoj konferencii, Tartu, 1970.

5 Andrej Belyj was the first to take into account the dependence of contrast on pauses. See his corrections for punctuation marks in the formula describing contrast in his book Ritm kak dialektika i "Mednyj Vsadnik", Moskva 1929. (Rhythm as a Dialectical Process and "The Bronze Horseman").

6 Puškin, Aleksandr Sergeevič. Polnoe sobranie sočinenij v 10-ti tomax, (Complete Collected Works in 10 Volumes). Volume 5, Moskva 1957.

7 This coefficient is described for instance, in B.L. Van der Waerden's Matematičeskaja statistika (Mathematical Statistics), Moskva 1960, p. 383. It is also widely used in V.S. Baevskij.

8 The calculations were carried out mainly by means of a "BESM-3M" computer according to a program compiled by the author. See my article "O programme dlja statističeskogo obsledovanija stixotvornyx tekstov" ("On a Program for the Statistical Research of Poetic Texts") in Metody vyčislenij, Leningrad 1974.

9 Bol'šev, L.M., Smirnov, N.V. Tablicy matematičeskoj statistiki. (Tables of Mathematical Statistics). Moskva 1965, p. 425.

10 Since pauses 2 and 3 show no correlation, separate data regarding them has not been provided. We have designated 0 and 1 as short pauses and 2 and 3 as long pauses.

11 Krasnoperova, M.A. "Zamečanija po povodu gipotezy o nezavisimosti ritmičeskix struktur v xudožestvennom tekste. ("Observations on the Hypothesis of the Independent Nature of Rhythmic Structures in Literary Texts") in Materialy V. vsesojuznogo simpoziuma po psixolingvistike i teorii kommunikacii, Moskva 1975. - In the present paper we have analyzed the works listed in the above-mentioned article.

12 The variants of the type 11.3x show no correlation by themselves. Therefore their relation to short pauses was ascertained on the basis of the sets including 10.3x.

13 In this way we indicate on the left-hand side of the equality the variant with a less degree of contrast than in any other variant on the right-hand side.

14 Inequality 7 forces us to consider Alternative 3 to be false when applied to 3x.10 (see IX).

THE ALGORITHM AND SOME RESULTS OF A STATISTICAL INVESTIGATION OF ALTERNATING RHYTHM ON THE MINSK-32 COMPUTER

V.S. Baevskij, L.Ja. Osipova

1. THE OBJECT OF INVESTIGATION

This article deals with a present-day trend in the automation of linguistic investigations in the field of research on verse rhythm (see KRASNOPEROVA 1970; HIRSCHMANN 1972; PALA 1972). The verse of alternating rhythm chosen for our investigation has been predominant during the whole history of new Russian poetry, i.e. for nearly two and a half centuries. It presents a regular one syllable alternation of strong positions (ictuses) and weak positions (non-ictuses). Usually all syllables in poetry are treated in terms of the binary opposition "stressed" / "non-stressed". Metrically ambiguous words (pronouns, auxiliary verbs, monosyllabic adverbs, interjections, polysyllabic prepositions, conjunctions, particles) are adapted artificially to it. Stressed syllables of polysyllables are always placed on ictuses only. Non-stressed syllables and monosyllables are placed on both ictuses and non-ictuses. The ictuses of a trochee are placed on odd syllables, those of the iambus are placed on even syllables.

2. SOME LINGUISTIC PRECONDITIONS OF THE INVESTIGATION OUR METHOD OF RHYTHM REPRESENTATION

"Potebnja's effect" introduced more than a hundred years ago, is well-known in linguistic literature (see POTEBNJA 1865; KOŠUTICH 1919, 21-63; BOGORODICKIJ 1930, 320-324; REFORMATSKIJ 1967, 196; PANOV 1967, 185, 360); on the distinction of syllables in Russian dialects see (AVANESOV 1949). Potebnja's effect makes it possible to define more exactly the verse rhythm with due regard to the gradual opposition of syllables on the basis of its

distinctness (vydelennost') (the concept of stressed/non-stressed is replaced by the concept of distinctness). The distinctness of the stressed syllable is 3.0, that of the first pretonic syllable, of the initial uncovered and final open syllable is 2.0, and of the rest of syllables - 1.0. In verse with alternating rhythm this effect becomes complicated by the influence of the metrical scheme, i.e. the position of the stressed syllable on the ictus or on the non-ictus. In addition the degree of significance of a word is taken into account.

The degree of distinctness is substituted for by every syllable m, where m assumes a value from the series of numbers

$$m = \{1.0;\ 1.5;\ 2.0;\ 2.5;\ 3.0\}$$

in accordance with the following rules introduced in BAEVSKIJ (1967).

The stressed syllable of a significant word on the ictus = 3.0.

The stressed syllable of a significant word on the non-ictus = 2.5.

The stressed syllable of a metrically ambiguous word on the ictus and non-ictus = 2.5.

The 1st pretonic, the 2nd initial uncovered, the 1st and the 2nd posttonic, the final open syllable = 2.0.

The 3rd pretonic initial uncovered and the 3rd posttonic final closed = 1.5.

The rest of the syllables = 1.0.

3. STATISTIC CHARACTERISTICS OF THE RHYTHM

3.1. SYLLABIC RHYTHM

Calculating average distinctness in a poem as a whole (here as well as below non-stressed syllables in the clausula are not taken into account) is done according to the formula:

$$\bar{m}_i = \frac{1}{n} \sum_{j=1}^{n} m_{ij}, \qquad (1)$$

where i is the position of the syllable if we count from the beginning of the line; j is the position of the line in the poem (lines of the meter under consideration are being investigated, lines of the different meters are omitted); m_{ij} is the distinctness of the i syllable in the j line. The average distinctness of every syllable $\bar{m}_i$ is an important rhythmic characteristic of a poem, a poet or a whole period. The method proposed in contradistinction to the traditional method, which only takes ictuses into account allows investigation of the participation of non-ictuses in the rhythm's organization as well.

3.2. The average distinctness of all ictuses in the line is calculated according to the formula:

$$\bar{m}_k = \frac{1}{k} \sum_{i}^{k} \bar{m}_i, \qquad (2)$$

where k is the number of ictuses in the line; i = s, s - 2, s - 4, s - 6, ...,r; s is the whole number of syllables in a line; r = 1 or 2.

3.3. The average distinctness of all non-ictuses in a line is calculated analogously as

$$\bar{m}_l = \frac{1}{l} \sum_{i}^{l} \bar{m}_i, \qquad (3)$$

where i = s - 1, s - 3, s - 5, ...,r; l is the number of non-ictuses.

3.4. The average distinction of the rhythm is the difference between the average distinctness of ictuses and non-ictuses:

$$\bar{m}_t = \bar{m}_k - \bar{m}_l.$$

This is another very important index. A strong average distinctness of ictuses ($\bar{m}_k \to 3.0$) and a weak average distinctness of non-

ictuses ($\bar{m}_1 \to 1.0$) emphasizes the meter by strong opposition of ictuses and non-ictuses. The reverse tendency reduces such "natural" opposition and creates opportunities for an increase in the rhythmical originality of the verse and for metric shifts that are peculiar to modern Russian poetry[1]. We shall call m_t the index of metric distinction of the rhythm (metričeskaja otčetlivost' ritma).

3.5. The average distinctness of the syllable in a poem as a whole is calculated according to the formula:

$$\bar{m} = \frac{1}{s} \sum_{i=1}^{s} \bar{m}_i . \qquad (4)$$

The average length of the word in the verse is characterized in directly by the value $\bar{m}$. An increase in $\bar{m}$ means that the number of stresses (and therefore of phonetic words) increases and the average syllabic length of the word decreases: the value of phonetic words is inversely proportional to their average syllabic length[2].

3.6. The rhythmic characteristic of the line. Let us turn from the average rhythmic characteristics of the whole poem to the indivudual characteristics of each line and calculate the mean squared deviation of this line's syllable distinctness from the average distinctness of these syllables on the whole:

$$\sigma_j = \sqrt{\frac{1}{s} \sum_{i=1}^{s} (m_{ij} - \bar{m}_i)^2} . \qquad (5)$$

Lines with extreme quadratic deviations are rhythmical rarities. They differ most from the ideal average rhythm and have the most noticeable individual rhythmical form. It is philological analysis that clears up this rhythmical deviation in the poem's semantics as a whole.

Observations show (see BOBROV 1965; KOLMOGOROV 1966; BAEVSKIJ 1972: 52f.) that rhythmically marked lines are as a rule of great importance for the organization of the meaning.

The indices of the mean squared deviation of the syllables in a line in an ensemble of poems written in the same meter by the same or different authors may be computed in the same way.

3.7. Variability of the rhythm. The mean value of the mean squared deviations of all the lines in the poem or rather the more preferable value of dispersion

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} \sigma_j^2 = \frac{1}{ns} \sum_{j=1}^{n} \sum_{i=1}^{1} (m_{ij} - \bar{m}_i)^2 \qquad (6)$$

is a general characteristic of rhythmic variability of a poem.

The value $\bar{\sigma}^2$ increases with an increase in the rhythmical originality of the poem.

Thus the value σ_j is the most significant characteristic of the rhythm of a certain line and the values $\bar{m}_i$, $\bar{m}_t$, $\bar{m}$ and $\bar{\sigma}^2$ are the most significant general characteristics of the rhythm of the poem as a whole.

4. THE INTERRELATION OF LINE RHYTHM AND BACKGROUND

One may assume that the average data under consideration shape the readers' perception of the rhythmic structure. Consequently the mean squared deviation of syllables' distinctness of the line shapes the process of perception as a result of interaction between what one expects as a result of previous experience and the individual rhythmic form of the present line. But the psychological approach towards perception in art is ineffective nowadays in clarifying how one perceives the rhythm of a whole line: either against the background of a poem, or against the background of an ensemble of poems written by the same poet in the same meter approximately at the same time or against the background of some mean rhythm of various poets writing in the same meter. For example note that Samojlov's line

I UŽEL' V DOMIŠKAX ČINNYX

differs from his individual rhythm but it does not differ noticeably from the rhythmical standard of the literary period. This line's σ_j = 0.78 exceeds the σ_j of all the other lines of the same poem. The mean squared deviation of the distinctiveness of the syllables from their average distinctness in the ensemble of poems by Samojlov and other modern poets is 0.67; it is less than the σ_j of the lines having the value on the level of 0.94; 0.88, etc. It is of great importance whether one or more lines turn out to be rarities in comparison with the other lines of the poem (the corpus of the 1st stage), of the ensemble of poems by the same poet (the corpus of the 2nd stage) and of the ensemble of poems by different poets (the corpus of the 3rd stage).

5. ORGANIZATION OF THE ALGORITHM AND THE COMPUTER PROGRAMME

The investigation was carried out on the "Minsk-32" computer. The operations discussed in §3 are simple enough but the analysis in accordance with §4 requires special organization of the algorithm and the programme.

We are aware that one may unite poems into ensembles in different ways, e.g. the poet's works may be divided into certain periods or poets of the period under consideration may be divided into certain groups so that the number of stages of the corpus may be changed according to the purpose of the investigation. They always form a hierarchial structure in which the corpora of higher stages include those of lower stages. Therefore it was necessary to make the programme an adaptable means of analysis with a view toward changing computation diagrams according to the problems in question.

With this aim in view the module principle is taken as the basis of the algorithm. The module is a standard programme. Each operation is accomplished by one or several modules.

The operating system is created immediately at the time of analysis by means of a dialogue between the investigator and the computer. The investigator selects an appropriate computing diagram of the analysis is given in Fig. 1.

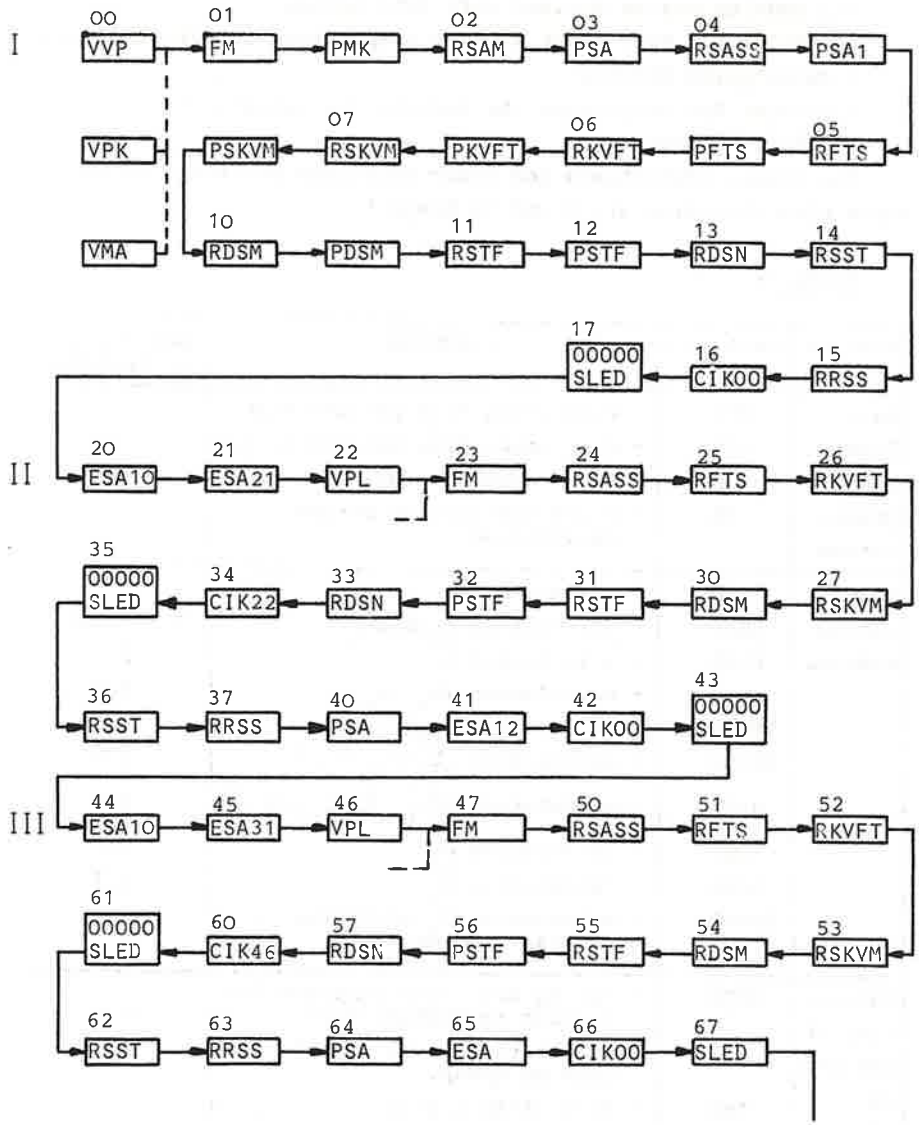


Fig. 1

All modules may be divided into four groups:

- modules for data input and its preparation for calculations;
- calculating modules;
- modules for outputting the results via typewriter;
- circuit operators.

The module identifiers and their functions according to the above classification are found in Table 1.

Table 1:

Modules type	Identifier	Function	Description of operation
Data input and its preparation	VPL	- File input from punched tape.	
	VPK	- File input from punched card.	
	VML	- File input from magnetic tape.	
	FM	- Array building up before calculation.	
Calculating modules	RSAN	- calculation $\bar{m}_i$	3.1
	RSST	- calculation $\bar{m}_k, \bar{m}_l, \bar{m}_t$	3.2 - 3.4
	RRSS	- calculation $\bar{m}$	3.5
	ROTS	- calculation $(m_{ij}-\bar{m}_i)$	3.6
	RKVOT	- calculation $(m_{ij}-\bar{m}_i)^2$	3.6
	RSKVM	- calculation $\sum_{i=1}^{s}(m_{ij}-\bar{m}_i)^2$	3.6
	RDSM	- calculation $\sigma_j^2=\frac{1}{s}\sum_{i=1}^{s}(m_{ij}-\bar{m}_i)^2$	3.6
	RSTO	- calculation σ_j	3.6
	RDSN	- calculation $\bar{\sigma}_j$	3.7
	RSASS	- calculation $\bar{m}_i$ according to recursive formula	5
Outputting of results on typewriter	PSTO	- Typing with line standard deviation (switching off)	
	PSA	- arithmetical mean typing of poem or group.	
	PMK	- Poem array typing	
	POTS		
	PKVOT		
	PSKVM	Control results typing	
	PDSM	of intermediate calculations.	
	PSA1		

Table 1 (Continuation)

Modules type	Identifier	Function	Description of operation
circuit = operator	CiKa	- Repetition operator of a part of the circuit in the cycle; a - module number from which the circuit repetition begins;	
	00000	Cycle finished (end of cycle).	
	SLED	- Go on further with the cycle.	
	(trace)		
	ESAbv	- Filling up of group arithmetical mean, b - field number, from which it is stored, v - field number in which it is stored.	

The modules may be put into operation from any data carrier. Each poem is stored as a separate unit (the corpus of the 1st stage).

Formula (1) is applied for computing the average distinctness of the syllables of each line.

The following recursive formula is used for computing the average distinctness of each syllable of the ensemble:

$$n_{m_i} = n-I_{\bar{m}_i} + \frac{I}{n}(n_{m_i}-n-I_{\bar{m}_i}).$$

This formula permits us to determine the average distinctness of the syllable with even more accuracy of calculation than formula (1) in case not all the elements m_{ij} are given simultaneously (see VAJT (White) 1971).

Typing the results is either combined with computing or organized in separate modules. A part of the modules of this type is used for the control of intermediate results only.

The diagram operators give the computer diagram adaptability. The operator permits automatic repetition of the working diagram more than once beginning with any module of the diagram; the arrangement of this operator in various positions of the diagram makes it possible to organize any number of cycles including internal loops.

The operator directed to store the arithmetic mean of the ensembles which may be used at the next stages of computing; it indicates the numbers of the fields used for storage.

The poem c_i^j is shown as the element of the 1st stage. Ensembles of elements of the 1st stage form the set of members of the 2nd stage A^j, $c_i^j \in A^j$, i.e.

$$A^1 = c_1^1 \cup c_2^1 \cup \ldots \cup c_x^1$$

$$A^2 = c_1^2 \cup c_2^2 \cup \ldots \cup c_x^2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$A^y = c_1^y \cup c_2^y \cup \ldots \cup c_x^y,$$

where x is the amount of elements (poems) in the set of numbers, y is the amount of set of numbers of the 2nd stage. The sets of the 2nd stage form the sets of numbers of the 3rd stage B^j, $A^j \in B^j$, i.e.

$$B^1 = A_1^1 \cup A_2^1 \cup \ldots \cup A_y^1$$

$$B^1 = A_1^2 \cup A_2^2 \cup \ldots \cup A_y^2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$B^z = A_1^z \cup A_2^z \cup \ldots \cup A_y^z,$$

where z is the number of sets of numbers of the 3rd stage, etc. The structure may be developed unlimitedly according to the principle of unification we have shown.

Computing within the limits of any ensemble is presented as a separate stage in the computation diagram.

The computation of the arithmetical mean necessary for computations at the next stage is carried out alongside the computation of the values for the given ensemble. One may observe that the structures of the computation diagram of the 2nd stage and the next stage differ only in the number of fields used for memorizing. After recall the diagram may be repeated more than once, automatically including more and more new poems.

These features constitute the advantage of our algorithm since they may be used even with changed computation operators.

6. THE RESULTS OF THE INVESTIGATION

6.1. Here we present the results of the investigation on the verse rhythm of 16 poets (40 poems, 843 lines). The computation time investigator computer dialogue (including information input and output of the results) was 4 hours. If the computation diagram is given beforehand, the computation time may be reduced considerably by changing conditions of information input and also by using magnetic tape instead of punched paper tape. The total analysis with the help of the programme described covered 442 poems (11282 lines) all in iambic tetrameter and pentameter and trochaic tetrameter and pentameter by 20 poets writing between 1956 and 1965. For the other results of the investigation see BAEVSKIJ (1975).

6.2. The syllabic rhythm is formed by the sequence of the average values of the distinctness of each syllable in the line (formula 1). The distinctness of the 7th syllable - of the next ictus of trochaic tetrameter - is always 3.0, therefore 6 other syllables are under consideration. The syllabic rhythm of various poems has an unpredictable atypical form. Poems differ in correlation of both ictuses and non-ictuses. Among 40 poems we find predominance of the 1st ictus over the 3rd (20 cases), of the 3rd ictus over the 1st (18 cases) and equality (2 cases), the predominance of the 1st non-ictus over the 3rd non-ictus (12 cases), of the 3rd non-ictus over the 1st ictus (18 cases), and equality (10 cases). Even the correlation of adjacent ictuses and non-ictuses is changed: the 1st ictus is more distinctive than the 1st non-ictus in 26 poems, the inverse, paradoxical correlation (a more distinctive non-ictus than an ictus) is observed in 8 poems; and the equality of the 1st ictus and of the 1st non-ictus is found in 6 poems. Approximately the same is true of the end of the line: the 3rd ictus is distinguished more than the 3rd non-ictus in 26 poems, less than the 3rd non-ictus in 11 poems, and they are equal in 3 poems.

The middle of the line, "the 2nd foot", is more fixed: the 2nd ictus is predominant over the 1st and the 3rd in all the poems except the poem "Solovjinaja ulica" by Samojlov; and the 2nd non-ictus is least distinguished (except for 6 cases).

The syllabic rhythm of trochaic tetrameter as a whole has the following characteristics. The greatest distinctness of the 4th and the 2nd ictus ensures the fulfilment of the laws of the regressive accentual dissimilation of the ictuses and of the stabilization of the ictus after the 1st non-ictus (on these laws see TARANOVSKIJ 1953: 334, 350; TARANOVSKIJ 1966: 184). The weakest 2nd non-ictus underlines still more the division of the line into two similar parts. The ictus and the non-ictus of the 2nd foot contrast sharply. The contrast is smaller in the 1st and 3rd feet. The correlations of the ictus and non-ictus of the 1st and 3rd feet are extremely diverse.

The average of all 40 poems give a "natural" picture (Tab. 2) the 2nd ictus ranks after the 4th; in its distinctness the 2nd non-ictus is the weakest; both ictuses of odd feet are more distinctive than their non-ictuses. But averages conceal very whimsic pictures of the syllabic rhythm. For example, both laws of the alternated rhythm are grossly violated in Šamojlov's "Solov'inaja ulica": the 1st ictus is more strangly distinguished there than the 2nd. The unique "image of the meter" A.N. Kolmogorov's term, see 1968 is formed by the unity of the rhythmic-syntactical features of the key-lines, where the 1st ictus is maximally distinguished and the middle ictuses maximally weakened.

ULICEJU LAKSTIGALAS,
ULICEJU SOLOV'INOJ (DS, 31).

An entirely different image of the meter of trochaic tetrameter is peculiar to the poem by L. Martynov "Otmečali vy, scholasty...": the distinctness of the 1st and 3rd ictuses is here lower than that of the 1st and the 3rd non-ictuses correspondingly, and two ictuses are distinguished more weakly than all other non-ictuses. An unusual syllabic rhythm is underlined by the poem's graphic form: almost each verse is divided into two parts, and the final ictus of the two-syllable half-lines thus formed is filled naturally by the stressed syllables, and the initial ictus (according to the law of regressive accentual dissimilation) are filled with non-stressed syllables:

OTMEČALI
VY, SCHOLASTY

PTOLOMEJA
JUBILEJ (...) (LM, 58).

The syllabic rhythm of "Pesenka Saguramo" by A. Mežirov is similar in some respects to that of A. Martynov (the two poems differ in strophic organization): the distinctness of the 1st and the 3rd ictuses is weaker than the distinctness of the corresponding non-ictuses. The 1st non-ictus is equal to the 3rd ictus, the 2nd non-ictus is distinguished least in the poems under consideration; it is characterized by an extremely limited distinctness of 1.19.

One might point out some poems in our material in the feet of which the ictus is constantly more distinguished than the non-ictus. This is the case for poems by B. Pasternak, P. Antokolskij, A. Prokofjev, E. Vinokurov. Other poets, expecially Martynov, Samojlov and Mežirov demonstrated the opposite, paradoxical picture more or less often.

Table 2

Poets	Syllable						
	1	2	3	4	5	6	7
NA	218	187	250	181	231	200	300
BP	224	166	272	153	201	185	300
PA	235	188	285	160	227	188	300
NU	205	176	284	149	187	188	300
AP	252	185	297	163	222	164	300
AT	212	194	287	144	181	199	300
JaS	210	185	294	151	180	194	300
LM	187	190	278	147	200	183	300
AJa	203	186	286	165	212	196	300
BS	227	187	267	135	172	200	300
DS	219	183	279	152	172	188	300
AM	198	191	282	145	183	195	300
EV	203	191	287	144	222	194	300
AV	189	193	298	159	198	191	300
EE	217	196	294	142	177	198	300
BA	204	191	286	151	189	193	300
$\bar{m}_i$	208	187	285	150	195	191	300
	on total material of						
LM-O	169	200	300	200	175	200	300
DS-S	256	170	253	168	165	188	300
AM-P	156	194	281	119	194	200	300

All numbers are multiplied by 100. LM-O: data for the poem by Martynov "Otmečali vy, scholasty..."; DS-S - for the poem by D. Samojlov " Solov'inaja ulica"; AM-P - for the poem by A. Mežirov "Pesenka Saguramo".

6.3. The metrical distinction of the rhythm and conjugate features are defined by the formulae (2), (3) and (4). Computation results are shown in Table 3. We see that in the verse of L. Martynov, A. Mežirov, B. Sluckij, D. Samojlov the ictuses are weakened and in the verse by A. Prokofjev, P. Antokolskij, N. Aseev, E. Vinokurov they are distinguished most strongly. The first tendency effectuates a decrease in the metrical distinction of the rhythm, the second effectuates its increase. The strong prominence of the non-ictuses that we see in verse of A. Jašin, N. Aseev, P. Antokolskij and E. Evtušenko serves also as an indication of the weakening metrical distinction of the rhythm. The non-ictuses are weakened in trochaic tetrameter of B. Pasternak, A. Prokofjev, N. Ušakov and L. Martynov.

One can detect an absence of determination between the average distinctness of the ictuses and non-ictuses. The metrical distinction of the rhythm is formed by their complex interaction. It is weakest in verse by A. Mežirov, B. Axmadulina, A. Jašin, L Martynov, D. Samojlov and B. Sluckij and the greatest in that of A. Prokofjev, P. Antokolskij, B. Pasternak and E. Vinokurov.

Computations on the verse of the 19th-20th centuries show a tendency to progressive weakening of the metrical distinction of the rhythm (GASPAROV 1974: 98). The strong ictuses (including the 2nd ictus of trochaic tetrameter) become weaker and the weak ictuses become stronger. We may add to this observation that non-ictuses also take part in the process of smoothening rhythmical curve. The metrical distinction of the rhythm diminishes sometimes because of the sharp increase of the non-ictuses' distinctness.

The average distinctness of a syllable in a poem as a whole has a tendency to decrease from the first half of the 19th century to the second half of the 20th century. It is manifested clearly by the trochaic tetrameter of L. Martynov, A. Mežirov, N. Ušakov, B. Sluckij, B. Pasternak and D. Samojlov. At the same time we can detect an orientation to the 19th century in the verse of A. Prokofjev, P. Antokolskij, N. Aseev and A. Jašin.

Table 3:

Poets	$\bar{m}_k$	$\bar{m}_l$	$\bar{m}_t$	$\bar{m}$	Poets	$\bar{m}_k$	$\bar{m}_l$	$\bar{m}_t$	$\bar{m}$
NA	255	182	073	223	AJa	250	183	067	221
BP	249	168	081	214	BS	242	174	068	213
PA	262	179	083	226	DS	243	175	068	214
NU	244	171	073	213	AM	241	177	064	213
AP	268	171	097	226	EV	253	176	077	220
AT	245	178	067	216	AV	246	181	065	218
JaS	246	177	069	216	EE	247	178	069	218
LM	241	173	068	212	BA	245	178	067	216
Average on total material						247	176	071	216

(All numbers are multiplied by 100.) The whole complex of the features studied permits us to consider the trochaic tetrameter of L. Martynov, B. Sluckij, D. Samojlov and A. Mežirov to be a natural result of rhythm's evolution, and the trochaic tetrameter of the late poems by P. Antokolskij and A. Prokofjev as a return to tradition of the 19th century. The peculiarities of both the tradition of the 19th century and of the modern rhythm are inextplatty entangled in the poems by the other poets.

6.4. The individual rhythmic characteristics of a line and the variability of the rhythm are computed by means of the formulae (5) and (6). Let us consider the poem by E. Evtushenko "Koroleva krasoty". The vision of the Cuban revolution is closely intershown here with the vision of a beauty competition. The greatest mean squared deviation of a line from the average distinctness of the syllables in the poem as a whole marks out the line rhythmically.

NU, A JA VLJUBILSJA NAČISTO (EE, 74).

The greatest mean squared deviation of the average distinctness in the line from the average distinctness in the total volume of trochaic tetrameter underlines rhythmically another line divided into two parts (the line quoted above is thereby relegated to the second rank).

ŽENŠČIN -

STRAŠNOE KOLIČESTVO! (EE, 72).

Both lines, marked by the rhythmical italics, designate two more significant themes of the poem.

Data on the rhythmical variability of poets under investigation, computed relative to a 1st stage corpus (of a poem), the 2nd stage (of an ensemble of poems if there are more than one) and the 3rd stage corpus (of all trochaic tetrameter in our material) agree well. The same poets occupy the identical place (ignoring unessential transpositions) in rank lists made according to the regression of the value $\bar{\sigma}^2$ relative to the average in the corpuses of several stages. Such a rank list relative to the 1st stage is presented below.

B. Pasternak (0.34), N. Ušakov and L. Martynov (0.30), A. Voznesenskij (0.28), A. Prokofjev and D. Samojlov (0.27), B. Sluckij, A. Mežirov and B. Achmadulina (0.26), A. Jašin (0.25), N. Aseev, A. Tvardovskij and E. Vinokurov (0.24), Ja. Smeljakov (0.23), P. Antokolskij (0.21), E. Evtušenko (0.16).

This list may be compared with that made for the 3rd stage: B. Pasternak (0.37), A. Prokofjev (0.34), L. Martynov (0.32), N. Ushakov (0.31), D. Samojlov and A. Voznesenskij (0.30), B. Sluckij and A. Mežirov (0.28), N. Aseev (0.27), B. Achmadulina (0.26), A. Tvardovskij, A. Jašin and E. Vinokurov (0.25), P. Antokolskij and Ja. Smeljakov (0.24), E. Evtušenko (0.20).

One can see that the same poets are at the beginning, in the middle, and at the end of the both rank lists. Data on the rhythmical variability computed relative to any stage indicate a general tendency of the verse: the rhythmical variability, the rhythmical restraint or the rhythmical limitation.

It should be noted that these lists and lists similar to these are not to be interpreted as dichotomic scales (e.g. "talent - non-talent"). Our criterion shows the way the rhythm is used by a poet, but it does not show if a poet's work is good or bad. We can enumerate at least the following aspects of the structure of a work written in verse: phonology, metrics, stanza, rhyme, composition, grammatical, lexical trope, figurative and non-figurative compound semantic complexes. All these aspects take part in the semantic organization of the work, and as usual some aspects become more complicated, while others get simplified. The rhythm is one of the means of the whole framework.

According to our rank lists, Pasternak's verse excels that of the other poets in rhythmic variability. It has also a very complicated system of imagery (see, for example BAEVSKIJ (1980), ŽOLKOVSKIJ, ŠČEGLOV (1971). At the same time, in spite of the complexity of the rhyme composition, lexical and phonological aspects of his poems are comparable in their complexity with those of some of his contemporary poets. But his metrics and stanza forms are strikingly simple in contrast to the verse of A Jašin or B. Sluckij.

The rank lists provide evidence for the rhythmical variability of B. Pasternak's, N. Ušakov's, A. Prokofjev's, D. Samojlov's and A. Voznesenskij's verse - poets of different generations and aesthetical positions. These poets continue the process of movement from limitation to increase of the rhythm's variability observed during the last century and a half. At the same time the verse of the late P. Antokolskij, A. Tvardovskij, Ja. Smeljakov, E. Vinokurov and E. Evtušenko - poets of different generations and aesthetic positions - is also characterized by rhythmic limitation. The rest of the authors of our list occupy intermediate positions. We define the rhythm of these poets' verses as restrained.

6.5. Comparison with other meters. The above observations (GASPAROV 1974, TARANOVSKIJ 1971) show that the same poets treat different meters similarly. But this is true only of the more general tendencies and primarily of the average distinctness of the ictuses. In other cases we see a considerable originality in the rhythmical treatment of different meters. For example the variability of B. Pasternak's trochaic tetrameter rhythm is 0.34; it exceeds considerably the variability of the trochaic tetrameter rhythm of all the other poets. But the variability of his iambic tetrameter rhythm is 0.30; it is less than the index of A. Achmatova, P. Antokolskij or L. Martynov and is at the level of Mežirov's verse. The variability of the iambic rhythm and trochaic tetrameter of P. Antokolskij is still more pronounced; they are 0.21 and 0.32 respectively. The whole complex of our data confirms that "each poet has his favourite meter through which the potentials of his rhythmical skill are expressed" (BELYJ 1910: 395).

The average indices of the metric distinction of the rhythm and of the conjugate characteristics within the limits of the

whole corpus of trochaic tetrameter in comparison with the same characteristics of the other meters (Table 4) show its stability. As it appears from §§2 and 3, these characteristics reflect indirectly the statistical organization of linguistic phenomena (stress, word-boundary, type of syllables) in speach. The similarity of the rhythmical characteristics of different meters is connected with the stability of these frequencies, given large enough excepts.

Table 4. (All numbers are multiplied by 100)

	$\bar{m}_k$	$\bar{m}_1$	$\bar{m}_t$	$\bar{m}$	$\bar{\sigma}^2$
Ch4	247	176	071	216	028
Ch5	246	176	070	214	031
Ja4	253	179	074	216	027
Ja5	255	182	073	218	029

7. STRUCTURE OF THE RHYTHM

Further it is necessary to consider the question of the correlation between various characteristics of trochaic tetrameter. We shall compare in pairs using the rank correlation analysis (see VAN DER VAERDEN 1957) the distribution of the indices of the average distinctness of the ictuses, the average distinctness of the non-ictuses, the metric distinction of the rhythm, the average distinctness of the syllable and rhythmic variability of all the poets under consideration. The coefficient of the rank correlation R = 1, if the connection of the indices is direct and determined. R = -1 if the connection of the indices is inverse and determined. $0 < R < 1$, if the connection is direct and stochastic (when one index increases the other index also increases on the average); $0 > R > 1$, if the connection is inverse and stochastic. R = 0, if the indices are independent and there is no connection at all.

The results of the rank correlation analysis are given in Table 5. To judge their significance we must bear in mind that

it is only values $R \leq -0.52$ or $R \geq 0.52$ that enter the critical domain (the significance level is 0.03). The correlation coefficients in the critical domain indicate the connection that exists between the pairs of indices.

Table 5

	$\bar{\sigma}^2$	$\bar{m}_k$	$\bar{m}_1$	$\bar{m}_t$
$\bar{m}_k$	0.38			
$\bar{m}_1$	<u>0.82</u>	0.30		
$\bar{m}_t$	0.02	<u>0.75</u>	-0.28	
$\bar{m}$	<u>0.56</u>	0.93	0.47	<u>0.57</u>

Five indices in pairs make ten pairs. Rank correlation coefficients of five of these pairs only get into the critical domain. They are all positive (those are underlined in Table 5). The more important indices of the rhythm - metric distinction and variability - do not correlate. It increases the value of each of them. The average distinction of the syllable is connected with all other indices except the average distinctness of the non-ictuses; its connection with the average distinctness of the ictuses is almost determinant, which confirms the importance of this index for the conception of rhythm in general. In the whole structure the metric distinction of rhythm is connected with the average distinctness of only the ictuses. But the average distinctness of the non-ictuses does not fall outside the whole system of indices: it is closely connected with the variability of rhythm.

On the whole the structure of the rhythmical indices of trochaic tetrameter is characterized partly by its independence, and partly by a direct stochastic connection. The same is true of the other meters. It is only the total set of indices that shows the verse rhythm fully enough.

The authors are expecially grateful to A. N. Kolmogorov and A.M. Šenderovič for their valuable advice on the methods of investigation and also to S.I. Gindin and M.A. Krasnoperova for very useful discussion of the results.

NOTES

1 See, for example, "Paxlo letom..." by L. Martynov, which is written in trochaic tetrameter; the line GDE-TO ZA UNIVERSITE-TOM... was omitted because of a "superfluous" syllable (LM, 34; an index of sources of the texts and abbreviations is at the end of the article).

2 Indeed, the picture is more complicated because according to §2 the distinctness of a syllable depends not only on its position with respect to the stress and word division but also on some other causes. But our analysis shows that these "other" causes play a minor role.

SOURCES

N. ASEEV, LAD. Moskva, Sov. pisatel' 1963(2), (NA).

B. PASTERNAK, Kogda razgvljaetsja. In: B. PASTERNAK, Stichotvorenija i poemy. Moskva-Leningrad, Sov. pisatel' 1965 (BP)

P. ANTOKOL'SKIJ, Četvertoe izmerenie. In: P. ANTOKOL'SKIJ, Izbrannoe, Bd. 2. Moskva, Chudožestvennaja Literatura 1966 (PA)

N. UŠAKOV, Teodolit. Moskva, Sov. pisatel' 1967 (NU)

A PROKOF'EV, Priglašenie k putešestvij. Leningrad, Chudožestvennaja Literatura 1969 (AP)

L. MARTYNOV, Novaja kniga. Moskva, Moskovskij Rabočij 1962 (LM)

A. TVARDOVSKIJ, Kniga liriki. Moskva, Sov. pisatel' 1967 (AT)

Ja. SMELJAKOV, Rabota i ljubov. Moskva, Molodaja gvardija 1963(2) (JaS)

A. JAŠIN, Bessonnica. Moskva, Sov. Rossija 1968 (AJa)

B. SLUCKIJ, Segodnja i včera. Moskva, Molodaja gvardija 1968 (BS)

D. SAMOJLOV, Vtoroj pereval. Moskva, Sov. pisatel' 1963 (DS)

A. MEŽIROV, Stichotvorenija. Moskva, Gichl 1963 (AM)

E. VINOKUROV, Slovo. In: E. VINOKUROV, Izbrannoe. Moskva Chudožestvennaja literatura 1968 (EV)

A. VOZNESENSKIJ, Antimiry. Moskva, Molodaja gvardija 1964 (AV)

E. EVTUŠEWKO, Vzmach ruki. Moskva, Molodaja gvardija 1962 (EE)

B. ACHMADULINA, Struna. Moskva, Sov. pisatel' 1962 (BA)

REFERENCES

(1) R.I. AVANESOV, Očerki russkoj dialektologii I. Moskva, UČPEDGIZ 1949

(2) V.S. BAEVSKIJ, Čislovye značenija sily slogov v stiche al'ternirujuščego ritma. Filologičeskie nauki 1967/3

(3) V.S. BAEVSKIJ, K evoljucii liričeskogo sticha Bloka. In: V.S. BAEVSKIJ, Stich russkoj sovetskoj poezii. Smolensk 1972

(4) V.S. BAEVSKIJ, Tipologija sticha russkoj liričeskoj poezii. Tartu, Diss. 1975

(5) V.S. BAEVSKIJ. Mif v poetičeskom soznanii i lirike Pasternaka. Izvestija AN SSSR, serija literatury i jazyka 39, 1980/2,

(6) A. BELYJ, Sravnitel'naja morfologija ritma russkich lirikov v jambičeskom dimetre. In: A. BELYJ, Simvolizm. Kniga statej. Moskva, MUSAGET 1910

(7) V.A. BOGORODICKIJ, Fonetika russkogo jazyka v svete eksperimental'nych dannych. Kazań, Izd. Doma tatarskoj kul'tury 1930

(8) S.P. BOBROV, Sintagmy, slovorazdely i litavridy. Russkaja literatura 1965/4,

(9) M.L., GASPAROV, Sovremennyj russkij stich. Moskva, Nauka 1974

(10) R. HIRSCHMANN, A Computational Approach to the Study of Verse-rhythm. - Computer Studies in the Humanities & Verbal Behavior, 1972,V. 3, No. 4

(11) A.N. KOLMOGOROV, O metre puškinskich "Pesen Zapadnych Slavjan. Russkaja literatura 1966/1,

(12) A.N. KOLMOGOROV, Primer izučenija metra i ego ritmičeskich variantov. In: Teorija sticha. Leningrad, Nauka 1968

(13) R. KOŠUTICH, Grammatika ruskog jezika. 1. Glasovi. Petrograd 1919

(14) M.A. KRASNOPEROVA, Ob ispol'zovanii EVM dlja izučenija metra i ritma stichotvornych tekstov. In: Informacionnye processy, evrističeskoe programmirovanie, problemy nejro kibernetiki, modelirovanie avtomatami, raspoznavanie obrazov, problemy semiotiki. Tbilisi 1970, 345-347

(15) K. PALA, On some Conflicts between Grammar and Poetics. - In: Prague Studies in Mathematical Linguistics, 4. Praha, 1972.

(16) M.V. PANOV, Russkaja fonetika. Moskva, Prosveščenie 1967

(17) A.A. POTEBNJA, O zvukovych osobennostjach russkich narečij. Filologičeskie zapiski (Voronež) 1865/1, 62

(18) A.A. REFORMATSKIJ, Vvedenie v jazykovedenie. Moskva, Prosveščenie 1967

(19) K. TARANOVSKI, Ruski dvodelni ritmovi. 1. Beograd, Srpska Akademia Nauka 1953

(20) K.F. TARANOVSKIJ, Osnovnye zadači statističeskogo izučenija slavjanskogo sticha. In: Poetics.Poetyka. Poetika. Warszawa, 1966

(21) K.F. TARANOVSKIJ, O ritmičeskoj strukture russkich dvusložnych razmerov. In: Poetika i stilistika russkoj literatury. Leningrad, Nauka 1971,

(22) B.L. van der WAERDEN, Mathematische Statistik. Springer-Verlag. Berlin-Göttingen-Heidelberg, 1957

(23) R. VAJT (WHITE), Bystryj metod rekursivnoj ocenki srednego na CVM. Vyčislitel'naja technika 1971/8,

(24) A.K. ŽOLKOVSKIJ, Ju.K. ŠČEGLOV, K opisaniju smysla svjaznogo teksta. Moskva, Institut russkogo jazyka AN SSSR 1971

DAS RHYTHMISCHE SYSTEM DER "ALEXANDRINISCHEN GESÄNGE"

G.S. Vasjutočkin

Die in der Sammlung "Alexandrinische Gesänge" von M.A. Kuzmin enthaltenen Gedichte gehören zu den interessantesten und kompliziertesten Beispielen des *vers libre*. Offen bleibt jedoch bis heute die folgende, schon 1921 von V.M. Žirmunskij gestellte Frage: "Was gibt uns das Recht, derartige Werke als Verse aufzufassen und sie wie Verse zu drucken (in einzelnen Zeilen), und nicht als einen Ausschnitt aus poetischer Prosa?" (ŽIRMUNSKIJ 1975: 528)

Nach einer Untersuchung von einigen Gesängen der Sammlung bemerkte Žirmunskij sehr richtig, daß "jeder Versuch, eine metrische Formel für einen solchen Vers zu ermitteln ... von vornherein aus theoretischen und allgemeinen Gründen als unergiebig angesehen werden muß" (ŽIRMUNSKIJ 1975: 529). Er zeigte auf, daß anstelle des fehlenden Metrums andere Formen zur Strukturierung der Gedicht-Segmente verwendet werden. Als wichtigsten rhythmusbildenden Faktor, der ein Gefühl der "Periodizität" des *vers libre* in den "Alexandrinischen Gesängen" hervorruft, bezeichnete Žirmunskij den syntaktischen Parallelismus. "Die Anordnung der syntaktischen Strukturen stellt somit die Basis der kompositorischen Gliederung des freien Verses dar." (ŽIRMUNSKIJ 1975: 530). Es erscheint jedoch zweifelhaft, daß das Motiv für die rhythmische Organisation aller Gedichte dieser Sammlung auf der syntaktischen Ebene zu finden ist. Die vorliegende Arbeit will Faktoren der rhythmischen Organisation in den kompliziertesten Beispielen für *vers libre* der "Alexandrinischen Gesänge" herausstellen, die eben nicht auf syntaktischen oder strophischen Parallelismus zurückführbar, sondern rein rhythmischer Natur sind. Wir wollen uns dabei auf formale Faktoren konzentrieren, wie etwa die Anzahl der Silben, der Betonungen, die Länge der Intervalle zwischen den Betonungen, d.h. auf alles, was dazu beitragen kann, den Eindruck von rhythmischer Inertheit entstehen zu lassen.

Bevor wir den Beweis führen, daß solche bislang unbekannten Faktoren zur Rhythmussteuerung existieren, soll festgestellt werden,

a) ob die Gedichte der Sammlung rhythmisch homogen sind oder ob

sie in Gruppen aufteilbar sind, wobei jede ihr eigenes System von wichtigsten Rhythmik-Definitoren hat, und

b) welche durchgängigen Wiederholungsmaße in den einzelnen Gedichten der "Alexandrinischen Gesänge" zu finden sind.

Der sechste Zyklus der Sammlung, "Kanopskie pesenki" läßt sich leicht aufgliedern. Die Gedichte dieses Zyklus (Gruppe A) (mit einer Ausnahme) können nicht zum *vers libre* gerechnet werden. In einem Fall liegt ein reiner Choreus vor: "Ach, naš sad, naš vinogradnik" (Ach, unser Garten, unser Weingarten), und in dem anderen Fall ein dreigliedriges alternierendes Versmaß ("Kružites', kružites'") (Dreht euch, dreht euch...). Das Gedicht "Adonisa Kiprida iščet" (Cyprida sucht Adonis) enthält dreibetonte Verse, wobei jeweils zwei aneinandergrenzende Zeilen einen Paarreim bilden (Reimschema: aab ccb). Das erste Gedicht des Zyklus ist ebenfalls metrisch: alle ungeraden Zeilen sind hier in dreifüßigen Jamben mit daktylischen Klauseln geschrieben, und alle geraden Zeilen folgen dem logaödischen Schema 'UU'U' (zweites Kolon des alkäischen Elfsilblers).

In den Zyklen "Vstuplenie" (Auftakt), "Ljubov'" (Liebe) und "Zaključenie" (Abschluß) sind alle Gedichte in rhythmischer Hinsicht gleich gebaut (Gruppe B). Man findet in ihnen durchgängige (d.h. für ein ganzes Gedicht geltende und sich von Zeile zu Zeile wiederholende) rhythmusorganisierende Faktoren. Die Klauseln sind hier konstant (weiblich), und dort, wo es bei der Klausel zu Schwankungen kommt, wird der Zeilenanfang fixiert und die Konstanz durch eine Anakrusis wiedergegeben. In fast allen Zeilen läßt sich Betonungsgleichheit nachweisen (eine Ausnahme bilden hier nur die beiden Gedichte "Esli b ja byl drevnym polkovodcem" (Wenn ich ein Feldherr alter Zeiten wär) und "Kak pesnja materi" (Wie der Mutter Lied).

Parallele Bildungsweisen, die noch durch den syntaktischen Parallelismus der Gedichtzeilen verstärkt werden, findet man gewöhnlich in denjenigen Gedichten der Gruppe B, die keine der oben genannten Rhythmus-Konstanten aufweisen. Dies wird besonders deutlich bei den Gedichten "Esli b ja byl drevnim polkovodcem", "Kak pesnja materi" sowie in dem Gedicht "Kogda mne govorjat Aleksandrija" (Wenn man mir von Alexandria spricht), wo in drei von sechzehn Zeilen die Konstanz der Anakrusis zerstört wird. Es ist möglich, daß der syntaktische Parallelismus hier den Verlust der durchgängigen rhythmischen Faktoren "kompensiert".

Es sei noch angemerkt, daß die Gedichte der Gruppe B (insgesamt 11) nahezu die gleiche mittlere Zeilenlänge aufweisen. Sie variiert in sehr engen Grenzen von 8.9 bis zu 11.5 Silben. Eine Ausnahme bildet hier ebenfalls wieder das Gedicht "Kak pesnja materi", das nach Žirmunskij als das deutlichste Beispiel für syntaktischen Parallelismus gilt. Die mittlere Zeilenlänge liegt in diesem Gedicht bei 7.8 Silben.

Schwieriger sind die Rhythmik-Konstituenten in den Gedichten nachzuweisen, die zu dem Zyklus "Ona" (Sie), "Mudrost'" (Weisheit) und "Otryvki" (Ausschnitte) gehören (Gruppe C). Fünf von siebzehn Gedichten dieser Gruppe, nämlich:

"Nas bylo četyre sestry"
"Vesnoju list'ja ronjaet topol'" — Zyklus "Ona"
"Ich bylo četvero v étot mesjac"
"Ja sprašival mudrecov Vselennoj" — Zyklus "Mudrost'"
"Čto za dožd'" — Zyklus "Otryvki"

ähneln in ihrer rhythmischen Konstitution der vorhergegangenen Gruppe. In drei von fünf Fällen liegt eine in allen Zeilen gleiche Betonung vor ("Nas bylo četyre sestry", "Čto za dožd'" und "Ja sprašival mudrecov Vselennoj"). In allen fünf Gedichten bleiben entweder die Klauseln (weiblich) oder die Anakrusis konstant. Als zusätzlicher Organisationsfaktor der Zeilen dient die Syntax.

Kehren wir wieder zum Ausgangspunkt unserer Fragestellung zurück: gibt es rhythmusorganisierende Faktoren (wenn ja, welche?) in denjenigen Gedichten der Sammlung, von denen die Forschung bislang annahm, daß sie keinerlei Kennzeichen einer wie auch immer gearteten Organisation aufweisen. Tatsächlich findet man in den übrigen zwölf Gedichten (Gruppe C) keine Dominanten innerhalb der Verse: die Anzahl der Betonungen, die Typen der Klauseln oder der Anakrusis, die Anzahl der Silben, sie alle ändern sich von Zeile zu Zeile. Die Beispiele aus der Gruppe C können mit voller Berechtigung als "völlig freie Verse, ohne strikte metrische und rhythmische Struktur" (ORLOV 1966: 128) oder als "Verse ohne jedes Versmaß" (BOBROV 1965: 124) bezeichnet werden. Es soll nun gezeigt werden, daß auch in den Gedichten dieser Gruppe tatsächlich ein rhythmusorganisierender Faktor wirkt, der den Eindruck einer - im Vergleich zur Prosa - starken Rhythmisierung des Textes entstehen läßt.

Durchgängige Wiederholungsmaße wird man in der Gruppe C vergeblich suchen: es gibt weder rhythmische noch phonetische Dominanten (Reim, Konsonanz, Binnenassonanzen u.ä.), die die einzelnen Zeilen etwa zu einem sinnvollen Ganzen verbinden könnten. Nichtsdestoweniger ist aber der Rhythmus nach den Worten von Žirmunskij gerade derjenige Faktor, der es uns erlaubt, solche Werke als Gedichte zu verstehen und wie Gedichte zu drucken. Diese Behauptung soll nun bewiesen werden, indem wir von der folgenden bekannten Annahme Kvjatkovskijs ausgehen: "Im freien Vers gibt es keine Periodizität der Wiederholungen. Die Zeilen sind hierbei unterschiedlich groß und quantitativ inkommensurabel." (KVJATKOVSKIJ 1963: 63).

Mit Hilfe einer Untersuchung des Wechsels von betonten und unbetonten Silben in den Zeilen von Gedichten überhaupt ohne Versmaß stellen wir fest, wie gerade die rhythmisch kommensurablen Zeilen in Gedichten angeordnet sind. Als kommensurabel werden die Zeilen betrachtet, die ein gemeinsames Maß haben, und als gemeinsames Maß wählen wir eine Periode von Wiederholungen von betonten Silben innerhalb einer Zeile. Wir werden nur solche kommensurablen Zeilen betrachten, die in unser Schema der regelmäßigen zwei- oder dreigliedrigen Versmaße hineinpassen. Die Grundannahmen für eine eindeutige Differenzierung der Zeilen sind die folgenden:

a) choreische und jambische Zeilen, wie z.B.

"podvesti glaza menja iskusnej
i každyj palec napitat'
otdel'nym aromatom ..."

werden im folgenden als äquivalent betrachtet, weil sie dasselbe Maß (zwei Silben) aufweisen. In ähnlicher Weise gelten auch Verse mit drei Silben als äquivalent:

"v tom že dome,
na toj že posteli,
gde rodilis' i umerli dedy";

b) verkürzte Zeilen, die - in Abhängigkeit von ihrer Umgebung - sowohl als zwei-, wie auch als dreigliedrige Versmaße gelesen werden können, wie etwa:

"...použ̌inat'",
"iz okeana",
"ja ostanovilsja",

gelten als neutral und werden bei der Zählung der kommensurablen Zeilen nicht berücksichtigt;

c) bei der Ausgliederung der äquivalenten Zeilen ist die Länge unwesentlich, d.h. es gelten auch die folgenden Zeilen als rhythmisch kommensurabel:

"...pirožkov i orechov,
mal'čik budet machat' opachalom,
a my budem smotret'".

Unsere Hypothese über den rhythmusorganisierenden Mechanismus in denjenigen Gedichten der "Alexandrinischen Gesänge", die reinen *vers libre* aufweisen, lautet folgendermaßen: rhythmisch kommensurable (äquivalente) Zeilen sind in einem Gedicht nicht zufällig verteilt, und es gibt eine Tendenz zu einer Häufung von Zeilen mit dem gleichen Maß.

Ju.I. LEVIN (1967) entwickelte ein mathematisches Verfahren, mit dem man den Kompaktheitsgrad eines Elementes aus einer beliebigen Folge quantitativ bewerten kann. Gegeben sei eine beliebige Folge, die aus Elementen der Typen α und β besteht. In unserem Fall könnte dies eine Folge von Zeilen eines von uns betrachteten Gedichts sein, in dem Zeilen vorkommen, die über das gegebene Merkmal verfügen (kommensurable), und solche, die es nicht aufweisen. Eine endliche Folge (Ausschnitt) kann dann aus n Exemplaren mit dem Symbol α und aus m mit dem Symbol β bestehen. Diese Symbole, z.B. α, können in dem Ausschnitt entweder völlig gleichmäßig verteilt sein (αβββαβββαβββα), oder sie können als kompaktes Segment des Ausschnitts auftreten (ββααααααββ). Im allgemeinen ist ein Symbol in einer Folge völlig willkürlich verteilt; nur gelegentlich kann man innerhalb eines Ausschnittes eine Tendenz zu einer kompakten Anhäufung (oder gleichmäßigen Anordnung) vorfinden.

Wir wollen die rhythmisch kommensurablen Zeilen der "Alexandrinischen Gesänge" mit α und alle übrigen mit β bezeichnen. Dann ergibt sich z.B. für das Gedicht "Ja sprašival mudrecov Vselennoj" die Folge βααββαββαβααβββββββ, wobei α die Zeilen mit dreigliedrigen Versmaßen und β alle übrigen bezeichnet.

Das Berechnungsschemata, das im folgenden zur Bestimmung des Kompaktheitsgrades der äquivalenten Zeilen in einem Gedicht verwendet wird, sieht so aus:

1. Es wird der Abstand d_i zwischen dem i-ten und dem i+1-ten Auftreten des Elements α in der Folge bestimmt; d_i ist also die Anzahl der Elemente β, die zwischen α_i und α_{i+1} liegen. Für die Folge der Zeilen aus dem Gedicht "Ja sprašival mudrecov Vselennoj" ist $d_1 = 0$, $d_2 = 2$, $d_3 = 2$ usw. Die Summe aller d_i ergibt die Gesamtzahl der β-Symbole der gegebenen Folge $\sum_{i=1}^{n} d_i = m$.

2. Es wird der Quotient $d_i/m = \delta_i$ berechnet.

3. Die δ_i-Werte werden zur Basis 2 logarithmiert, und die Werte von $\log \delta_i$ werden mit δ_i mit demselben Index i multipliziert.

4. Es wird die Summe $\sum_{i=1}^{n} -\delta_i \log \delta_i$ berechnet.

5. Der Wert n wird zur Basis 2 logarithmiert.

6. Die errechneten Werte werden in die Formel

$$L = 1 - \frac{\sum_{i=1}^{n} (-\delta_i \log \delta_i)}{\log n}$$

eingesetzt, die den Zahlenwert des gesuchten Kompaktheitsmaßes L für das Element α liefert.

7. Der Wert L gibt noch nicht den tatsächlichen Kompaktheitsgrad an. Er muß zunächst mit einem anderen Wert, dem Hilfswert $\bar{L}(n)$, verglichen werden. Dieser Wert entspricht einem "zufälligen" Auftreten von α in der Folge. Der analytische Ausdruck $\bar{L}(n)$, die Funktion der Anzahl der Symbole n in der Folge, wird mit der Formel

$$\bar{L}(n) = 1 - \frac{1/2 + 1/3 + \ldots + 1/n}{\ln n}$$

berechnet.

8. Erst der Quotient $L/\bar{L}$ ergibt den Zahlenwert des Kompaktheitsgrades. Wenn $L \ll \bar{L}(n)$ ist, bedeutet dies, daß α in unserer Folge unnatürlich gleichmäßig verteilt ist. Wenn die Werte L und $\bar{L}(n)$ nahezu übereinstimmen, gilt die Verteilung von α in dem betrachteten Ausschnitt als zufällig, und schließlich nimmt man bei

$L > \bar{L}(n)$ an, daß α in unserer Folge in Form von kompakten "Inseln" verteilt ist, was beim Fehlen von organisierenden Faktoren nicht vorkommen kann.

Die Ergebnisse der Berechnungen sind in den Tabellen 1 und 2 dargestellt. Für fast alle Gedichte, die im Ganzen in freien Versen verfaßt sind, wurden die Quotienten $L/\bar{L}(n)$ für kommensurable zwei- bzw. dreigliedrige Zeilen getrennt berechnet. Die Gedichte, für die keine Berechnung des Kompaktheitsgrades vorliegt, enthielten entweder gar keine "richtigen" Zeilen, oder nur so wenige, daß eine Interpretation des erhaltenen Zahlenwertes nicht gerechtfertigt gewesen wäre.

Die Mehrzahl der berechneten Quotienten liegt signifikant über Eins, und dies ist eine quantitative Bestätigung unserer Hypothese. Die kommensurablen (äquivalenten) Zeilen haben die Tendenz, nicht zufällig in den Gedichten angeordnet zu sein und Anhäufungen zwischen den übrigen Zeilen zu bilden. Hier eines der deutlichsten Beispiele für solche Anhäufungen (die äquivalenten Zeilen sind unterstrichen):

"Choče$\check{\text{š}}$',
pobežim vperegonku,
po dorožke, obsaženn$\check{\text{o}}$j želtymi rozami,
k ozeru, gde plavajut zelotye rubki?
Choče$\check{\text{š}}$',
pojdem v besedku,
nam dadut sladkich napitkov,
pirožkov i orechov,
mal'čik budet machat' opachalom,
a my budem smotret'
na dalekie ogorody s kukuruzom?
Choče$\check{\text{š}}$',
ja spoju grečeskuju pesnju pod arfu
tol'ko ugovor:
ne zasypat'
i po okončanii pochvalit' pevca i musykanta.
Choče$\check{\text{š}}$',
ja stancuju osu
odna na zelennoj lužajke
dlja tebja odnogo? ("Segodnja prazdnik").

Tab. 1: Rhythmikdefinitoren der ersten Gruppe

Gedichte mit fixierten (durchgängigen) Rhythmikdefinitoren	Rhythmikdefinitoren: Betonungen in einer Zeile	Rhythmikdefinitoren: Anakrusis	Rhythmikdefinitoren: Klausel	syntaktische Parallelismen	Ausmaße des Gedichtes: mittlere Zeilenlänge	Ausmaße des Gedichtes: Anzahl der Zeilen	Kompaktheitsgrad der Zeilen: zweigliedrige	Kompaktheitsgrad der Zeilen: dreigliedrige
Kogda ja tebja v pervyj raz	3-4		W		10.4	9	wurde nicht berechnet	nein
Ty - kak u gadatelja otrok	3		W		9.7	10		"
Naverno, v polden' ja byl zacat	3	W			8.9	8		"
Ljudi vidjat sady ç domami	3-4		W	+	9.7	14		"
Kogdla utrom vychožu iz doma	3		W	+	10	14		"
Ne naprasno my čitali bogoslovov	3-4		W		11.1	11		"
Eslib ja byl drevnim polkovodcem	1-5		W	+	9.0	32		"
Kak pesnja materi				+	7.8	24		1.97
Kogda mne govorjyt Aleksandrija	4-3	W	W	+	11.5	16		nein
Večernij sumrak nadteplym morem	4		W	+	11.3	10		"
Ach, pokidaju ja Aleksandriju	4-3		W	+	9.7	19		2.03
Nas bylo četyre sestry	6	W	W	+	16.3	20		nein
Vesnoja list'ja menjaet topol'		W	W	+	9.2	12		"
Ich bylo četvero v ètot mesjac			W	+	12.6	23		"
Ja spraŝival mudrecov vselennoj'	3	W/d	W	+	8.8	18		1.96
Čto za dožç'	3	W			8.7	20		2.62

Tab.2: Rhythmikdefinitoren der zweiten Gruppe

Gedichte mit schwankender Metrizität	syntaktische Parallelismen	Ausmaße des Gedichtes		„regelmäßige" Zeilen in %		Kompaktheitsgrad der Zeilen	
		mittlere Länge	Anzahl der Zeilen	zweigliedrige	dreigliedrige	zweigliedrige	dreigliedrige
Sladko umeret'		6.2	37	40	35	2.47	1.95
Segodnja prazdnik	+	7.2	37	30	30	2.76	2.60
Čto ž delat'		6.3	40	20	35	3.00	1.89
Razve nepravda	+	8.4	37	27	32	2.32	1.66
Solnce, Solnce		7.5	30	30	23	1.50	2.30
Kogda menja proveli skvoz' sad		8.0	29	20	14	1.43	2.47
Kak ljublju ja, večnye Bogi	+	8.0	26	27	23	2.43	0.86
Ne znaju, kak èto slučilos'		8.0	60	-	38	nein	2.51
Syn moj		7.9	31	10	22	"	3.05
Snova uvidel ja gorod		8.2	48	17	15	2.00	-

Manchmal kann man auch eine Umsetzung der rhythmischen Inertheit beobachten: hier alternieren die Kompaktheits-"Inseln" von zwei- und dreigliedrigen Zeilen. Als Beispiel sei das Gedicht "Razve nepravda" angeführt:

"Razve nepravda,
čto ja - pervaja v Aleksandrii,
po roskoši dorogich uborov,
po cennosti belych konej i serebrjannoj sbruj,
po dline černych kos chitrospletennych?
Čto nikto ne ummet
podvesti glaza menja iskusnej
i každyj palec napitat'
otdel'nym aromatom?"

Hier wird deutlich sichtbar, wie die Zeilen des dreigliedrigen Versmaßes (durchgezogene Linie) allmählich den zweigliedrigen weichen, d.h., es findet ein Wechsel der Periode statt.

Wir wollen nun die Behauptung, daß die kommensurablen Zeilen in einem Gedicht "Anhäufungen" bilden, an einem Beispiel verdeutlichen, wobei wir auf die gesonderte graphische Darstellung der

Zeilen verzichten. "Razve men'še ja budu ljubit'/eti milye chrupkie vešči/ za ich tlennost'".

Das Schema dazu lautet:

UU'UU'UU'UU'UU'UU'U UU'U.

Es ist nahezu unmöglich, daß in einem Prosatext ein Abschnitt mit solch zahlreicher Wiederholung eines Dreifußes auftritt. Ein weiteres Beispiel:

"Možet byt', guby moi/ i kosnulis' ego, ja ne znaju..."

Das Schema dazu:

'UU'UU'UU'UU'UU'U.

Damit ist bewiesen, daß sich die "Alexandrinischen Gesänge" in vier Gruppen gliedern, die jeweils ihren eigenen rhythmusregulierenden Mechanismus haben. Vier von fünf Gedichten des Zyklus "Kanopskije pesenki" (Totengesänge) sind in "metrischen" Versen (KVATKOVSKIJ 1966: 160) geschrieben. Die - im Unterschied zu den "Kanopskije pesenki"-im *vers libre* geschriebenen Gedichte sind ihrerseits wieder in zwei Gruppen gegliedert. Zur ersten Gruppe gehören alle Gedichte der Zyklen "Auftakt", "Liebe", "Abschluß", drei Gedichte des Zyklus "Sie" und je eines aus den Zyklen "Weisheit" und "Ausschnitte" (Tab. 1). Die Gedichte dieser Gruppe sind bestimmt durch die gleichbetonten Zeilen und die Konstanz der Klauseln und der Anakrusis (weiblich). Die Mehrzahl dieser Gedichte besteht aus syntaktisch ähnlichen Zeilen. Meistens äußert sich diese Ähnlichkeit in einem syntaktischen Parallelismus der Zeilen, der manchmal noch durch die anaphorische Wiederholung eines Wortes am Anfang benachbarter Zeilen verstärkt wird: "Raznuju krasotu ja uvižu, / v raznye glaza nasmotrjus', / raznye guby celovat' budu, / raznym kudrjam dam svoi laski, / i raznye imena ja šeptat' budu / v ožidan'i svidanij v raznych roščach".

Die zweite Gruppe von Gedichten besitzt keine sich wiederholenden (durchgängigen) rhythmischen Dominanten. Der Eindruck der rhythmischen Organisiertheit entsteht hier durch die Verdichtung der kommensurablen metrischen Zeilen, die die Erwartung eines Metrums erwecken. Innerhalb dieser Gruppe findet man sechs bimetrische und vier monometrische Gedichte (die Erklärungen dieser Termini wurden oben gegeben).

Es bleiben noch zwei Beispiele des "reinen" *vers libre*, die in keine der genannten Gruppen der Sammlung passen. Möglicherweise hat A.A. Žovtis recht, wenn er behauptet, daß lediglich die graphische Darstellung die Zeilen solcher Gedichte in ein bestimmtes, nicht prosaisches Verhältnis zueinander bringt.

Zum Schluß soll betont werden, daß es in dieser Arbeit hauptsächlich darum ging zu beweisen, daß der *vers libre* in den "Alexandrinischen Gesängen" rhythmisch nicht homogen ist. Außerdem sollte die rhythmusorganisierende Funktion gezeigt werden, die die veränderliche Metrizität hat, indem sie die Lesereinstellung auf die Erwartung eines Metrums hinlenkt und genau in den Fällen wirkt, wo durchgängige Wiederholungsmaße, die die Gedichtreihen zu einem Ganzen aneinanderfügen, nicht mehr vorhanden sind.

Die mathematischen Verfahren, die zum Beweis dieser veränderlichen Metrizität herangezogen wurden, präzisieren nur das, was Žirmunskij vor einem halben Jahrhundert bereits vorausahnte: "... die Schwankungen der rhythmischen Welle, die bald stärker anschwillt, bald schwächer abfällt, ... stellen die wichtigste rhythmische Besonderheit des freien Verses dar" (ŽIRMUNSKIJ 1975: 531).

LITERATUR

BOBROV, S.P., Tesnota stichovogo rjada. "Russkaja literatura" 3, 1965

KVJATKOVSKIJ, A.A., Russkij svobodnyj stich. "Voprosy literatury" 12, 1963

KVJATKOVSKIJ, A.A., Poetičeskij slovar'. Moskau 1966

LEVIN, Ju.I., O količestvennych charakteristikach raspredelenija simvolov v tekste. "Voprosy jazykoznanija" 6, 1967, 112-121

ORLOV, V.N., Na rubeže dvuch ėpoch. "Voprosy literatury" 10, 1966

ŽIRMUNSKIJ, V.M., Kompozicija liričeskich stichotvorenij. In: Teorija sticha. Leningrad 1975

REVIEW

Ipo Tapani PIIRAINEN and Mariann SKOG-SÖDERSVED, Untersuchungen zur Sprache der Leitartikel in der Frankfurter Allgemeinen Zeitung . Vaasa: Vaasan Korkeakoulun Julkaisuja. Tutkimuksia No. 84. Philologie 10, 1982, 95pp.

The investigation of mass communications media is currently one of the most interesting and controversial issues in computerized information processing. The main feature of the book reviewed here is its double-barelled approach, combining consistent analysis of mass media as a whole with the quantitative analysis of editorials printed in the Frankfurter Allgemeine Zeitung. In accordance with such an approach this interesting book consists of three parts: a) a general analysis of mass communications and the press, b) quantitative analysis of the lexical system of newspaper German and c) quantitative analysis of the syntax of the editorials.

First the authors investigate mass media as a multi-step flow, which presupposes the existence of such components as

- an initiator of the information
- a reporter
- a "gatekeeper" (who examines preliminary informational material, means of communication, and the utterance itself), and
- the recipients (opinion leaders and readers in the general public.)

In carrying out their research the authors have framed a hypothesis involving a two-step mass media flow: a) the reception of information by opinion leaders and b) the reception of secondary information by readers from the general public. From the authors' point of view editorials are addressed to opinion leaders.

The next part of the preliminary investigation concerns the main functions of the mass media process and (as a consequence) of the special characteristics of mass media language as distinct from literary language (as defined according to the standard of the aouthors) and "language for special purposes."

Detailed analysis of the mass media process and its special characteristics is determined by the special character of editorials and the authors' particular interest in its comprehensive quantitative analysis.

The statistical aspects of the research undertaken include detailed analysis of fourteen editorials published in the Frankfurter Allgemeine Zeitung in September 1978. The sample size is 10,333 running words (=516 sentences), in other words a rather modest corpus, one which can hardly be considered representative, nor large enough to make the results of the investigation anything like statically reliable. The fact of this small sample may explain the differences between the quantitative results obtained by the authors and those of other investigators of German.

Problems of sampling and sample size have been discussed for many years and their necessity is in the meantime beyond all question. Thus it seems strange that the authors of such a complex and detailed investigation disregard in this fashion the criterion of the statistical reliability of the results.

It is absolutely incomprehensible that - involved as they are in statistical research - the authors do not utilize the results of such fundamental works as Hofmann, L., Piotrowski, R., Beiträge zur Sprachstatistik, Leipzig: VEB Enzyklopädie 1979, or Aleksejew, P., Statistische Lexikographie, Bochum: Studienverlag Dr. Brockmeyer 1984.

In spite of these manifest deficits, due to its scrupulous analysis of each lexical and syntactic form the book itself and its quantitative results are of great interest. The authors examine four basic parts of speech - nouns, adjectives, verbs and adverbs. In addition they undertake a statistical description of the occurrence of personal, relative and substantivized pronouns.

Their quantitative research on syntax embraces an analysis of verb forms, the length of sentences, sentence paterns and so on. But at the same time the quantitative analysis was not supplemented by an investigation of the social basis of the results received, although such an investigation can be of importance and although a good example of it is available:

Naro, A.T., "The Social and Structural Dimensions of a Syntactic Change", Language 57, 1981, 63-98.

The book under review is a very interesting example of complex research conducted with the help of content analysis, lexical, morphological and syntactical analysis, supplemented by quantitative analysis.

Our main criticism of the work is this: the authors of such a complex investigation should be familiar with and use not only literature appearing in German, but fundamental results achieved in the field elsewhere in the world. Ignoring this necessity in research makes it nearly impossible to use the results as a basis for further investigation.

I. Beliaeva
E. Shingariova

NEW PROJECTS

QUANTITATIVE AND QUALITATIVE RESEARCH BASED ON THE LINGUISTIC DATABASE OF FRISIAN

Hans Stellingsma

As the linguistic database project of the Frisian Academy is still in knee breeches and as yet probably completely unknown, we have opted for this occasion for an introductory kind of paper in which we hope to present a global outline of our project. We will therefore start by saying something about Frisian, the relatively unknown target language, this is followed by a brief survey of the general set-up of the database; we then continue by relating how the input is selected and how a first pilot corpus is being compiled. Finally, we suggest a number of quantitative and qualitative analyses which we want to undertake on the basis of the collected materials.

0. GENERAL INTRODUCTION

Frisian is a West Germanic language that - in its oldest stage - belongs, together with Old English and Old Saxon, to the Northsea Germanic group. Traditionally, three stages of the Frisian language are recognized: Old Frisian, Middle Frisian and Modern Frisian. These three developmental stages do, however, not coincide with the periodization of the other Germanic dialects and in this respect the terms 'Old' and 'Middle' need specifying. Old Frisian is the language found in a number of manuscripts, charters and one incunabulum that date from the 11th to the 16th centuries. The Old Frisian sources derive from an area ranging from the Weser in the East to the IJsselmeer in the West. Though the oldest MSS are relatively young compared to other Old Germanic sources, they often contain texts that are very much older. Linguistically, these early texts reflect features that justify calling this stage Old Frisian and make it comparable with neighbouring Germanic dialects.

Middle Frisian is the term used for the language of the poet Gysbert Japicx, who lived from 1603 to 1666, and that of the literature of the 17th and 18th centuries. Though there are no specific linguistic reasons for considering the period 1550-1800 a separate

stage in the process of linguistic evolution, it has been designated so and will for the time being be continued to be referred to accordingly.

Post-1800, Modern Frisian is commonly split up in three dialects, viz. West Frisian, North Frisian and East Frisian. As we shall only be concerned here with West Frisian, the latter two dialects will only be briefly mentioned.

North Frisian is spoken by some 10.000 people living either in the coastal areas of Northern Schleswig Holstein (bordering Southern Denmark) or on the islands before these coasts, like Helgoland, Sylt, Amrum and Föhr.

East Frisian is nowadays spoken by less than a thousand people living in three villages in the Saterlandic region (close to the German town of Leer).

Modern West Frisian, then, is the language spoken in the Dutch province of Friesland. This province has approximately 600.000 inhabitants of which there are some 400.000 speakers of Frisian. It is this variety of Frisian which was recognized as the second official language in the Netherlands some 30 years ago and that forms the object of investigation for the linguistic department of Frisian Academy. This institute occupies itself with the scientific research of Friesland in aspects relating to its inhabitants, their language or their history.

1. AN LDB OF FRISIAN

It was only in the beginning of last year that the data base of Frisian was started by the linguistics department of our Academy. Right from the outset its aims were formulated as: establishing, maintaining and improving a central computer archive & research tool for doing Frisian linguistics.

This meant that a data base system needed to be set up that met at least the following requirements:

- It should be a representative inventory of Frisian in all its aspects and developmental stages.
- It should be a flexible and optimally accessible system enabling linguists to investigate the inventory in a variety of ways.

- A minimum set of basic facilities should be offered relating to: storage of primary texts, enhancing primary tests, parsing, lexicometric research and lexicography.

With these starting-points we can now proceed to relate how the LDB will be set up by discussing trajects and ways of input, selection criteria and representativity.

2. INPUT

We distinguish three separate trajects of input, which together form our master corpus (MC).

The first traject is the one formed from a string of sampled quanta of language. With this traject we aim to regulate the overall representativity of the database as a whole by selecting samples on the basis of findings from specially designated representativity studies which will take place at regular intervals. The second traject is contracted externally and concerns machine-readable output from daily newspapers, magazines, book publishers and the like. In this way we intend to build up the bulk part of our archive. Here selection plays only a minor role, for one can be but grateful for every piece of machine-readable material one can get.

Finally, a third traject is formed by, and at the same time will form, the lexicon. For input we have designated at least our two existing major dictionaries and an as yet undetermined collection of wordlists, glossaries and minor lexicographical works. Part of this lexicon will be carefully constituted to form our basic-lexicon (BL) and this BL serves as a first filter through which all input is directed.

3. SELECTION AND REPRESENTATIVITY

When one wants to compile a linguistic archive that will - after a certain period of time - have to meet a preset configuration of requirements, one will have to take into account all sorts of weighing factors which determine the extent to which the eventual collection will be balanced.

The dichotomy between spoken and written forms of language is the first one would like to respect.

For all ways of input for spoken language one is faced with inherent extra steps. This poses a few more or less serious problems, notably the transcription of the spoken materials. One does, however, realize that if any claim for linguistic representativity is to be made, it involves a considerable portion of spoken discourse to be taken up into a LDB.

If and when money allows us, we will constitute a small corpus, say one of some hundred thousand spoken words. This will at least enable us to undertake some basic comparative research, whereby one of the first things that will need looking into, is finding the proper balance between spoken and written language. Leaving this first binary classification be for now, we will turn to the question of the representativity of the collection.

Representativity is a notion one can measure only if one knows what the universe is one wants to represent. Nature and size of the linguistic universe of any one language are incredibly hard to define. Frisian, being a threatened minority language, is even more difficult to delimit. It undergoes all sorts of influences from Dutch, slowly but certainly it is incringed upon by the much more powerful neighbouring language. Any language as a closed system is frayed at its linguistic edges, Frisian is frayed outside and inside. Within the geographical area, in which Frisian is being spoken, linguists have tried to establish a limited set of definable dialects, these too are at an ever increasing speed more and more influenced from outside. Still, it is possible to reach a consensus on what can be considered as standard Frisian. This model, artificial as it may be, is found in the majority of newspaper items, literature and a great number of publications dealing with a wide variety of every-day topics. A very small number of people even use Frisian for special purposes like publishing research results or advertising. A growing number of civil servants working in the province of Friesland use Frisian as their primary medium for their administration. To capture most of the dimensions along which we define linguistic variety, we have formulated the following - as yet provisional - set of variables that will serve to find the selection criteria for each new sample corpus and which will allow us to define the changing character of the master corpus.

4. VARIABLES

First of all we employ a temporal variable. Starting from the principle that one of the first concrete products of our LDB will be a concise monolingual dictionary of contemporary Frisian, and knowing that the *terminus ad quem* for the largest dictionary of Frisian yet, the WFT, is 1950, it has been decided to take this very year as the *terminus a quo* for the collection of our LDB. As soon as it is thought feasible to do so, we will work back in time per decade in order to eventually cover the whole period of modern, i.e. post-1800 Frisian. While in this process, we are able to study the coverage of existing dictionaries in retrospect. A balanced selection of sources that are to constitute the sample corpora for a decade, will thus enable us to compare the lexicon arising from an 'LDB-decade' with existing dictionary coverage.

A second variable is a dialect-geographical one. Here we will need to study how and in what proportion we want to incorporate individual dialects of Frisian. In the present constellation this is of low-priority, as we cannot spare the proper personnel to undertake such specific dialectal studies. In the next few years, however, we intend to start a pilot-study to this end which will at least enable us to define what and how much material can be considered as being non-standard.

A third variable is a stylistic/sociolinguistic one. With this variable we will try to capture variation of register, i.e. along the axis formal - informal; while at the same time idiolectal variation will be reckoned to belong here. Then, there are Labovian questions like age, sex, education and social position of the authors that need to be taken into account.

A major question here, as with the dialectal variable is the one of differentiation, especially when one only takes written material into consideration. Where does one find enough and more importantly, enough differentiated material on which linguistic research of different kinds could be based?

A fourth variable deals with content matter. Here we start from the SISO classification and divide the Frisian materials according to these headings. Within a particular heading we want to respect a proportional representation of subheadings just as with the overall representation. One has to realize, however, that there is an -

as yet - unknown relationship between a classification as to subject matter and the exact nature of the vocabulary found in sources belonging to a specific heading. Intuitively, one aims for the widest possible spread over all imaginable sub-vocabularies, as this will yield the highest coverage of the absolute vocabulary of the language. How high the chance is of finding particular terms, constructs, phrases or idioms, when only rating a classification according to subject, remains undetermined until specific attention has been paid to this aspect.

One will readily understand that it is possible to devise a grid whose meshes are extremely fine so that the danger of missing sub-vocabularies (or even words!) is minimized. We will opt for a common-sense approach in this respect, until there are counter-indications that force us to use a tighter grid. In order to maintain a workable grid and yet satisfy particular lexicographical needs, the method of excerption is still available.

5. PILOT CORPUS

After a set of variables has been thus determined, one can start to select a first sample corpus. For our purposes we chose some 20 novels which amongst them contain approximately one million word tokens. The sampling principle was dictated by the temporal variable, viz. the period 1950-1970; we furthermore selected one work per annum and aimed at a certain spread within the category literary prose (e.g. not more than one book per author, not too long or short a novel, both complete novels and short story collections to be represented over the period). For the time being, all our texts will be taken up into the LDB integrally, later we may want to decide to change sampling techniques in such a way that only parts of a complete text will be processed for input. As this is our first sample we determined its size to be at least one million word tokens. This quantity enables us to cover a few basic aspects such as the graphic system (letter frequencies, word-length, punctuation and spelling variation) and the phonological system (again with relatively few discrete units and fairly simple rules of combination). Furthermore, we needed a manageable quantity of data by means of which we could get some experience in the computing aspects of linguistic data.

In the way of quantitative analysis we want to undertake at least the following investigations:

1. the relation between wordlength (measured in number of syllables) and syllablelength (measured in phonemes/characters) (Menzerath Law)
2. the relation between wordlength (measured either in syllables or phonemes/characters) and the number of word-meanings. (Here too we have the Menzerath Law)
3. type/token relations
4. frequency of types/lemmas
5. dispersion indices, usage coefficients (cf. those proposed by Juilland, Rosengren and Carroll)
6. rank-frequency relation (Zipf Law)
7. wordlength distribution
8. relation between the number of native Frisian words (i.e. predecessors present in Old Frisian) and the number of non-native ones
9. relation between the number of Dutch words occurring in (basically) Frisian texts

6. PROSPECTS AND CONCLUSION

We cannot elaborate on all aspects of the project as this falls outside the scope of this paper. Yet we shall definitely mention some of them here, as they form essential steps in the process of building the LDB system into a definite form.

First of all there is the suite of programs that processes the texts from their raw state of entry to a standardized format from which lists can be compiled, and on the basis of which further linguistic research can be done. Here we have envisaged a series of standard facilities which serve the linguist for his basic toolbox. Next to lists stating absolute and relative frequencies of types, we offer a lemmatised list, either alphabetical or retrograde, a Key Word In Context (KWIC) list, or a fully detailed concordance. All these are available for any subset from the Master Corpus one might want to define, for instance on the basis of the variables mentioned above. Then there are several programs for different

sorts of linguistic analysis that we want to offer our prospective clients. Among these the most important ones will be a syntactic parser and a lexicographically oriented databank for storage and retrieval of lexidata which will eventually lead to the compilation of new dictionaries.

Some two near-future goals have already been formulated here:

- the production of a monolingual dictionary of present-day West Frisian
- The compilation of a series of bilingual dictionaries, starting with Frisian into English

A series of prepatory studies will eventually lead to the production of a comprehensive grammar of modern West Frisian, while other projects - syntactic and lexicographical, that deal with earlier stages of the language - have been contemplated, but not yet definitively formulated.

H. Stellingsma
Frisian Academy
Doelestraat 8
8911 DX Leeuwarden
The Netherlands

MUSIKOMETRIKA

A new series on mathematical musicology

M.G. Boroda

Among the problems of contemporary linguistics, concerned with the investigation of non-verbal communicative systems, the problem of general principles and regularities of musical language is one of the most interesting and at the same time most complicated. From time to time it has attracted linguists' attention, but few linguists have conducted serious research on it, due to the real complexity of music as a semiotic and semantic system and to the ambiguity of musicological terminology concerning the basic components of musical language: structural units, metro-rhythm, etc.

The development of "mathematical musicology" has resulted in a certain amount of progress in this area - especially as regards the investigation of the "language of musical style", generative grammar modelling of musical compositions, etc. Journal publications show that there is a real necessity for and a real possibility of intensifying and expanding the investigation of musical language, to study these phenomena from various points, with various methods, including those of "traditional", "structural", and "quantitative" linguistics.

Bearing all this in mind and being interested in promotion of qualitative and quantitative mathematical investigations of musical language, the editorial board of QUANTITATIVE LINGUISTICS announces a special series "Musikometrika". The main aim of the new subseries is not to give the specialists in mathematical musicology more publication possibilities, but to make an attempt to organize, on a general basis, a range of research programs aimed at the broad and detailed structural/quantitative investigations of musical texts and musical language, and specially oriented towards the cooperation between musicology and linguistics - a cooperation which undoubtedly would be stimulating for both areas.

The topics and research areas foreseen for MUSIKOMETRIKA are the following:

(i) Structural units of musical language and musical text: metastylistic principles of formation; rules, definitions and algorithms of segmentation; regularities of functioning.

(ii) Statistical and information theoretical characteristics of the "language of individual style", the "language of a particular school of composition", the "language of a particular epoch": intonational vocabulary (on the basic levels of musical text), structural type of language (in the linguistic sense), grammatical regularities.

(iii) Principles, regularities and motive forces of the development of language of European music of the 10th to 20th centuries: quantitative and qualitative-mathematical models; hypotheses on the causes of the new "intentional crises" in the development of European music in the last four centuries.

(iv) Connections and relationships between music and speech:
  (a) in vocal music: correlations between melody- and text-segmentation on various levels of basic musical units; the analogies and differences between words and musical units respectively in the rhythmic structure, principles of expressing the "phonetic-expressive" program of verse in the romance melody;
  (b) in instrumental music: the problem of using structural-intentional units formed in vocal music (the problem of "semantic transplantation").

(v) Common principles and regularities of the organisation of musical and literary text on various levels: the quantitative mathematical approach, Zipf-Mandelbrot' Law, Menzerath's Law, Fuchs's and Grotjahn's distribution, etc. in musical text.

(vi) The influence of the phonetic- and structural type of language on the music of a given nation (folk music, composed music).

(vii) The analogies and differences between the folk and composed music: rhythm and pitch organisation, structural units.

(viii) Structural and quantitative characteristics of the functioning of mode, rhythm and pitch, as the main components of musical language, in musical composition: stylistic and general principles, regularities of development.

(ix) Mathematical models of musical text, based on the hypotheses and experimental results of investigations of musical perception. Mathematical models of musical language (within the framework of a given composer's style, the style of a particular epoch, etc.) and musical communication.

(x) Applications of psycho- and socio-linguistic models in musicological investigations.

In MUSIKOMETRIKA there would be also published "frequency vocabularies" of various musical texts, dissertations and reviews on the above-mentioned problems, as well as discussions of articles previously published in it. Within the framework of the "psychological approach" a special issue could be dedicated to the problem of multidisciplinary investigation into the creative process of music.

Contributions to MUSIKOMETRIKA are to be sent to M.G. Boroda, State Conservatory, Griboedova 8, 380004 Tbilisi (USSR) or to the editors of Quantitative Linguistics.

QUANTITATIVE TYPOLOGY OF TEXTS

P.M. Alekseev

Quantitative typology of text (QTT) denotes typological studies based on linguistic, probabilistic, systemic and semiotic concepts of language and speech. The QTT aims at obtaining probabilistic and informational models that could describe and explain typological features of complex linguistic objects represented by text. The subject of the QTT is further developing the Saussurean triad "langage= langue + parole" into a tetrad "language phenomenon as a whole=language system + language norm + usage + verbal act and its result, i.e. text".

The text with all linguistic units that it contains is the immediate material for study as the only linguistic reality subject to observation. Text realises and forms the system of language, its norm, functional styles (registers), sublanguages, idiolects. Therefore the QTT reflects typologies of all linguistic systems and subsystems being observed in text.

The theory of QTT comprises the above mentioned linguistic, probabilistic and other compacts of language and speech. The methodology of QTT is built of principles of linguistic sampling and linguistic distribution analysis. The technology of QTT is a system of statistical techniques of collecting and processing linguistic data. The factographic base of QTT is the statistical lexicography that deals with frequency dictionaries and other statistical inventories of linguistic units.

A vital tool of quantitative linguistics is the distribution analysis, but up to now there has not been offered a well-defined, compact and comprehensible set of concepts and notions underlying systematic analysis of distributions of linguistic phenomena. Thus an actual problem is to develop some "fundamentals" of linguistic distribution analysis embodying as much as possibble related information and presenting it in a way understandable to linguists.

LINGUISTIC QUANTA IN QUANTITATIVE LINGUISTICS

V.N. Byčkov

This project offers possible approaches to the choice of linguistic units intended for complex linguistic objects description and evaluation. The main concepts under consideration and the very results obtained due to application of presented methodology may be of interest for all those dealing with the problems of quantitative linguistics.

LINGUOSTATISTICAL METHODS OF KASAKH CHILDREN SPEECH STUDIES

K.B. Bektaev, K.M. Moldabekov, R.G. Piotrowski

This project details some new advances in semiotic and computational and statistical studies of children oral speech and fiction for the children.

ENGINEERING LINGUISTICS AND SYSTEMATIC-STATISTICAL STUDIES OF UZBEK TEXTS

S.A. Muhamedov, R.G. Piotrowski

Over a decade has now passed since there were published the first computer generated indexes, frequency dictionaries, and MT-results concerning Turkic languages, and time is ripe for an overview of the developments in the field. This project will evaluate solutions (and attempted solutions) to a variety of the contemporary Turkology semiotic and computational problems. The project is also intended to serve as an introduction for the novice and as a concise summary for the expert who may not have had the opportunity to examine all the transactions in the field of engineering Turkology first-hand.

Language Synergetics

The sponsor of the above project is the foundation Stiftung Volkswagenwerk, Hannover, West Germany. It is to run from 1 October, 1986 on.

A group of linguists and mathematicians in various parts of the world will be investigating the selfregulation of language in the domain of the lexicon (Part I). They will proceed from the assumption that all properties of words (or lexical items) are in mutual dependence, either directly or by means of feedback, so that it is possible to set up synergetic models of language processes.

Since in order to achieve valid results a number of languages is required, additional collaborators are warmly welcomed. For further information contact Prof. G. Altmann, Sprachwissenschaftliches Institut der Ruhr-Universität Bochum, Postfach 102148, 4630 Bochum, West Germany.

CURRENT BIBLIOGRAPHY

General

1. ALTMANN, G.: On the dynamic approach to language. Ballmer, Th.T. (ed.): Linguistic Dynamics (Dynamics, Procedures and Evolution). Berlin: de Gruyter, 1985, 181-189.

2. ALTMANN, G., BURDINSKI, V.: Towards a law of word repetitions in text-blocks. Lehfeldt, W., Strauss, U. (eds.): Glottometrika 4. Bochum: Brockmeyer, 1982, 147-167.

3. GUITER, H., ARAPOV, M.V. (eds.): Studies on Zipf's Law. Bochum: Brockmeyer, 1982.

4. LAUBER, J., LINHART, J.: Kanonische Analyse von Kontingenztafeln. Lehfeldt, W., Strauss, U. (eds.): Glottometrika 4. Bochum: Brockmeyer, 1982, 168-182.

5. MARTIN, W.: Kwantitatieve Taalkunde. de Geest, W. et al. (eds.): Twintig facetten van de taalwetenschap. Leuven, 1981, 273-291.

6. MARTIN, W.: Möglichkeiten und Grenzen der quantitativen Linguistik beim Studium der wissenschaftlichen Fachsprachen. Bungarten, Th. (ed.): Wissenschaftssprache. München, 1981, 169-184.

Phonology

7. WINKLER, P.: Quantitative Analyse phonetischer Transkripte (Umgangssprache). Lehfeldt, W., Strauss, U. (eds.): Glottometrika 4. Bochum: Brockmeyer, 1982, 1-79.

WRITING

8. DIETZE, J.: Grapheme und Graphemkombinatorik der russischen Fachsprache. Eine phonostatistische Untersuchung. Lehfeldt, W., Strauss, U. (eds.): Glottometrika 4. Bochum: Brockmeyer, 1982, 80-94.

9. GVOZDOVIČ, B.N.: Stabil'nost' častot nemeckich grafem [The stability of German graphem frequencies]. Kvantitativnaja lingvistika i avtomatičeskij analiz tekstov. Trudy po lingvostatistike. Tartu, 1985, 3-8.

MORPHOLOGY

10. DILLER, H.-J.: Forum: Rosy cheeks and bloody shrouds. Das Schicksal der Adjektive des Typs N+ -y in der englischen Dichtungsgeschichte. Anglistik und Englischunterricht 27. 1985, 197-218.

11. GERLACH, R.: Zur Überprüfung des Menzerathschen Gesetzes im Bereich der Morphologie. Lehfeldt, W., Strauss, U. (eds.): Glottometrika 4. Bochum: Brockmeyer, 1982, 95-102.

12. KRÁMSKÝ, J.: A stylostatistical investigation of possessive pronouns in modern English. Philologica pragensia 28. 1985, 152-158.

13. KUZNECOVA, A.I.: Periferijnye javlenija v morfologii russkogo jazyka [Peripheral phenomena in the Russian morphology]. Aktual'nye voprosy strukturnoj i prikladnoj lingvistiki. Moskva, 1980, 128-141.

14. LÜDTKE, H.: The place of morphology in a universal cybernetic theory of language change. Jacek Fisiak (ed.): Historical morphology. The Hague: Mouton, 1980, 273-281.

15. RUSKOVA, M.P.: Statističeskie parametry slovoobrazovania v bolgarskom jazyke 18-go veka [Statistical parameters of word-formation in the 18th century Bulgarian language]. Kvantitativnaja lingvistika i avtomatičeskij analiz tekstov. Trudy po lingvostatistike. Tartu, 1985, 81-92.

SYNTAX

16. BABA, T.: Developmental difference in linguistic intuition of word relatedness in a sentence. Mathematical Linguistics, Vol. 15, No. 5. 1986, 155-167.

17. ŠTĚPÁN, J.: Složité souvětí s řetězcovou závislostí [The Compaind Sentence with Chain Dependency]. Praha, 1977 (1980).

SEMANTICS

18. ROTHE, U.: Die Semantik des textuellen 'et'. (Europäische Hochschulschriften, Reihe 13: Französische Sprache und Literatur, Bd. 112). Frankfurt a.M. - Bern - New York: Peter Lang, 1986, 242 pp.

19. WILDGEN, W.: Archetypensemantik - Grundlagen für eine dynamische Semantik auf der Basis der Katastrophentheorie. Tübingen: Narr, 1985.

LEXICOLOGY

20. BRANCH, M., NIEMIKORPI, A., SAUKKONEN, P.: A Student's Glossary of Finnish. The Literary Language Arranged by Frequency and Alphabet. English-French-German-Hungarian-Russian-Swedisch. Porvoo-Helsinki - Juva: Werner Söderström Osakeyhtiö, 1983.

21. Frekvenční slovník současné českе publicistiky [Frequency Dictionary of Present-Day Newspaper Czech]. Praha, 1980, 189 pp.

22. Frekvenční slovník současné administrativy [Frequency Dictionary of Present-Day Administration]. Praha, 1980, 86 pp.

23. Frekvenční slovník současné odborne češtiny [Frequency Dic-

tionary of Present-Day Technical Czech]. Praha, 1982, 229 pp.

24. Frekvenční slovníku češtiny věcného stylu [Frequency Dictionary of Non-Fictional Style in Czech]. Praha, 1983, 239 pp.

25. GRÄBNITZ, V.: Häufigkeitsuntersuchungen zum Korpus der deutschen Gegenwartssprache. Diss., München, 1982.

26. HOFLAND, K., JOHANSSON, S.: Word Frequencies in British and American English. The Norwegian Computing Centre for the Humanities. Bergen, 1982.

27. MANASYAN, N.S.: Primenenie dzeta-funkcii Rimana pri prognozirovanii slovarnogo sostava pod"jazyka [The application of Riemann's zeta-function in vocabulary structure prediction]. Kvantitativnaja lingvistika i avtomatičeskij analiz tekstov. Trudy po lingvostatistike. Tartu, 1985, 67-80.

28. MARTIN, W.: On the construction of a basic vocabulary. Burton, S.K., Short, D.D. (eds.): Proceedings of the 6th International Conference on Computers and the Humanities. Rockville, 1983, 410-414.

29. NOMURA, M., TSURUOKA, A.: Research on vocabulary in compositions by school children and junior high school students. Mathematical Linguistics, Vol. 15, No. 5. 1986, 168-179.

30. SAUKKONEN, P., HAIPUS, M., NIEMIKORPI, A., SULKALA, H.: Suomen kielen taajuussanasto. A Frequency Dictionary of Finnish. Porvoo - Helsinki - Juva: Werner Söderström Osakeyhtiö, 1979.

31. ŠTĚPÁN, J.: Frekvenční slovník současné odborné češtiny [A Frequency Dictionary of Present-Day Technical Czech]. Jazykovědné Aktuality 20. 1983.

32. ŠTĚPÁN, J.: M. Těšitelová a kol., Nové frekvenční slovníki dvou stylu [New frequency dictionaries of two styles]. Jazykovědné Aktuality 18. 1981, 149-150.

33. ŠUNEVIČ, B.I.: Castotnyj anglo-russkij slovar' po robototechnike [English-Russian frequency list on robotics]. Kvantitativnaja lingvistika i avtomatičeskij analiz tekstov. Trudy po lingvostatistike. Tartu, 1985, 117-126.

Text Analysis

34. ALEKSEEV, R.M.: Metodika kvantitativnoj tipologii teksta [Methods of Quantitative Text Typology]. Leningrad, 1983, 76 pp.

35. ALTMANN, G., BURDINSKI, V.: Ein Gesetz des Textaufbaus. Lehfeldt, W., Strauss, U. (eds.): Glottometrika 4. Bochum: Brockmeyer, 1982, 147-167.

36. CHISHOLM, D.H.: Phonological patterning in English and German verse: A computerassisted approach. Lehfeldt, W., Strauss, U. (eds.): Glottometrika 4. Bochum: Brockmeyer, 1982, 114-146.

37. GRIGOR'EVA, A.S.: Statističeskaja struktura russkogo epistoljarnogo teksta (leksika častnych pisem) [The Statistical Structure of Russian Epistolary Text (Lexicon of Frequent Letters)]. AKD., Leningrad, 1981.

38. KRYLOV, Ju.K.: K voprosu o dinamike narastanija ob"ema slovarja teksta [On the growth of vocabulary size in random samples and connected texts]. Kvantitativnaja lingvistika i avtomatičeskij analiz tekstov. Trudy po lingvostatistike. Tartu, 1985, 55-66.

39. KRYLOV, Ju.K., JAKUBOVSKAJA, M.D.: On some possibilities of applying the quantummechanical formalism when constructing stochastic models of text generation and recognition. Symposium on Grammars of Analysis and Synthesis and their Representation in Computational Structures. Summaries. Tallin, 1983, 49-51.

40. PETR, J.: O jazyce současné české publicistiky v číslech [Figures on present-day newspaper Czech]. Naše Řeč 65.

1982, 248-253.

41. TĚŠITELOVÁ, M.: K jazyku věcného stylu z hlediska kvantitativního [On the language of non-fictional style from the quantitative point ov view]. Slovo a Slovesnost 44. 1983, 275-283.

42. TIEKEN, I.: Do-support in the writings of Lady Mary Wortly Montagu: a change in progress. Folia Linguistica Historica VI. 1985, 127-151.

43. TULDAVA, Ju.A.: Častotnaja struktura teksta i zakon Cipfa [The statistical structure of text and Zipf's law]. Kvantitativnaja lingvistika i avtomatičeskij analiz tekstov. Trudy po lingvostatistike. Tartu, 1985, 93-116.

44. TULDAVA, Ju.A.: Dlina slova i raspredelenie slov po dline v tekste i v slovare [On word-length distribution in text and vocabulary]. Mutt, O. (ed.): Issledovanija po obščemu i sopostavitel'nomu jazykoznaniju (Acta et Commentationes Universitatis Tartuensis). Tartu: Tartuskij Gosudarstvennyj Universitet, 1986, 150-166.

45. TULDAVA, Ju.A.: K voprosu ob analitičeskom vyraženii svjazi meždu ob"enom slovarja i ob"enom teksta [Questions of the analytic expression of the relation between the size of vocabulary and text length]. Učen. zap. Tart. un-ta, vyp. 549. Trudy po lingvostatistike. Tartu, 1980, 113-144.

Psycholinguistics

46. GÜNTHER, H.: Zur methodischen und theoretischen Notwendigkeit zweifacher statistischer Analyse sprachpsychologischer Experimente. Sprache & Kognition 2. 1983, 279-285.

47. GÜNTHER, H., GFROERER, S., WEISS, L.: Inflection, frequency, and the word superiority effect. Psychological Research 46. 1984, 261-281.

48. SPIVAK, D.L.: Lingvistika izmenennych sostojanij soznanija [The Linguistics of Altered States of Consciousness]. Leningrad: Nauka, 1986.

Sociolinguistics

49. WILDGEN, W.: Synergetische Modelle in der Soziolinguistik. Zur Dynamik des Sprachwechsels Niederdeutsch-Hochdeutsch in Bremen um die Jahrhundertwende (1880-1920). Zeitschrift für Sprachwissenschaft 5. 1986, 105-137.

Historical Linguistics

50. ARAPOV, V.M., HERZ, M.M.: Mathematische Modelle in der historischen Linguistik. Bochum: Brockmeyer, 1982.

51. GERD, A.S.: Drevneslavjanskij jazyk i ego tipy po lingvostatističeskim dannym [Old Slavonic and its types on the statistical basis]. Kvantitativnaja lingvistika i avtomatičeskij analiz tekstov. Trudy po lingvostatistike. Tartu, 1985, 9-22.

52. LASS, R.: On Explaining Language Change. Cambridge: Cambridge U.P., 1980.

Language Teaching

53. PETIT, J.: De L'enseignement des langues secondes à l'apprentissage des langues maternelles. Paris-Genève: Champion, 1985, 693 pp.