

SEBASTIAN KEMPGEN<sup>1</sup>

## Syntagmatische Phonemtypologie

### *Syntagmatic Typology of Phonemes*

*Abstract:* Based on the recent developments of the distributional calculus of HARARY and PAPER, a method is proposed to discover the constructional mechanism underlying the phonemic distribution of a language. The phonemes are classified hierarchically and the mutual interdependence of their distributional characteristics is studied. To this end, the technics of correlation and regression are being used. In this way, a “typology” of phonemes in the sense of HEMPEL and OPPENHEIM is arrived at.

### I

Den Äußerungen der Sprecher einer Sprache liegt auf der lautlichen Ebene eine zweifache Gliederung zugrunde, erstens die materielle Implementierung der Laute, zweitens die phonemische Strukturierung, durch die konventionell festgelegt ist, welche phonetischen Unterschiede in einer Sprache funktionalen Charakter tragen. Diese Unterscheidung betrifft den Bereich der *sprachlichen Wirklichkeit*. Eine der Aufgaben des Linguisten ist es, das phonemische System einer Sprache, das nicht direkt beobachtet werden kann, durch ein phonologisches Modell zu beschreiben. Dies ist die Ebene der linguistischen Wirklichkeit<sup>2</sup>. Das phonologische Modell ist also eine hypothetische Annäherung an das phonemische System der sprachlichen Wirklichkeit, darf aber nicht mit ihm gleichgesetzt werden. Selbstverständlich läßt sich ontologisch nicht *beweisen*, daß eine Sprache ein phonemisches System besitzt, aber im Rahmen einer linguistischen Theorie ist es sinnvoll, eine entsprechende Annahme zu machen. Weiterhin geht man gewöhnlich davon aus, daß die Menge der Phoneme einer Sprache kein ungeordnetes Ensemble darstellt, sondern daß zwischen ihnen bestimmte Relationen bestehen, woraus sich überhaupt erst der Gedanke eines „Systems“ ergibt. Zur Beschreibung dieser Relationen kann man sich verschiedener Methoden bedienen, deren Fruchtbarkeit sich aus den mit ihnen erzielten Ergebnissen ablesen läßt.

---

<sup>1</sup> Ich danke W. LEHFELDT, Konstanz, und G. ALTMANN, Bochum, für ihre konstruktive Kritik.

<sup>2</sup> Die Distinktion von sprachlicher und linguistischer Wirklichkeit wird besonders von GRUCZA (1970, 40f.) herausgearbeitet.

Sehen wir von den internen Eigenschaften der Phoneme ab (also der Frage, wie sie sich aus phonologischen Merkmalen konstituieren), so können die Relationen der Phoneme untereinander vor allem in zweierlei Hinsicht untersucht werden. Einmal kann man sie darauf betrachten, wie sie selbst wiederum höhere Einheiten (Silbe, Morphem, Wort) konstituieren. Diesen *funktionellen* Relationen stehen die *distributionellen* Relationen gegenüber (vgl. JUILAND 1961, 24). Diese „Erforschung der Regeln des Vorkommens der Phoneme und der Phonemverbindungen“ (TRUBETZKOY 1932, 40) ohne Rücksicht auf ihre Position in einer höheren Einheit gilt seit langem als fester Bestandteil phonologischer Beschreibung (vgl. auch TRNKA 1936), erfolgt bei TRUBETZKOY charakteristischerweise aber stets unter funktionellem Aspekt, wie seine Beschreibung des Attischen deutlich zeigt (vgl. TRUBETZKOY 1971, 219ff.) – daher auch seine Forderung, die Phonemanalyse hinsichtlich der phonologischen Merkmale durch „funktionelle Einteilungen“ zu ergänzen. Löst man sich von funktionalen Gesichtspunkten, so ergibt sich allgemein als Problem, das sich hinter einer Aufgabe wie der oben genannten eigentlich verbirgt, eine „Typologie des betreffenden Phonemsystems“ (TRUBETZKOY 1971, 219) und seiner Teile. Dies kann nur heißen, daß die *allgemeinen Gesetzmäßigkeiten*, die die syntagmatischen, distributionellen Relationen der Phoneme einer Sprache steuern, aufgedeckt werden sollen, d.h. daß das allgemeine Bauprinzip des syntagmatischen Teils des Phonemsystems einer Sprache ermittelt werden soll, so daß die Phoneme als „Typ“ gekennzeichnet werden können. Wie dies geschehen kann, soll in diesem Beitrag exemplarisch erläutert werden.

Um eventuelle Mißverständnisse von vornherein auszuschließen, sei betont, daß unsere Aufgabe nicht gleichzusetzen ist mit der Untersuchung der „strukturellen“ und der „zufälligen“ Lücken der Phonemdistribution, obwohl auch diese wohl nur dadurch ermittelt werden können, daß Kombinationstendenzen berücksichtigt werden (vgl. ALTMANN 1973a), die aber für einzelne Phonempaare oder -klassen gelten und nicht mit den allgemeinen Gesetzmäßigkeiten des gesamten Phonemsystems identisch sind.

Die Formulierung der gestellten Aufgabe weist schon auf die Parallelität zu den Zielen einer allgemeinen Sprachtypologie hin, so wie sie von ALTMANN und LEHFELDT (1973, 15) programmatisch entworfen wurden. In Paraphrasierung dieser Ziele läßt sich unsere Aufgabe noch einmal folgendermaßen formulieren:

Unter den Zielen einer *syntagmatischen Phonemtypologie* werden wir im folgenden nur zwei Probleme verstehen: a) die *Phonemklassifikation*, d.h. den Aufbau eines Ordnungssystems für die Phoneme einer natürlichen Sprache aufgrund ihrer globalen syntagmatischen Ähnlichkeit; b) die *Aufdeckung des Konstruktionsmechanismus* der Phonemdistribution, d.h. den Aufbau eines Beziehungssystems, eines „Netzes“, an dem man nicht allein die offensichtli-

chen, kategorischen, sondern auch die latenten Mechanismen der Phonemdistribution ablesen kann.

Da dies nur mit quantitativen Methoden geschehen kann, ist es naheliegend, die Prinzipien und Verfahren, nach denen eine allgemeine Sprachtypologie genannter Art ihre Ziele erreichen will, auf den innersprachlichen Bereich zu übertragen und ein Teilsystem, eben die Regelmäßigkeiten der Phonemdistribution, auf diese Weise zu erfassen. Eine ausführliche Begründung der theoretischen Prämissen und der Wahl der praktischen Mittel braucht deshalb in diesem Rahmen nicht gegeben zu werden.

## II

Der erste Schritt besteht darin, die Distribution der Phoneme einer Sprache<sup>3</sup> zu erfassen und in einer Matrix zu fixieren. Als Beispiel ziehen wir im folgenden das Attische heran, da sich TRUBETZKOY auf diese Sprache bei der Erläuterung seiner Vorstellungen von „funktionellen (Phonem-)Einteilungen“ stützt. Die Angaben zur Distribution, wie sie Tabelle I wiedergibt, sind AFENDRAS (1970, 98) entnommen<sup>4</sup>. Daran schließt sich eine exakte Beschreibung und Erfassung der Phonemdistribution an. Von ALTMANN und LEHFELDT (1980, Kap. 5) übernehmen wir zwölf grundlegende Maße, die die Distribution eines jeden Einzelphonems numerisch erfassen<sup>5</sup>. In den Formeln der folgenden Maßdefinitionen steht  $|A_i|$  für die Zahl der Elemente in der Menge der *Vorgängerphoneme* von  $i$ ,  $|B_i|$  für die entsprechende Zahl der *Nachfolgerphoneme* von  $i$ ;  $k$  steht für die Zahl der Phoneme im Inventar, ist also für das Attische gleich 27. Durch die bekannten Mengenverknüpfungen  $\cap$  (Durchschnitt),  $\cup$  (Vereinigung) und  $\ominus$  (symmetrische Differenz) lassen sich daraus die verschiedenen Maße ableiten<sup>6</sup>:

---

<sup>3</sup> Das Problem, welches Phoneminventar der Beschreibung zugrunde gelegt werden soll, lassen wir außer acht, da es für das Folgende unerheblich ist.

<sup>4</sup> AFENDRAS' Angaben wurden insoweit abgewandelt, als es sich bei der Untersuchung distributioneller Relationen ja ausdrücklich *nicht* um das Vorkommen in bestimmten Positionen einer (höheren) Rahmeneinheit handelt; die Spalte und die Reihe, die bei AFENDRAS durch # gekennzeichnet sind, wurden deshalb getilgt.

<sup>5</sup> Die ersten, einfachsten Maße stammen bereits von HARRY und PAPER (1957).

<sup>6</sup> Für genauere Angaben zu der mengentheoretischen Schreibweise vgl. ALTMANN und LEHFELDT (1973, 93 ff.) und ALTMANN (1973b).

Tabelle I. Phonemdistribution im Attischen

	I	I:Y	Y:E	EIH	A	A:W	O	U:P	T	K	B	D	G	PH	TH	KH	S	(	M	N	L	R	
I	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
I:		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Y	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Y:	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
E	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EI			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
H	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A:	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
W	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
O	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
U:			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
T	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
K	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
B	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
D	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PH	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
TH	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
KH	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
S	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
(	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
M	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
L	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
R	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

- (1)  $At(i) = \frac{|Ai|}{k}$ , das Maß der Attraktivität von i
- (2)  $Ag(i) = \frac{|iB|}{k}$ , das Maß der Aggressivität von i
- (3)  $As(i) = \frac{|Ai \cup iB|}{k}$ , das Maß der vollständigen Assoziativität von i
- (4)  $As^*(i) = \frac{|Ai| + |iB| - G(i)}{2k - 1}$ , das Maß der Paarbildung von i
- (5)  $At_1(i) = \frac{|Ai|}{|Ai \cup iB|}$ , das Maß der internen Attraktivität von i

- (6)  $Ag_1(i) = \frac{|iB|}{|Ai \cup iB|}$ , das Maß der internen Aggressivität von i
- (7)  $As^*_1(i) = \frac{|Ai|+|iB|-G(i)}{2(|Ai \cup iB|) - G(i)}$ , das Maß der internen Paarbildung von i
- (8)  $S(i) = \frac{|Ai \cap iB|}{k}$ , das Maß der Symmetrie von i
- (9)  $S_1(i) = \frac{|Ai \cap iB|}{|Ai \cup iB|}$ , das Maß der internen Symmetrie von i
- (10)  $S^*_1(i) = \frac{2|Ai \cap iB|}{|Ai|+|iB|}$ , das Maß der (additiven) Symmetrie von i
- (11)  $S'(i) = \frac{|Ai \circ iB|}{k}$ , das Maß der Amtsymmetrie von i
- (12)  $G(i)$ , das Maß der Reflexivität von i, das den Wert 1 annimmt, wenn das Paar ii auftritt, sonst den Wert 0

Neben den Anzahlen der Elemente in den Ausgangsmengen  $A_i$ ,  $iB$ ,  $A_i \cup iB$ ,  $A_i \cap iB$  und  $A_i \circ iB$  (im Computerausdruck durch AI, IB, VE, DU und SD gekennzeichnet) finden sich Werte aller genannten 12 Maße für die Phöneme des Attischen in Tabelle II<sup>7</sup>.

Damit ist eigentlich schon eine Antwort auf unsere Frage nach dem syntagmatischen „Typ“ eines jeden Phonems gegeben, denn alle im weiteren verwendeten Informationen sind in diesen 12 Meßreihen enthalten. Nach HEMPEL und OPPENHEIM (1936, 66f.) nämlich „entspricht es einer elastischen typologischen Begriffsbildung, die Frage nach der typologischen Zugehörigkeit eines Individuums nicht mit der Feststellung zu beantworten, das Individuum gehöre dem und dem Typus an, sondern die Stellung des Individuums in jeder der Abstufungsreihen anzugeben, die zu den einzelnen in der betreffenden Theorie unterschiedenen abstufbaren Merkmalen gehören“. Jedem Phonem wird auf diese Weise „ein Ort im typologischen *Merkmalraum* individuell zugewiesen“ (HEMPEL und OPPENHEIM 1936, 67). Die Ziele einer Typologie sind aber damit natürlich nicht erschöpft, denn bisher stehen die einzelnen Werte beziehungslos nebeneinander, und von einer „Erklärung“ irgendwelcher Sachverhalte kann noch nicht die Rede sein. Diese Frage werden wir in Paragraph IV wieder aufgreifen.

<sup>7</sup> Im Computerausdruck wird ein langer Vokal durch nachgesetzten Doppelpunkt gekennzeichnet, also z.B. /ī/ = I:; H steht für /ŋ/, ( für /h/, den Spiritus asper, W steht für /ω/.

Tabelle II. Distributionsmaße der Phoneme des Attischen

I	AI	IB	VE	DU	SD	AT(I)	AG(I)	AS(I)	AS*(I)	ATI(I)	AGI(I)	ASI*(I)	S(I)	SI(I)	SI*(I)	S'(I)	G(I)
I	24	26	27	23	4	0.89989	0.96296	1.00000	0.92453	0.88889	0.96296	0.92453	0.85185	0.85185	0.92000	0.14815	1
I:	16	24	26	14	12	0.59259	0.88889	0.96296	0.75472	0.61538	0.92308	0.76923	0.51852	0.53846	0.70000	0.44444	0
Y	22	23	25	20	5	0.81481	0.85185	0.92593	0.84306	0.88000	0.92000	0.90000	0.74074	0.80000	0.88889	0.18519	0
Y:	17	21	24	14	10	0.62963	0.77778	0.88889	0.71598	0.70833	0.87500	0.79167	0.51852	0.58333	0.73684	0.37037	0
E	27	24	27	24	3	1.00000	0.88889	1.00000	0.94340	1.00000	0.88889	0.94340	0.88889	0.88889	0.94118	0.11111	1
EI	25	22	27	21	6	0.96296	0.81481	1.00000	0.88679	0.96296	0.81481	0.88679	0.77778	0.77778	0.87500	0.22222	1
H	27	21	27	21	6	1.00000	0.77778	1.00000	0.88679	1.00000	0.77778	0.88679	0.77778	0.77778	0.87500	0.22222	1
A	26	24	27	23	4	0.95296	0.88889	1.00000	0.92453	0.96296	0.88889	0.92453	0.85185	0.85185	0.92000	0.14815	1
A:	22	16	23	15	8	0.81481	0.59259	0.85185	0.71598	0.95652	0.69565	0.82609	0.55556	0.55217	0.78947	0.29630	0
W	26	22	26	22	4	0.96296	0.81481	0.96296	0.88679	1.00000	0.84615	0.92157	0.81481	0.84615	0.91667	0.14815	1
O	25	24	27	22	5	0.92593	0.88889	1.00000	0.90566	0.92593	0.88889	0.90566	0.81481	0.81481	0.89796	0.18519	1
U:	25	18	26	17	9	0.92593	0.66667	0.96296	0.79245	0.96154	0.69231	0.82353	0.62963	0.65385	0.79070	0.33333	1
P	18	19	21	16	5	0.65567	0.70370	0.77778	0.67925	0.85714	0.90476	0.87805	0.59259	0.76190	0.86486	0.18519	1
T	19	18	21	16	5	0.79370	0.66667	0.77778	0.67925	0.90476	0.85714	0.87805	0.59259	0.76190	0.86486	0.18519	1
P	17	23	23	17	6	0.62963	0.85185	0.85185	0.73565	0.73913	1.00000	0.86667	0.62963	0.73913	0.85000	0.22222	1
K	14	15	18	11	7	0.51952	0.55556	0.66667	0.54717	0.77778	0.83333	0.80556	0.40741	0.61111	0.75862	0.25926	0
B	17	15	19	13	6	0.62963	0.55556	0.70370	0.60377	0.89474	0.78947	0.84211	0.48148	0.68421	0.81250	0.22222	0
D	16	19	20	15	5	0.59259	0.74074	0.74074	0.64151	0.80000	0.95000	0.87179	0.55556	0.75000	0.85714	0.18519	1
PH	17	16	20	13	7	0.62963	0.59259	0.74074	0.62264	0.85000	0.80000	0.82500	0.48148	0.65000	0.78788	0.25926	0
TH	20	16	21	15	6	0.74074	0.59259	0.77778	0.67925	0.95238	0.76190	0.85714	0.55556	0.71429	0.83333	0.22222	1
KH	17	17	19	15	4	0.62963	0.62963	0.70370	0.64151	0.89474	0.89474	0.89474	0.55556	0.78947	0.86235	0.14815	0
S	18	26	26	18	8	0.65567	0.96296	0.96296	0.81132	0.69231	1.00000	0.84314	0.66667	0.69231	0.81818	0.29630	1
(	1	13	13	1	12	0.03704	0.48148	0.48148	0.26415	0.07692	1.00000	0.53846	0.03704	0.07692	0.14286	0.44444	0
M	22	17	26	13	13	0.81481	0.62963	0.96296	0.71698	0.84615	0.65385	0.74510	0.48148	0.50000	0.66667	0.48148	1
N	24	20	24	20	4	0.89989	0.74074	0.88889	0.81132	1.00000	0.83333	0.91489	0.74074	0.83333	0.90909	0.14815	1
L	22	24	24	22	2	0.81481	0.88889	0.88889	0.84906	0.91667	1.00000	0.95745	0.81481	0.91667	0.95552	0.07407	1
R	24	26	26	24	2	0.89989	0.96296	0.96296	0.92453	0.92308	1.00000	0.96078	0.88889	0.92308	0.96000	0.07407	1
MASS DER TOTALEN ASSOZIATIVITAET = 0.7530864197530864																	
MASS DER TOTALEN SYMMETRIE = 0.6378600823045267																	
MASS DER TOTALEN INTERNEN SYMMETRIE = 0.7345971465110777																	
MASS DER TOTALEN INTERNEN SYMMETRIE * = 0.8469945192337035																	
MASS DER TOTALEN ANTISYMMETRIE = 0.2304526748971193																	
MASS DER TOTALEN REFLEXIVITAET = 0.6296296119689941																	

## III

Das eine Ziel unseres Vorgehens ist eine Klassifikation der Phoneme (vgl. oben). Damit kann zweierlei gemeint sein: Einmal kann jedes einzelne Phonem isoliert betrachtet und aufgrund gewisser Merkmale einer von mehreren Klassen zugewiesen werden, zum zweiten kann man die gegenseitige Ähnlichkeit aller Phoneme untereinander zur Grundlage einer Klassifikation machen. Letzteres ist an dieser Stelle gemeint. Die gegenseitige Phonemähnlichkeit kann selbst wiederum auf verschiedene Weise festgestellt werden. Zum Beispiel können je zwei Phoneme daraufhin überprüft werden, wieviel gemeinsame Elemente unter ihren Vorgänger- und/oder Nachfolgerphonemen sind usw. Spricht man hier von einer „distributionellen Ähnlichkeit“ (vgl. LEHFELDT 1973), so sollte man zum Unterschied davon das hier angewandte Verfahren die Bestimmung der „syntagmatischen Ähnlichkeiten“ nennen.

Bei dieser Vorgehensweise wird zunächst für alle Phoneme die Ähnlichkeit berechnet, die sie hinsichtlich der Ausprägung der von uns verwendeten 12 Merkmale, der Maße ihrer Distribution, aufweisen. Dazu benutzen wir das bekannte Maß der euklidischen Distanz:

$$D(A, B) = \left[ \sum_{i=1}^{12} (a_i - b_i)^2 \right]^{1/2}$$

Phoneme, denen eine ähnliche Ausprägung hinsichtlich vieler Merkmale zukommt, werden demgemäß eine geringere Distanz voneinander aufweisen als Phoneme, deren Distribution stark unterschiedliche Eigenschaften aufweist. So berechnet sich beispielsweise die Distanz zwischen /h/<sup>8</sup> und /e/ als

$$[(0,03704-1,0)^2 + (0,48148-0,88889)^2 + \dots + (0,44444-0,11111)^2 + (1-0)^2]^{1/2} = 2,44648,$$

während sie für /ei/ und /η/ nur 0.07407 beträgt. Da alle Phoneme über die numerische Ausprägung von 12 Merkmalen verglichen werden, liegt die theoretisch maximale Distanz zwischen zwei Phonemen bei  $\sqrt{12} = 3,4641$ , die dann auftreten würde, wenn bei dem einen Phonem jeweils genau ein Maß den Wert 1 (bzw. 0) annimmt, wenn dieses Maß bei dem anderen Phonem den Wert 1 (bzw. 0) annimmt<sup>9</sup>. Transformiert man die Werte der euklidischen Distanzen durch Division durch  $\sqrt{12}$  in das Einheitsintervall  $\langle 0; 1 \rangle$ , so erge-

<sup>8</sup> /h/ bezeichnet den Spiritus asper.

<sup>9</sup> Der Fall, daß ein Phonem bei allen Maßen den Wert 1 oder bei allen Maßen den Wert 0 aufweist, ist natürlich logisch unmöglich. Ein Phonem kann z.B. nicht gleichzeitig maximal symmetrisch und maximal asymmetrisch sein; auch muß in der Regel mindestens ein Maß  $>0$  sein, sonst hieße das ja, daß das Phonem gar nicht auftritt.

ben sich für das Attische die Werte der Tabelle III. Die Distanzen für die Phoneme /h/ und /e/ bzw. /ei/ und /η/ sind jetzt durch 0.70624 bzw. durch 0.02138 gegeben, die gleichzeitig die höchste bzw. die niedrigste der auftretenden Distanzen darstellen.

Die Werte dieser (symmetrischen) Distanzmatrix sind die Grundlage für eine *hierarchische Klassifikation* der Phoneme. Dazu bedienen wir uns des Klassifikationsalgorithmus von JOHNSON (1967) in seiner Maximumvarianante<sup>10</sup>. Durch ihn werden jeweils *die* Elemente zu einer größeren Gruppe zusammengefaßt, die die jeweils geringste Distanz voneinander aufweisen, bis schließlich alle Phoneme in einem einzigen Cluster zusammengefaßt sind. So werden z.B. die Phoneme des obigen Beispiels, /ei/ und /η/, aufgrund ihrer minimalen Distanz als erste zu einem Cluster zusammengefaßt, während /h/ und /e/ erst beim letzten Schritt vereinigt werden. Das Ergebnis dieser Klassifikationsprozedur zeigt für die Phoneme des Attischen Abbildung 1.

Legen wir auf dem Distanzniveau 0.25 einen horizontalen Schnitt durch dieses Dendrogramm, so ergeben sich vier *syntagmatische* Phonemklassen<sup>11</sup>:

- 
- A:     { h }  
 B:     { y ī ŷ ā t<sup>h</sup> b d p<sup>h</sup> k<sup>h</sup> }  
 C:     { p t g k s ū m }  
 D:     { ei η n e a ω o l i r }
- 

Dies ist eine *syntagmatische Klassifikation* der Phoneme des Attischen (vgl. damit TRUBETZKOYS „funktionelle Einteilung“ (1971, 219f.), in der ein direkter Bezug von phonologischen Merkmalen und Distributionsbeschränkungen hergestellt wird). Auf die Interpretation dieser Klassifikation kommen wir später zurück.

---

<sup>10</sup> Zu einer Einschätzung dieses Verfahrens vgl. SPÄTH (1975, 189); genauere Erläuterungen bei einer linguistischen Anwendung und ein Beispiel eines Computerausdrucks der hierarchischen Klassifikation findet man bei ALTMANN und LEHFELDT (1972).

<sup>11</sup> Vergleicht man die vier sich hier ergebenden Klassen mit denkbaren Alternativen, so zeigt es sich, daß der Schnitt bei 0,25 die am besten interpretierbare Klassifikation ergibt. Verlegt man die Schnittstelle auf 0,35, so fallen die Klassen C und D zusammen, so daß weit über die Hälfte aller Phoneme in einer Klasse versammelt wäre. Dies wäre aber nicht sinnvoll. Andersherum würde sich bei einem Schnitt auf dem Distanzniveau 0,2 oder 0,15 die Zahl der Klassen so erhöhen, daß der Sinn der Klassifikation, nämlich die Gruppierung ähnlicher Elemente, praktisch wieder aufgehoben wäre.

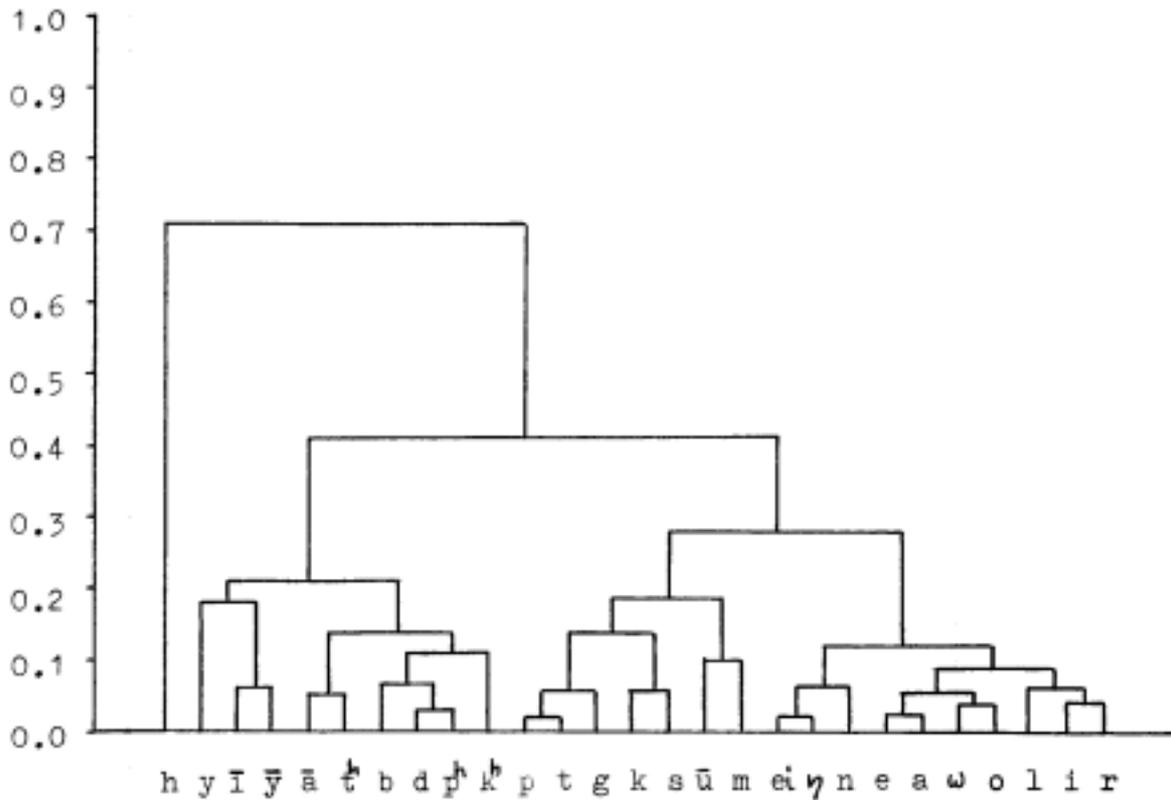


Abb. 1. Hierarchische Phonemklassifikation.

#### IV

Oben hatten wir als zweites Ziel einer syntagmatischen Phonemtypologie die „Aufdeckung des Konstruktionsmechanismus der Phonemdistribution“ genannt<sup>12</sup>. Dies läßt sich auch in einer etwas anderen Weise verdeutlichen. In unserem Falle haben wir – analog zu der Zahl der Maße – einen zwölfdimensionalen Merkmalraum, in dem jedes Phonem seine individuelle Position aufweist. Es ist klar, daß schon bei einer solchen Zahl von Dimensionen (und erst recht dann, wenn man noch weitere Maße entwickeln würde) die genannte Phonemposition hinsichtlich aller Merkmale zugleich nicht mehr anschaulich erfaßbar ist. So besteht denn „eines der Hauptziele jeder Typologie gerade darin, die vieldimensionale Ordnung, von der sie ausgeht, durch eine, wie wir sagen wollen, gleichwertige *reduzierte Ordnung* zu ersetzen, d.h. durch eine solche, die nur eine oder jedenfalls möglichst wenige Dimensionen besitzt, die aber gleichwohl dieselben Tatsachen zu beschreiben gestattet wie die kompliziertere ursprüngliche Ordnung“ (HEMPEL und OPPENHEIM 1936,

<sup>12</sup> Ein ähnliches Ziel stellt sich auch in anderen Wissenschaften. So sieht es LEWIN (1971, 45) als Aufgabe einer Psychologie, die nicht mehr aristotelischen Denkschemata nachhängt, die «inneren Kräfte» und «Wirkungsfaktoren» abzubilden.

70f.). Eine solche Reduzierung kann natürlich nur vorgenommen werden, wenn es möglich ist, Gesetze aufzustellen, die eine Verknüpfung zwischen den einzelnen Merkmalen herstellen. Dabei sind zwei Fälle zu unterscheiden: Einmal kann es sich um strenge Implikationen handeln, d.h. zwei Maße variieren in genau gleicher Weise. Zum zweiten aber kann es sich um mehr oder weniger starke Korrelationen zwischen den Merkmalen handeln, die Varianz eines Merkmals kann also nur zum Teil durch die Varianz eines anderen Merkmals erklärt werden. Liegt der letztere Fall vor, so versucht man, zu „einer *reduzierten Ordnung mit einer möglichst kleinen Zahl von Dimensionen* zu gelangen, indem man die Gesamtheit der berücksichtigten Merkmale in Bündel zusammenfaßt, derart, daß die Merkmale jedes Bündels ... parallel zueinander variieren. Die Ordnung läßt sich dann auf eine solche reduzieren, deren Dimensionszahl gleich der Anzahl der so entstehenden Bündel ist“ (HEMPEL und OPPENHEIM 1936, 75).

Der nächste Schritt unserer Untersuchung gilt also dem Ziel, herauszufinden, welche Relationen zwischen den einzelnen Maßen bestehen. Dazu berechnen wir die Korrelation zwischen allen 12 Maßen. Der Korrelationskoeffizient  $r$  wird dabei nach folgender Formel bestimmt (nach KREYSZIG 1973, 304):

$$r = \frac{\sum x_j y_j - \sum x_j \sum y_j}{\sqrt{\left[ \sum x_j^2 - \frac{1}{k} (\sum x_j)^2 \right] \left[ \sum y_j^2 - \frac{1}{k} (\sum y_j)^2 \right]}}$$

Hierbei steht  $x$  für die Werte des Maßes  $X$ ,  $y$  für die Werte des Maßes  $Y$ .  $r$  kann Werte zwischen  $-1$  und  $+1$  annehmen. Dabei bedeutet (bei normal verteilten, intervallskalierten) Variablen ein Wert von  $0$ , daß die Variabilität des Maßes  $Y$  stochastisch unabhängig ist (und umgekehrt). Ein Wert, der sich  $+1$  nähert, zeigt an, daß zwischen den beiden Maßen  $X$  und  $Y$  eine starke lineare Abhängigkeit besteht, wobei hohe Werte des einen Maßes zusammen mit hohen Werten des anderen Maßes auftreten. Ein Wert von  $-1$  zeigt im Unterschied dazu nur an, daß hohe Werte des einen Maßes zusammen mit niedrigen Werten des anderen Maßes auftreten. Der Korrelationskoeffizient ist also ein Maß für die *lineare Abhängigkeit* zwischen zwei Variablen (vgl. KRIZ 1973, 217). Die Ergebnisse der Berechnungen sind für das Attische in Tabelle IV festgehalten.



Da ja  $r=0$  anzeigt, daß zwischen den beiden betrachteten Merkmalen überhaupt keine lineare Abhängigkeit besteht, testen wir alle Werte der Tabelle IV zunächst daraufhin, ob sie signifikant von 0 verschieden sind, d.h. ob überhaupt angenommen werden kann, daß zwischen den betreffenden Merkmalen eine statistisch signifikante Korrelation besteht. Dies ist der Fall (nach WETZEL et al. 1967, 28), wenn

$$t_1 - \frac{a}{2}; k-2 \leq \sqrt{\frac{r^2}{1-r^2}}(k-2).$$

Wählen wir als Signifikanzniveau  $\alpha$  den Wert 0.05, so finden wir in der Tabelle der t-Verteilung unter  $k-2 = 25$  und  $\alpha = 0,05$  den Wert  $t_{0,975;25}$ . Setzen wir für  $t_1 - \frac{a}{2}; k-2$  der Einfachheit halber  $t$ , so läßt sich der gesuchte kritische Wert von  $r$  folgendermaßen berechnen (vgl. auch BEYER 1968, 389):

$$r_s = \sqrt{\frac{t^2}{t^2 + (k-2)}}.$$

Daraus folgt im Falle des Attischen, daß  $|r| \geq 0,38094$  sein muß, damit die betreffende Korrelation als signifikant betrachtet werden kann. Da die kritischen Werte des Korrelationskoeffizienten in statistischen Werken für  $k > 30$  meist nur unzureichend genau tabelliert sind (vgl. etwa bei OWEN 1962, 510; BEYER 1968, 390), und um ähnliche Berechnungen für andere Sprachen zu erleichtern, sind diese Werte in Tabelle V in dem für die Phonologie interessanten Bereich zusammengestellt.

Um das „Beziehungsnetz“ für die 12 distributionellen Merkmale darstellen zu können, entnehmen wir Tabelle IV zunächst die Angaben darüber, ob, und wenn ja, mit welchen anderen Merkmalen ein jedes der 12 Merkmale seine höchste positive bzw. negative Korrelation aufweist, wobei nach dem oben Gesagten klar sein dürfte, daß nur signifikante Korrelationen berücksichtigt werden. Für das Attische ergibt sich das Bild der Tabelle VI. Aus diesen Angaben läßt sich eine graphische Darstellung der im Attischen herrschenden Gesetzmäßigkeiten konstruieren, indem alle Merkmale durch Kanten miteinander verbunden werden, die durch ein „+“ oder ein „-“ als eine höchste positive oder negative Korrelation gekennzeichnet sind (Abb.2). Die Pfeilspitze zeigt dabei jeweils auf das Merkmal, *mit dem* diese Korrelation besteht. So ist z.B. „ $X \rightarrow Y$ “ gleichbedeutend mit „ $X$  hat seine höchste Korrelation mit  $Y$ “, und „ $X \leftrightarrow Y$ “ heißt soviel wie „ $X$  und  $Y$  sind durch gegenseitige höchste Korrelation verbunden“.

Tabelle V. Kritische Werte des Korrelationskoeffizienten

k	r <sub>s</sub>						
5	0,87832	24	0,40441	43	0,30086	62	0,25000
6	0,81136	25	0,39613	44	0,29731	63	0,24807
7	0,75455	26	0,38826	45	0,29400	64	0,24607
8	0,70675	27	0,38094	46	0,29066	65	0,24411
9	0,66645	28	0,37396	47	0,28755	66	0,24231
10	0,63190	29	0,36731	48	0,28454	67	0,24044
11	0,60205	30	0,36095	49	0,28161	68	0,23871
12	0,57596	31	0,35502	50	0,27876	69	0,23691
13	0,55295	32	0,34933	51	0,27600	70	0,23526
14	0,53245	33	0,34403	52	0,27330	71	0,23353
15	0,51392	34	0,33880	53	0,27068	72	0,23184
16	0,49735	35	0,33392	54	0,26813	73	0,23029
17	0,48207	36	0,32908	55	0,26565	74	0,22877
18	0,46830	37	0,32456	56	0,26323	75	0,22717
19	0,45557	38	0,32121	57	0,26087	76	0,22571
20	0,44378	39	0,31601	58	0,25857	77	0,22417
21	0,43286	40	0,31196	59	0,25644	78	0,22276
22	0,42272	41	0,30818	60	0,25424	79	0,22128
23	0,41332	42	0,30439	61	0,25210	80	0,21992

(k steht für den Umfang des Phoneminventars)

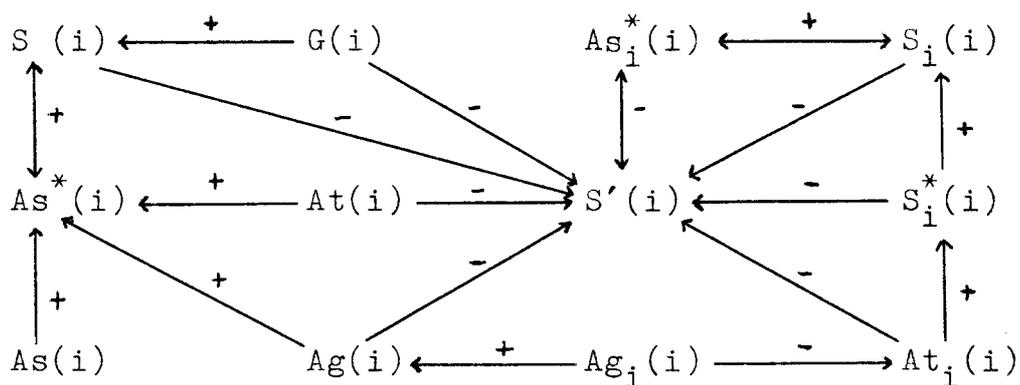


Abb. 2. «Netz der Beziehungen»

Tabelle VI. Maximalkorrelationen im Attischen

Maß	höchste Korrelation		niedrigste Korrelation	
	mit Maß	r	mit Maß	r
( 1) At (i)	( 4) As* (i)	0,9214	(11) S' (i)	-0,5022
( 2) Ag (i)	( 4) As* (i)	0,8272	(11) S' (i)	-0,4135
( 3) As (i)	( 4) As* (i)	0,9398	–	
( 4) As* (i)	( 8) S (i)	0,9677	(11) S'(i)	-0,5226
( 5) At <sub>i</sub> (i)	(10) S* <sub>i</sub> (i)	0,8909	(11) S'(i)	-0,6134
( 6) Ag <sub>i</sub> (i)	( 2) Ag (i)	0,5344	( 5) At <sub>i</sub> (i)	-0,4215
( 7) As* <sub>i</sub> (i)	( 9) S <sub>i</sub> (i)	0,9998	(11) S(i)	-0,8717
( 8) S (i)	( 4) As* (i)	0,9677	(11) S'(i)	-0,7183
( 9) S <sub>i</sub> (i)	( 7) As* <sub>i</sub> (i)	0,9998	(11) S'(i)	-0,8690
(10) S* <sub>i</sub> (i)	( 9) S <sub>i</sub> (i)	0,9786	(11) S'(i)	-0,7727
(11) S'(i)	–		( 7) As* <sub>i</sub> (i)	-0,8717
(12) G (i)	( 8) S (i)	0,6344	(11) S'(i)	-0,3965

Dieses „Netz der Beziehungen“ zwischen den distributionellen Merkmalen stellt das gesuchte allgemeine Bauprinzip dar, das die Distribution der Phoneme des Attischen bestimmt, es sind die allgemeinen Gesetzmäßigkeiten der Phonemdistribution des Attischen. Eine solche Darstellung besitzt eine „interne Prädiktivität“ (wir können auch sagen, sie sei „reduzierbar“ im Sinne von HEMPEL und OPPENHEIM 1936), d.h. kennt man den Wert eines Merkmals bei einem beliebigen Phonem, so kann man einen Bereich abschätzen, in dem mit einer im voraus festzulegenden Wahrscheinlichkeit der Wert eines anderen Merkmals liegt, sofern beide Merkmale durch eine signifikante Korrelation miteinander verbunden sind.

Um dies praktisch durchführen zu können, muß man die „Regression“ des abzuschätzenden Merkmals Y auf das Merkmal X, dessen Wert als bekannt angenommen wird, kennen. Durch die Gleichung

$$y = k + bx$$

ist die sogenannte Regressionsgerade der y-Werte bezüglich der x-Werte definiert. Hierbei bezeichnet b den „Regressionskoeffizienten“. Die genannte Gleichung geht durch einfache Umformung aus

$$y - \bar{y} = b(x - \bar{x})$$

hervor, wobei  $\bar{y}$  und  $\bar{x}$  für die Mittelwerte von Y und X stehen. Der Regressionskoeffizient berechnet sich (nach KREYSZIG 1973, 261) als:

$$b = \frac{S_{xy}}{S_x^2}$$

$$\frac{\frac{1}{k-1} \left[ \sum_{j=1}^k x_j y_j - \frac{1}{k} \left( \sum_{i=1}^k x_i \right) \left( \sum_{j=1}^k y_j \right) \right]}{\frac{1}{k-1} \left[ \sum_{j=1}^k x_j^2 - \frac{1}{k} \left( \sum_{j=1}^k x_j \right)^2 \right]}$$

Beispielsweise können wir Tabelle IV entnehmen, daß die Maße (4) As\*(i) und (8) S (i) gegenseitig durch ihre höchste positive Korrelation verbunden sind. Zunächst berechnen wir die Werte

$$\sum_{i=1}^k x_i = 17,22223, \quad \sum_{j=1}^k y_j = 20,39642,$$

$$\sum_{j=1}^k x_j^2 = 11,90535, \quad \sum_{j=1}^k x_j y_j = 13,72552$$

sowie  $\bar{x} = 0,63786$  und  $\bar{y} = 0,75542$  aus Tabelle II, wobei wir X als S (i) und Y als As\*(i), das aufgrund von X abgeschätzt werden soll, festlegen; daraus ergibt sich ein Regressionskoeffizient

$$b = \frac{\frac{1}{26} \left[ 13,72552 - \frac{1}{27} (17,22223)(20,39642) \right]}{\frac{1}{26} \left[ 11,90535 - \frac{1}{27} (17,22223)^2 \right]} = 0,77769,$$

so daß die gesuchte Regressionsgerade den Wert

$$y - 0,75542 = 0,77769 (x - 0,63786)$$

bzw.

$$y = 0,25936 + 0,77769 x$$

annimmt, oder, wenn wir für x und y die Benennungen der Maße einsetzen:

$$\text{As}^* (i) = 0,25936 + 0,77769 \text{ S} (i).$$

Für /i/ ist beispielsweise ein  $S(i)=0.85185$  ermittelt worden. Versuchen wir nun, den zugehörigen Wert des Maßes  $As^*(i)$  für /i/ abzuschätzen, so ergibt sich ein geschätztes  $\hat{y}=0,92184$ ; tatsächlich liegt der Wert bei 0.92453. Weitere Beispiele dieses Abschätzens von  $As^*(i)$  aufgrund der Kenntnis von  $S(i)$ :

i	$S(i) = x$	$As^*(i) = \bar{y}$ (geschätzt)	$As^*(i) = y$ (tatsächlich)
/i/	0,51852	0,66261	0,75472
/b/	0,40741	0,57620	0,54717
/h/	0,03704	0,28817	0,26415
/r/	0,88889	0,95064	0,92453

Solche Schätzungen (vgl. auch KRUPA und ALTMANN 1966) werden, was unmittelbar einleuchtend ist, um so genauer sein, je stärker die lineare Abhängigkeit zwischen den betrachteten zwei Merkmalen ist, d.h. je größer der Absolutbetrag des Korrelationskoeffizienten  $r$  ist. Dann nämlich wird der Standardfehler der Regressionsschätzung (vgl. KRIZ 1973, 223, 328) von  $Y$  aufgrund von  $X$ , der durch

$$s_{x/x} = s_y \sqrt{1-r^2}$$

definiert ist, geringer.  $s_y$  ist hierbei die Standardabweichung

$$\sqrt{\frac{1}{k} \sum_{i=1}^k (y_i - \bar{y})^2}$$

In unserem Beispiel für die Schätzung von  $Y=As^*(i)$  aufgrund von  $X=S(i)$  ergibt sich ein Standardfehler der Regressionsschätzung von

$$s_{x/x} = 0,19612 \sqrt{1-0,9677^2} = 0,04944.$$

Hatten wir oben für /i/ ein  $\hat{y} = 0,92184$  abgeschätzt, so zeigt sich, daß der tatsächliche Wert (0.92453) in dem Bereich

$$\hat{y} \pm s_{x/x} = < 0,87240; 0,97138 >$$

liegt.

Führt man eine solche Berechnung der Regressionsgeraden für alle die Korrelationen durch, die oben in das „Netz der Beziehungen“ eingegangen

sind, so läßt sich die gewünschte Reduktion auf möglichst wenige Merkmalbündel durchführen.

In Tabelle VI ist ja angegeben, mit welchen anderen Maßen ein jedes Maß seine höchste bzw. niedrigste signifikante Korrelation aufweist. Dies heißt aber nichts anderes, als daß das betreffende Merkmal durch diese beiden Maße am besten abgeschätzt werden kann. Für die Reduktion wird man natürlich dasjenige Maß nehmen, bei dem der Absolutbetrag von  $r$  größer ist. So werden z.B. die Maße (1) bis (3) am besten durch Maß (4) abgeschätzt, dieses wiederum durch Maß (8) usw. Man kann nun hier zwei Wege gehen. Einmal könnte man - der größeren Genauigkeit halber - fordern, jedes Maß solle nur *direkt* abgeschätzt werden, also nicht durch ein Maß, das seinerseits wieder abgeschätzt ist. In diesem Falle läßt sich eine Reduktion von 12 auf 5 Meßreihen, zu denen 7 Angaben über die Regressionsgeraden hinzutreten, erzielen:

Maß			Abschätzung durch		
(1)	At	(i)	(4)	As*	(i)
(3)	As	(i)	(10)	S* <sub>1</sub>	(i)
(5)	At <sub>1</sub>	(i)	(2)	Ag	(i)
(6)	Ag <sub>1</sub>	(i)			
(9)	S <sub>1</sub>	(i)	(7)	As* <sub>1</sub>	(i)
(11)	S'	(i)			
(12)	G	(i)	(8)	S	(i)

Gibt man dagegen größtmöglicher Reduktion der Ausgangsdaten den Vorzug (und dies ist theoretisch fruchtbarer), so versucht man, von möglichst wenigen Maßen ausgehend, alle anderen Maße zu erreichen. Jetzt läßt sich für das Attische eine Reduktion von 12 auf 2 Meßreihen, zu denen 10 Regressionsgeraden hinzutreten, erzielen (mit anderen Worten: wir betrachten den Graphen der Abbildung 2 als gerichtet): Es werden nur die Werte der beiden Maße (4) As\*(i) und (7) As\*<sub>1</sub>(i) vorgegeben, alle anderen Werte werden mittels der Regressionsgeraden abgeschätzt:

	Y	X
(1)	At (i) = -0,1928 + 1,2522	As* (i)
(2)	Ag (i) = 0,1617 + 0,7829	As+ (i)
(6)	Ag <sub>1</sub> (i) = 0,5903 + 0,3695	Ag (i)

(3)	As	(i) = 0,2405 + 0,8311	As*	(i)
(8)	S	(i) = -0,2717 + 1,2040	As*	(i)
(12)	G	(i) = -0,4289 + 1,6595	S	(i)
(9)	S <sub>1</sub>	(i) = -0,9994 + 2,0025	As* <sub>1</sub>	(i)
(10)	S* <sub>1</sub>	(i) = 0,3655 + 0,7170	S <sub>1</sub>	(i)
(5)	At <sub>1</sub>	(i) = -0,0229 + 1,0626	S* <sub>1</sub>	(i)
(11)	S'	(i) = 1,1859 - 1,1128	As* <sub>1</sub>	(i)

---

Bei einer solchen auch indirekten Schätzung einzelner Maße pflanzt sich natürlich die Ungenauigkeit des Schätzens nicht nur fort, sondern vergrößert sich bei jedem Schritt, wie leicht einzusehen ist. Im Falle des Attischen werden drei Maße durch ein Maß abgeschätzt, das selbst einmal abgeschätzt wurde; ein Maß durch ein anderes, das durch zweifaches Abschätzen ermittelt wurde. Nach HEMPEL und OPPENHEIM (1936, 104f.) ist erst jetzt die Stufe einer *Theorie* über die Ausgangsdaten erreicht: Erst wenn empirische Gesetzmäßigkeiten (hier in Form von Regressionsgeraden) angegeben sind, kann man nicht mehr von einem beziehungslosen Nebeneinander der Maße sprechen. Die Frage nach der Stabilität solcher Regelmäßigkeiten kann natürlich erst in einer sprachtypologischen Arbeit beantwortet werden.

## V

Bisher haben wir zweierlei erreicht: Die Phoneme des Attischen wurden auf der Basis ihrer distributionellen Relationen, die in den 12 Maßen zum Ausdruck kommen, hierarchisch klassifiziert. Es ergaben sich vier syntagmatische Phonemklassen (A–D, vgl. oben). Zweitens wurde ein „Netz von Beziehungen“ ermittelt (vgl. Abb.2), das den Distributionstendenzen der einzelnen Phoneme unterliegt. Die Ergebnisse beider Schritte gilt es nun miteinander zu verknüpfen.

Indem wir zu den Ausgangsdaten zurückgehen, d.h. zu den Werten der 12 Distributionsmaße, soll jede der genannten vier Klassen, A–D, durch einen „diagnostischen Schlüssel“ von allen anderen unterschieden und bezüglich ihrer Struktur gekennzeichnet werden. „Zum Ausgangspunkt des diagnostischen Schlüssels wählen wir das Merkmal, bei dem die Summe aller Korrelationskoeffizienten im absoluten Wert am größten ist“ (ALTMANN und LEHFELDT 1973, 46). Dies ist im Attischen für S (i) der Fall, da

$$\sum_{j=1}^{11} |r_{ij}| = 8,4239.$$

Dieses Merkmal ist mit  $As^*(i)$  durch gegenseitige höchste Korrelation verbunden. Weiter weist  $G(i)$  zu  $S(i)$  seine höchste positive Korrelation auf. Es fällt zunächst auf, daß /h/, das einzige Phonem der Klasse A, den kleinsten aller auftretenden  $As^*$ -Werte hat: 0.26415. Das Kriterium  $As^* < 0,3$  reicht also aus, um die Klasse A von den anderen Klassen (B,C,D) eindeutig zu unterscheiden. Weiter stellt man fest, daß alle Phoneme der Klasse D durch  $S(i) > 0,7$  ausgezeichnet sind; aber auch /y/, das nicht in diese Klasse gehört, weist einen  $S(i)$ -Wert  $> 0,7$  auf. Auf den ersten Blick scheint dieses Kriterium nicht zur Klassenunterscheidung tauglich. Wir wenden uns deshalb dem Merkmal (12) zu, dem Maß der Reflexivität. Wie Tabelle II zeigt, sind alle Phoneme der Klassen C und D reflexiv, während alle Phoneme der Klasse B irreflexiv sind.

Man braucht nun diese drei Kriterien nur in eine richtige Reihenfolge zu bringen, um eine eindeutige Unterscheidung aller Klassen zu erreichen. Das geht aus folgender Übersicht hervor:

Kriterium	Klasse			
	A	B	C	D
1. $As^*(i)$	$< 0,3$	$> 0,3$	$> 0,3$	$> 0,3$
2. $G(i)$	(-)	0	1	1
3. $S(i)$	(-)	(-)	$< 0,7$	$> 0,7$

Dieselben Prinzipien können auch in Form eines einfachen Algorithmus, der die eindeutige richtige Zuordnung eines jeden Phonems zu einer der vier Klassen erlaubt, gekleidet werden:

- 
1. Wenn  $As^*(i) < 0,3$ , → A, sonst → 2
  2. Wenn  $G(i) = 0$ , → B, sonst → 3
  3. Wenn  $S(i) < 0,7$ , → C, sonst → D
- 

Diese drei Merkmale können also *a posteriori* insofern als „wesentlich“ für eine distributionelle Klassifikation der Phoneme des Attischen bezeichnet werden, als sie es gestatten, diese Klassifikation einfach zu interpretieren und jedes Phonem als Angehörigen einer Klasse, mithin als „Typ“, zu kennzeich-

nen. Will man die einzelnen Klassen benennen, so ist Klasse A die Klasse der Phoneme mit geringer interner Paarbildung, Klasse B die der irreflexiven Phoneme; Klasse C umfaßt Phoneme mit einem höchstens mittleren Grad an Symmetrie, Klasse D solche mit hoher Symmetrie.

## VI

Mit den 12 Maßen haben wir quantitative Feststellungen über die Ausprägung distributioneller Relationen für die Phoneme des Attischen getroffen. Die Berechnung dieser Werte ist aber kein Ziel in sich; sie dient klassifikatorischen, „typologischen“ Zwecken, indem, wie oben durchgeführt, alle Phoneme über die numerischen Werte verglichen und (hierarchisch) klassifiziert werden, woraus eine „Typologie“ der betreffenden Phoneme gewonnen werden kann. Es kann aber auch der Wunsch bestehen, beurteilen zu wollen, ob sich zwei Phoneme hinsichtlich der Ausprägung eines Merkmals signifikant unterscheiden, ob ein bestimmtes Phonem einen hohen, einen mittleren oder einen niedrigen Grad der gemessenen Eigenschaften aufweist. Dazu müssen statistische Modelle herangezogen werden (sofern geeignete Modelle schon existieren). Mit ihrer Hilfe ist es möglich, „quantitative Feststellungen in kategoriale umzuwandeln“ (ALTMANN und LEHMANN 1973, 16). Bisher existieren solche Modelle für vier Eigenschaften von Phonemen:

1. Das Modell 1 gestattet über die Größe  $|A_i|$  eine Klassifikation des Phonems  $i$  als *attraktiv*, *semiattraktiv* oder *nichtattraktiv*;

2. Das Modell 2 gestattet über die Größe  $|iB|$  eine Klassifikation des Phonems  $i$  als *aggressiv*, *semiaggressiv* oder *nichtaggressiv*.

Modelle 1 und 2 (in Wirklichkeit zwei Varianten *eines* Modells) sind bei ALTMANN und LEHFELDT (1972, 16 ff.) beschrieben.

3. Aufgrund des Zusammenhangs der Größen  $|A_i|$ ,  $|B_i|$  und  $|A_i \cap B_i|$  kann in Modell 3 entschieden werden, ob ein Phonem  $i$  *symmetrisch*, *semisymmetrisch* oder *asymmetrisch* ist.

4. Aufgrund der Differenz  $|A_i| - |B_i|$  gestattet das Modell 4 eine Klassifikation des Phonems  $i$  als *quantitativ-symmetrisch* oder als *quantitativ-asymmetrisch*.

Modelle 3 und 4 sind bei ALTMANN (1973b) beschrieben, Modell 3 verbessert in ALTMANN und LEHFELDT (1980, Kap.5).

Angewandt auf das Attische ergibt sich folgendes Bild:

*Modell 1*

Attraktiv (T):	{ i y e ei η a ā ω o ū t <sup>h</sup> m n l r }
Semiattraktiv (ST)	{ ī ȳ p t k b d g p <sup>h</sup> k <sup>h</sup> s }
Nichtattraktiv (NT)	{ h }

*Modell 2*

Aggressiv (G)	{ i ī y ȳ e ei η a ω o k s n l r }
Semiaggressiv (SG)	{ ā ū p t b d g p <sup>h</sup> t <sup>h</sup> k <sup>h</sup> h m }
Nichtaggressiv (NG)	{ Ø }

*Modell 3*

Symmetrisch (S)	{ p t k b d g p <sup>h</sup> t <sup>h</sup> k <sup>h</sup> n l }
Semisymmetrisch (SS)	{ i ī y y e ei η a ā ω o ū s h m r }
Asymmetrisch (AS)	{ Ø }

*Modell 4*

Quantitativ-symmetrisch (QS)	{alle Phoneme bis auf 3}
Quantitativ-asymmetrisch (QAS)	{ī s h}

Oben haben wir die Ergebnisse der hierarchischen Klassifikation mit den numerischen Ausprägungen der Distributionsmaße verknüpft und eine eindeutige Klassifizierung erreichen können. Wie sieht es nun aus, wenn wir die quaternäre Klassifikation der Phoneme, die ja ihrerseits auch wieder auf der Ausprägung einiger numerischer Maße beruht, heranziehen? Dazu führen wir noch einmal alle vier Klassen mit ihren Elementen auf und schreiben hinter jedes Phonem die Symbole, die seiner oben gegebenen Einstufung entsprechen:

A	/h/	NT, SG, SS, QAS	C	/p/	ST, SG, S, QS
				/t/	ST, SG, S, QS
B	/y/	T, G, SS, QS		/g/	ST, SG, S, QS
	/i/	ST, G, SS, QAS		/k/	ST, G, S, QS
	/ȳ/	ST, G, SS, QS		/s/	ST, G, SS, QAS
	/ā/	T, SG, SS, QS		/ū/	T, SG, SS, QS
	/t <sup>h</sup> /	T, SG, S, QS		/m/	T, Sg, SS, QS

/b/	ST, SG, S, QS	D	/ei/	T, G, SS, QS
/d/	ST, SG, S, QS		/ɲ/	T, G, S, QS
/p <sup>h</sup> /	ST, SG, S, QS		/n/	T, G, S, QS
/k <sup>h</sup> /	ST, SG, S, QS		/e/	T, G, SS, QS
			/a/	T, G, SS, QS
			/ω/	T, G, SS, QS
			/o/	T, G, SS, QS
			/l/	T, G, SS, QS
			/i/	T, G, SS, QS
			/r/	T, G, SS, QS

Daraus läßt sich zum Beispiel ablesen, daß in der Klasse A das einzige nichtattraktive Phonem zu finden ist, daß in der Klasse D alle Phoneme sowohl attraktiv, aggressiv als auch quantitativ-symmetrisch sind. Diese beiden Klassen lassen sich somit leicht von den übrigen abgrenzen, für die auf dieser Stufe der Untersuchung keine Kriterien gefunden werden können, die eine Zuordnung der Phoneme zu der einen oder anderen Klasse stets eindeutig erlauben würden. Die Klassen B und C können beide nur durch  $(\neg NT \wedge \neg NG)$  charakterisiert werden. Man darf aber vermuten, daß eine eindeutige Unterscheidung der Klassen B und C möglich sein dürfte, wenn weitere statistische Modelle zur Bewertung distributioneller Einzelmaße entwickelt sind. Im Augenblick können wir deshalb nur folgende Vorschrift zur Zuordnung der Phoneme geben:

1. Wenn  $i \in NT$ , dann  $\rightarrow A$ , sonst  $\rightarrow 2$
2. Wenn  $i \in (T \wedge G)$ , dann  $\rightarrow D$ , sonst  $\rightarrow B \vee C$

Falsch eingeordnet wird /y/, das der Klasse D zugefügt wird.

Die Phoneme des Attischen sind weiterhin dadurch gekennzeichnet, daß u.a. folgende Implikationen gelten, die gleichzeitig die Redundanz der genauen Spezifizierung deutlich machen:

- (1)  $(T \wedge G) \rightarrow QS$
- (2)  $(ST \wedge SG) \rightarrow SS$
- (3)  $QAS \rightarrow SS$
- (4)  $(TS \wedge G) \rightarrow QS$
- .
- .
- .

## VII

Ohne dies auch praktisch durchzuführen, wollen wir aufzeigen, wie die oben dargelegten Eigenschaften des Phonemsystems in seinem syntagmatischen Aspekt für den Sprachvergleich benutzt werden können. Dazu sind mehrere Möglichkeiten denkbar:

1. In der Tabelle IV sind insgesamt

$$\binom{12}{2} = 1(12) / 2 = 66$$

verschiedene Werte des Korrelationskoeffizienten  $r$  aufgeführt. Diese kann man als einen Vektor verstehen und in beliebiger, aber für alle Sprachen gleicher, Reihenfolge anordnen, so daß jede Sprache also durch 66 Werte gekennzeichnet ist. Will man nur die Stärke der Korrelationen für den Sprachvergleich berücksichtigen, so setzt man die Absolutbeträge des Korrelationskoeffizienten in diesen Vektor ein. Soll auch die Richtung des Zusammenhangs zum Ausdruck kommen, so transformiert man die Werte des Korrelationskoeffizienten aus dem Intervall  $\langle -1; +1 \rangle$  in das Einheitsintervall  $\langle 0; 1 \rangle$ , indem man

$$r' = (r+1)/2$$

berechnet und einsetzt. Dann berechnet man die euklidischen Distanzen und klassifiziert die Sprachen (hierarchisch), so wie oben für die Phoneme des Attischen beschrieben.

2. Oben haben wir gezeigt, wie die Signifikanz der Werte des Korrelationskoeffizienten getestet werden kann. Man kann demnach eine einfache Distinktion treffen, indem man zwischen signifikanten (1) und nichtsignifikanten (0) Korrelationen unterscheidet. Anstelle der numerischen Ausprägungen setzt man die auf statistischer Grundlage getroffenen kategorischen Entscheidungen über die Korrelationen in den Vektor ein, der damit die Gestalt eines Binärvektors annimmt, d.h. seine Elemente nehmen nur die Werte 0 oder 1 an. Für zwei Sprachen bestimmt man ihre Ähnlichkeit bzw. ihre Distanz, indem man auszählt, wie oft die 66 Elemente der beiden Vektoren übereinstimmen (11 oder 00) bzw. nicht übereinstimmen (10 oder 01). Auch auf die Werte der hier erhaltenen Ähnlichkeits- bzw. Distanzmatrix ist dann ein klassenbildender Algorithmus anwendbar.

3. Schließlich kann man die Sprachen, wie schon angedeutet, hinsichtlich ihres Beziehungsnetzes, wie es in Abbildung 2 dargestellt ist, vergleichen. Dazu stellt man die höchste signifikante positive und negative Korrelation eines jeden der 12 Merkmale fest (vgl. Tab. IV) und zählt aus, wie oft zwei

Sprachen hierin übereinstimmen. Auch dies führt zu einer (hierarchischen) Klassifikation der Sprachen.

Diese Andeutungen mögen genügen, um aufzuzeigen, welche Aufgaben zu lösen sind, wenn eine hinreichend große Zahl von Sprachen unter dem hier betrachteten Aspekt bearbeitet ist, d.h. wenn für sie zuverlässige Angaben über die Phonemdistribution vorliegen.

### *Zusammenfassung*

Aufbauend auf den Weiterentwicklungen des Distributionskalküls von HARRY und PAPER wird eine Methode vorgeschlagen, um den Konstruktionsmechanismus der Phonemdistribution einer Sprache aufzudecken. Die Phoneme werden hierarchisch klassifiziert und der gegenseitige Zusammenhang ihrer Distributionsmerkmale untersucht. Dazu wird die Korrelations- und Regressionstechnik verwendet. Auf diese Weise wird eine „Typologie“ der Phoneme im Sinne von HEMPEL und OPPENHEIM ermöglicht.

### Résumé

#### **Typologie syntagmatique des phonèmes**

En se basant sur les développements du calcul distributionnel dûs à HARRY et PAPER, l'auteur propose une méthode pour découvrir le mécanisme constructif de la distribution des phonèmes dans une langue. Il procède à une classification hiérarchique des phonèmes et analyse leurs traits distributionnels en appliquant les techniques de corrélation et de régression. On obtient ainsi une „typologie“ des phonèmes dans le sens de HEMPEL et OPPENHEIM.

## Literatur

- AFENDREAS, E.A.: Can one measure a Sprachbund? A calculus of phonetic distribution for language contact. In: *Folia linguistica* 4, 1970, 93–103.
- ALTMANN, G.: Probabilistische Klassifikation von Konsonantenverbindungen des Indonesischen. In: *Zeitschrift der deutschen morgenländischen Gesellschaft*, 123, 1973a, 98–116.
- ALTMANN, G.: Asymmetrie der Distribution linguistischer Einheiten. In: *Phonetica* 28, 1973b, 86–96.
- ALTMANN, G. und LEHFELDT, W.: Typologie der phonologischen Distributionsprofile. In: *Beiträge zur Linguistik und Informationsverarbeitung*, No. 22, 1972, 8–32.
- ALTMANN, G. und LEHFELDT, W.: *Allgemeine Sprachtypologie. Prinzipien und Meßverfahren*. München: Fink 1973.
- ALTMANN, G. und LEHFELDT, W.: Einführung in die quantitative Phonologie (*Quantitative Linguistics*, vol 7.). Bochum: Studienverlag Dr. N. Brockmeyer 1980.
- BEYER, W.H. (ed.): *Handbook of tables for probability and statistics*. 2nd. ed. Cleveland: Chemical Rubber 1968.
- GRUCZA, F.: *Sprachliche Diakrise im Bereich der Ausdrucksebene des Deutschen*. Beiträge zur Allgemeinen Sprachtheorie. Poznan: PAN 1970.
- HARARY, F. and PAPER, H.H.: Towards a general calculus of phonetic distribution. In: *Language* 33, 1957, 143–169.
- HEMEPL, C.G. und OPPENHEIM, P.: *Der Typusbegriff im Lichte der neuen Logik*. Wissenschaftstheoretische Untersuchungen zur Konstitutionsforschung und Psychologie. Leiden: Sijthoff 1936.
- JOHNSON, S.C.: Hierarchical clustering schemes. In: *Psychometria* 32, 1967, 241–254.
- JUILLAND, A.: *Outline of a general theory of structural relations*. The Hague: Mouton 1961.
- KREYSZIG, E.: *Statistische Methoden und ihre Anwendungen*. Göttingen: Vandenhoeck & Ruprecht 1973.
- KRIZ, J.: *Statistik in den Sozialwissenschaften. Einführung in die kritische Diskussion*. Reinbek–Hamburg: Rowohlt 1973.
- KRUPA, V. und ALTMANN, G.: Relations between typological indices. In: *Linguistics* 24, 1966, 29–37.
- LEHFELDT, W.: Distributionelle Phonemähnlichkeit. In: *Phonetica* 27, 1973, 82–99.
- LEWIN, K.: *Der Übergang von der aristotelischen zur galileischen Denkweise in Biologie und Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft 1971.
- OWEN, D.B.: *Handbook of statistic tables*. Reading: Addison–Wesley 1962.
- SPÄTH, H.: *Cluster-Analysen–Algorithmen zur Objektklassifizierung und Datenreduktion*. München: Oldenbourg 1975.
- TRNKA, B.: General laws of phonemic combinations. In: *TCLP* 6, 1936, 57–62.
- TRUBETZKOY, N.S.: Charakter und Methode der systematischen phonologischen Darstellung einer gegebenen Sprache (1932). In: HEIKE, G. *Phonetik und Phonologie*. München 1974, 40–44.
- TRUBETZKOY, N.S.: *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht 1971.
- WETZEL, W., JÖHNK, M.–D., NAEVE, P.: *Statistische Tabellen*. Berlin: de Gruyter 1967.