



117 QUANTITATIVE TYPOLOGIE

1. Die quantitativen morphologischen Indizes von J.H. Greenberg
2. Zur Weiterentwicklung des Greenbergschen Ansatzes
3. Taxonomische morphologische Sprachklassifikation und Suche nach Zusammenhängen
4. Quantifizierung der Eigenschaften grammatischer Subsysteme

1. Die quantitativen morphologischen Indizes von J.H. Greenberg

Die quantitative morphologische Typologie, wie sie sich in den vergangenen 30-40 Jahren entwickelt hat, kann als Nachfolgerin und ‘Erbin’ der morphologischen Sprachtypologie des 19. Jahrhunderts betrachtet werden. Das zeigt sich besonders deutlich bei J. H. Greenberg, dessen Arbeit ‘A Quantitative Approach to the Morphological Typology of Language’ (1960, zuerst 1954) den eigentlichen Beginn der Entwicklung der quantitativen morphologischen Typologie bezeichnet. Es sollte freilich beachtet werden, daß auch andere Autoren Vorschläge zur quantifizierenden Erfassung morphologischer Eigenschaften erarbeitet haben, so etwa Thorndike (1943), Hamp (1958), Kelemen (1970) und Robins (1965). Ein ganzes System von quantifizierbaren Eigenschaften des Morphems ist von Skalička (1935) und Slavičková (1968) entwickelt worden.

Greenberg betrachtet die morphologische Wortstruktur als “something of fundamental importance to the over-all characterization of a language” (1960, 180) und stellt sich die Aufgabe, “to re-examine the nineteenth-century approach to the problem” (das.; vgl. auch 1960, 179). Das Verbindungsglied zwischen Greenbergs Ansatz und der älteren morphologischen Typologie ist die Sprachtypologie Edward Sapirs, an die Greenberg unmittelbar anknüpft: “My [morphological] typology is a pragmatic [dazu s.u.] approach using Sapir’s attribute space with some modification” (Greenberg 1957, 72; vgl. auch 1960, 187 f.). Ungeachtet aller wichtigen Modifikationen an den Verfahren der ‘klassischen’ morphologischen Typologie, die Sapir vornimmt, läßt auch er sich leiten von einer “logic of metonymy in which a few parts (i.e. a low number of morphosyntactic features) represent the whole” (Voegelin et al. 1960, 198).

Greenberg (1960) schlägt zehn Indizes vor, mit deren Hilfe verschiedene morphologische Eigenschaften natürlicher Sprachen in Form quantitativer Begriffe erfaßt werden können. Damit formalisiert er aber nicht nur den Ansatz von Sapir, sondern erweitert ihn um die Dimension einer pragmatischen Frequenzuntersuchung, da sich die Indizes auf die Vorkommenshäufigkeit von morphologischen Einheiten in Texten bestimmter Länge beziehen.

Die zehn Indizes von Greenberg entsprechen insgesamt fünf morphologischen Parametern, von denen drei von Sapir übernommen sind. Zwei der Sapirschen Parameter sind formaler Natur, und so bezieht sich Greenberg ohne weiteres auf sie, nämlich die Parameter “Synthesis” und “Technique”, die allerdings bei Sapir nur eine nachgeordnete Rolle spielen.

Auf Sapirs Parameter “Synthesis” bezieht sich der bei Greenberg an erster Stelle stehende “index of synthesis” (M/W), der die “gross complexity of the word” mißt, d.h. “the degree of complexity exhibited on the basis of the number of subordinate meaningful

elements it contains” (Greenberg 1960, 182). Dieser Index bezieht die Anzahl der Morpheme eines Textes auf die Anzahl der Wörter. Sein kleinstmöglicher Wert – für den Fall, daß $W = M$ – ist 1; je größer er ist, desto ‘synthetischer’ ist die betreffende Sprache.

Auf Sapirs Parameter “Technique” bezieht sich Greenbergs “index of agglutination” (A/J), der die Anzahl der agglutinierenden morphematischen Junktoren zur Anzahl der morphematischen Junktoren überhaupt ins Verhältnis setzt. Hier ist 1 der höchste Indexwert. Je weiter sich der für eine Sprache ermittelte Indexwert von 1 entfernt, um so weniger agglutinierend, d.h., um so ‘fusionierender’ ist diese Sprache. Der Unterschied zwischen agglutinierenden und fusionierenden Sprachen ist der “between languages in which the subordinate elements are added to the root elements mechanically, i.e., without either of the elements being modified (...), and those involving a process of fusion by which the constituent elements become difficult to recognize and separate” (Greenberg 1960, 183).

Der Hauptgesichtspunkt von Sapirs Schema ist der des Ausdrucks bestimmter Begriffe: radical concepts (I), derivational concepts (II), concrete relational concepts (III), pure relational concepts (IV) (vgl. hierzu Sapir 1949, Ch. V, v.a. S. 101) (s. Art. 115). Greenberg kritisiert Sapir: “..., Sapir seems to be talking about concepts, but in reality his test is a formal, not a semantic one” (Greenberg 1960, 184). Daher modifiziert Greenberg diesen Sapirschen Hauptparameter, indem er sich zu einem formalen Vorgehen entschließt (vgl. 1960, 184): “..., the present treatment is based on the possibility of the exhaustive division of morphemes into three classes, root, derivational, and inflectional” (1960, 186).

Die Quantifizierung von Sapirs formalisiertem Hauptparameter führt Greenberg zur Aufstellung von drei Indizes: der “compounding index” (R/W) bezieht die Anzahl der Wurzeln in einem Text auf die Anzahl der Wörter; der “derivational index” (D/W) setzt die Anzahl der Ableitungsmorpheme eines Textes zur Anzahl der Wörter in Beziehung, und der “gross inflectional index” (I/M) bezieht die Anzahl der Flexionsmorpheme in einem Text auf die Anzahl der Wörter.

Den drei Sapirschen Parametern fügt Greenberg zwei weitere hinzu. Der erste davon “is the order of subordinate elements in relation to the root”, “a topic discussed by Sapir as important for the morphological structure of a language but not included in his final formulation” (Greenberg 1960, 186). Die beiden diesem Parameter entsprechenden Indizes – der “prefixal index” P/W und der “suffixal index” S/W – beziehen die Anzahl der Präfixe bzw. der Suffixe in einem Text auf die Anzahl der Wörter.

“The final parameter has to do with the devices employed for relating words to each other. It therefore brings in syntactical as well as morphological considerations. There are three devices that language may use, inflectional morphemes without concord, significant order, or concord (agreement)” (Greenberg 1960, 187). So ergeben sich drei weitere Indizes: der “isolational index” O/N bezieht die Anzahl der lediglich durch die Stellung signalisierten syntaktischen Konnexionen (“order”) auf die Anzahl der syntaktischen Konnexionen (“nexus”). Der “pure inflexional index” Pi/N bezieht die Anzahl der durch Flexion (aber nicht durch Kongruenz) signalisierten syntaktischen Konnexionen auf die Anzahl aller Konnexionen. Der “concordial index” Co/N schließlich setzt die Anzahl der durch Kongruenz (“concord”) signalisierten Konnexionen in einem Text zur Anzahl aller Konnexionen in ein Verhältnis.

In ihrer Gesamtheit bilden die zehn Indizes einen “multidimensional attribute space” (Greenberg 1957, 76), in dem jede untersuchte Sprache lokalisiert werden kann. Die Auswahl der gerade hier gewählten Attribute steht unter der Annahme, “that they represent closely related aspects of the same general class of phenomena which may reasonably be expected to reveal significant mutual connections” (das.). Greenberg wendet sein Verfahren an, indem er die Werte seiner zehn Indizes für acht Sprachen – Sanskrit, Angelsächsisch (Altenglisch), Persisch, Englisch, Jakutisch, Swahili, Annamitisch, Eskimo – anhand von Textstichproben berechnet, die jeweils 100 Wörter lang sind. Es handelt sich also um eine pragmatische Frequenzuntersuchung. Bei einer solchen Untersuchung wird die Häufigkeit einer Einheit aus Texten ermittelt, während sich die systemische Häufigkeit auf ein Inventar von Einheiten bezieht.

2. Zur Weiterentwicklung des Greenbergschen Ansatzes

Greenbergs Aufsatz von 1960 bezeichnet den eigentlichen Beginn der Entwicklung der quantitativen morphologischen Typologie. Diese Entwicklung bezog ihren Impetus weitgehend aus der Auseinandersetzung mit Greenbergs Ansatz. Sie ist in verschiedenen Richtungen verlaufen, von denen einige im folgenden behandelt werden sollen.

Einige Linguisten haben Greenbergs Indizes um weitere Indizes ergänzt. So schlägt Cowgill (1976, 123) einen “index of infixes per word” (Inf/W) vor. Auch andere Typologen – etwa Ju.Ja. Glazov (1965) und M. Mejlach (1973) – verwenden in ihren Untersuchungen den “Infigierungsindex”. V.B. Kasevič und S.E. Jachontov (1982, 37) betrachten die Nichtunterscheidung von autosemantischen und synsemantischen Wörtern (s. Art. 27) als wichtigen Mangel von Greenbergs Typologie und schlagen deshalb einen “Index der Analytizität” (Aux/W) vor, der die Anzahl der synsemantischen Wörter in einem Text auf die Anzahl aller Wörter bezieht (vgl. auch Sil’nickij et al. 1986, 106). Silnitsky (1993, 159) kommt allerdings zu dem Ergebnis, daß dieser zusätzliche Index keinerlei signifikante klassifikatorische Funktion erfülle und deshalb nicht weiter zu berücksichtigen sei. Von Pierce (1962) stammt ein Index C/W, der die Anzahl der Konsonanten in einem Text auf die Anzahl der Wörter bezieht, also die durchschnittliche Anzahl von Konsonanten pro Wort mißt (s. Art. 44).

Kasevič und Jachontov modifizieren gleichfalls einige Bestimmungen der durch die Indizes erfaßten sprachlichen Einheiten, beispielsweise indem sie den Begriff der Agglutination weiter fassen als Greenberg und die Worteinheit anders bestimmen (vgl. 1982, 3). Dies führt natürlich zu einer Veränderung der Indexwerte. M. Mejlach (1965, 165) berücksichtigt in seinem System nur die ersten vier von Greenbergs fünf Parametern. Den fünften vernachlässigt er deswegen, weil die zu ihm gehörenden Indizes “im Bereich der reinen Morphologie” fakultativ, in dem der Syntax hingegen unzureichend seien.

A.L. Kroeber (1960) machte die von Greenberg für acht Sprachen berechneten Indexwerte zur Grundlage einer “rank order” zwischen diesen Sprachen, ohne Greenbergs Vorgehen selbst in irgendeiner Hinsicht zu modifizieren. Jede Sprache bekommt bezüglich jeden Indexes einen Rang zwischen 1 und 8 zugeordnet, der sich jeweils aus dem Vergleich der Werte für einen Index ergibt. Beispielsweise erhält das Eskimo im Hinblick auf den “index of synthesis” den Rang 1, weil es für diesen Index unter allen betrachteten Sprachen den höchsten Wert aufweist. Die Gruppierungen, die sich aus der sich so ergebenden Rangtabelle

ablesen lassen, versucht Kroeber zu den Klassen der morphologischen Sprachtypologie des 19. Jh. in Beziehung zu setzen und – für die indogermanischen Sprachen – entwicklungs-geschichtlich zu interpretieren.

Eine weitere, vieldiskutierte Frage ist die nach der Stichprobengröße. Greenberg selbst hatte sich mit Texten im Umfang von jeweils 100 Wörtern begnügt. Es fragt sich, ob aufgrund derartig kleiner Stichproben signifikante Daten gewonnen werden können, d.h. Aussagen, die jeweils für die betreffende Sprache gelten. Auf dieses Problem macht bereits Cowgill aufmerksam: “Another weakness of the material used here is the small size of the samples, 100 words of text from each language. ... One of the urgent desiderata of current typology is to determine the minimum size of sample needed to be reasonable sure of having an accurate picture of a language” (1976, 116 f., zuerst 1963; vgl. auch van Son, Peters 1989, 20). Mit diesem Problem hat sich J. E. Pierce beschäftigt. Pierce will folgende Frage lösen: “how large a sample is necessary for computation of indices and how should this sample be secured, i.e., a series of small samples drawn from different texts or one large one, and if the former, how small” (1966, 44). Zu diesem Zweck werden drei von Greenbergs Indices – M/W, R/W, S/W – für drei verschiedene Stichproben aus türkischen Texten berechnet: (a) für 14 Stichproben von je 100 Wörtern, wobei diese Stichproben vier Stilarten und sechs Autoren repräsentieren; (b) in zehn Stichproben von je 200 Wörtern aus denselben Quellen wie unter (a), jedoch so, daß eine Überschneidung vermieden wurde; (c) in zehn Stichproben von je 100 Wörtern aus einem zusammenhängenden 1000 Wörter umfassenden Text.

Pierce testet die Differenzen zwischen den Indexmittelwerten seiner drei Stichprobenmengen auf ihre Signifikanz. Die Differenz zwischen den Mittelwerten der ersten und der dritten Stichprobenmenge erweist sich als hochsignifikant, d.h., mehrere 100-Wörter-Stichproben aus mehreren verschiedenen Texten sind für das Türkische als Ganzes repräsentativer als ebensolange Stichproben aus einem zusammenhängenden Text. Die Differenz zwischen den Mittelwerten der 100- und der 200-Wörter-Stichproben ist nichtsignifikant. “This would seem to indicate that Greenberg’s 100-word-sample size, arbitrary though it may have been, was large enough, i.e. nothing of significance was gained by increasing the sample size to 200-words, ...” (1966, 47). Davon unberührt bleibt natürlich der Umstand, daß statistische Signifikanz nur dann erreicht werden kann, wenn man nicht nur, wie Greenberg das getan hat, eine einzige Stichprobe untersucht.

Letzteres Problem ist von A. V. Stepanov aufgegriffen worden, der in einer 1995 erschienenen Arbeit konstatiert, die Anzahl der bisher untersuchten Textstichproben sei nicht ausreichend gewesen, “to make conclusions about real statistical distribution of index values for any index” (Stepanov 1995, 143). Um diesem Mangel abzuwehren, schlägt Stepanov vor, der Untersuchung der Greenbergschen Indizes für jede hierfür ausgewählte Sprache ein großes Textkorpus zugrundezulegen, das alle möglichen Textformen und -arten berücksichtigen sollte: “...the corpus will include hundreds of texts from all existing genres, sources, historical periods etc. as one large sample” (144).

In bezug auf ein solches Korpus soll nun nicht einfach für einen gegebenen Index dessen Durchschnittswert berechnet werden, vielmehr soll die Verteilungsfunktion eines jeden Indexes durch die Parameter Mittelwert, Standardabweichung und relativer Fehler beschrieben werden. Dieses Vorgehen ermögliche es, zu analysieren, in welcher Weise zufällige und systematische Faktoren das von dem jeweiligen Index repräsentierte morphologische

Merkmal in einer bestimmten Sprache beeinflussen. Jedesmal nämlich, wenn die empirische Verteilung nicht einer Normalverteilung entspreche, trage für diese “Störung” ein systematischer Faktor die Verantwortung, der anschließend ermittelt werden müsse. Die exemplarischen Analysen von Texten aus einigen semitischen Sprachen vermögen allerdings nicht zu überzeugen, da Stepanov seine empirisch gewonnenen Verteilungen nicht mit Hilfe der einschlägigen statistischen Verfahren testet, sondern nur impressionistisch auswertet. Seine an und für sich anregenden Vorschläge bedürfen daher weiterer Präzisierung.

Eine wichtige, leider aber noch immer nicht hinreichend beachtete Modifikation des Greenbergschen Ansatzes stammt von V. Krupa (1965). Dieser kritisiert, daß Greenberg nicht die mathematischen Eigenschaften seiner Indizes untersucht habe. Diese seien inkommensurabel, da ihre Werte in drei unterschiedliche Intervalle fielen:

(a) Die Werte der Indizes A/J , O/N , P_i/N und Co/N fallen in das geschlossene Intervall $< 0; 1 >$, weil $A \leq J$, $O \leq N$, $P_i \leq N$ und $Co \leq N$.

(b) Die Werte der Indizes M/W und R/W fallen in das halbgeschlossene Intervall $< 1; \infty >$ – wobei 0 natürlich nur eine theoretische Größe ist –, weil $M \geq W$ und $R \geq W$.

(c) Die Werte der Indizes I/W , D/W , P/W , Inf/W und S/W bewegen sich – theoretisch – zwischen 0 und 1, d.h. innerhalb des halbgeschlossenen Intervalls $< 0; 1 >$, weil $I \geq W$, $D \geq W$, $P \geq W$, $Inf \geq W$, $S \geq W$ oder $I < W$, $D < W$, $P < W$, $Inf < W$ und $S < W$.

So ist also beispielsweise der höchste Wert von A/J gleichzeitig der niedrigste Wert von M/W ; d.h., die Werte beider Indizes können nicht unmittelbar miteinander verglichen werden. Um diese Komplizierung zu vermeiden, schlägt Krupa vor, die Indizes so zu konstruieren, daß sämtliche Werte in das Einheitsintervall $< 0; 1 >$ fallen. Diese Bedingung ist für die unter (a) genannten Indizes bereits erfüllt. Die übrigen Indizes sind wie folgt zu ersetzen: $M/W \rightarrow W/M$; $R/W \rightarrow W/R$, $D/W \rightarrow D/M$, $I/W \rightarrow I/M$; $P/W \rightarrow P/M$; $Inf/W \rightarrow Inf/M$; $S/W \rightarrow S/M$.

Der neue Index W/M mißt nicht, wie M/W , direkt den Grad des Synthetismus, sondern den des Analytismus. Dieser ist am höchsten, wenn $M = W$. Der Synthetismus berechnet sich komplementär als $1 - W/M$. W/R ist ein direktes Maß der Einwurzeligkeit von Wörtern. Der Komplementärindex $1 - W/R$ ist ein Maß für den Grad der Komposition. Die Ersetzung von D/W durch D/M bedeutet, daß die Anzahl der Derivationsmorpheme nicht mehr auf die Anzahl der Wörter, sondern die der Morpheme im Text bezogen wird; entsprechend bei den übrigen Indizes.

Obwohl die Werte aller Indizes jetzt in das Einheitsintervall fallen, gliedern sich die Indizes in drei Gruppen, und zwar unter dem Gesichtspunkt der Fähigkeit bzw. der Unfähigkeit, die Extremwerte der Skala tatsächlich zu erreichen. Beispielsweise kann der Index A/J sowohl den Wert 0 wie den Wert 1 annehmen, da es in einem Text gar keine agglutinierenden morphematischen Junktoren zu geben braucht bzw. die Anzahl der agglutinierenden morphematischen Junktoren gleich der Anzahl der morphematischen Junktoren überhaupt sein kann. Der Index D/M hingegen kann nur den Extremwert 0 erreichen – etwa in analytischen Sprachen ohne jegliche Derivationsmorpheme –, nicht aber den Extremwert 1, da es keine Texte geben kann, die nur aus Derivationsmorphemen bestünden (weitere Einzelheiten vgl. bei Krupa 1965, 34). Eine Übersicht über die Indizes zeigt Tab. 1. Die für 20 Sprachen berechneten Werte der Greenberg–Krupaschen Indizes sind in Tab. 2 zusammengestellt (nach Altmann & Lehfelddt 1973, 40).

Mit den von Krupa modifizierten Indizes arbeiten etwa Altmann & Lehfeldt (1973, 39 ff.), van Son & Peters (1989) u.a. Aber noch in dem von V. B. Kasevič und S. E. Jachontov 1982 herausgegebenen Band, der ganz der Anwendung von Greenbergs Ansatz gewidmet ist, werden Greenbergs Indizes in ihrer ursprünglichen Form verwendet und die Werte der Indizes ohne Zuhilfenahme mathematischer Verfahren interpretiert. Sil'nickij et al. (1986) verwenden den "index of synthesis" und den "compounding index" in der von Krupa vorgeschlagenen Form und ersetzen im Nenner der Indizes O/N, Pi/N und Co/N N durch W (so auch bereits Kasevič & Jachontov 1982; vgl. auch Silnitsky 1993; in dieser Arbeit werden Krupas Modifikationen der Greenbergschen Indizes ignoriert). Letztere Modifikation berührt freilich nicht nur die Form der drei Indizes, da die Anzahl der syntaktischen Konnexionen nicht gleich der Anzahl der in einem Text vorhandenen Wörter zu sein braucht.

Eine für die Interpretation der Werte von Greenbergs Indizes äußerst wichtige Frage ist die nach deren gegenseitigem Verhältnis: Sind die von den Indizes erfaßten Eigenschaften voneinander logisch unabhängig, oder gibt es zumindest zwischen einigen von ihnen bestimmte logische Implikationen? Greenberg selbst weist auf dieses Problem hin, wenn er sagt: "In general, the lower the first, or synthetic, index, the fewer the morph junctures which occur and the less the importance of this second index [d.h. des "index of agglutination" A/J] in characterizing the language" (1960, 185 f.). Vgl. auch Cowgill (1976, 130): "If a language has a high morpheme-per-word index, it will automatically have high indices in at least two, and probably all four, of the indices 4 through 7" (vgl. auch Kasevič & Jachontov 1982, 307).

Krupa, der selber auf die Notwendigkeit hinweist, die "mutual relationships linking the individual indices" (1965, 35) zu studieren, vernachlässigt diesen Gesichtspunkt, wenn er das Einheitsintervall in vier Bereiche einteilt und den Wert, den ein Index annimmt, ohne Bezugnahme auf die Werte anderer Indizes beurteilt; vgl.: "Thus, for example, a value of 0.88 for W/M is highly analytical, since $0.76 < 0.88 < 1$, while a value of 0.27 for the same index is moderately synthetic, since $0.26 < 0.27 < 0.50$."

Instances where the resulting value of an index is equal to 1 or 0 may be called purely A or purely $\neg A$." Was wir bei Krupa finden, ist also eine absolute, nicht-relationale Klassifikation, die nicht die Relationen zwischen den verschiedenen Indizes berücksichtigt.

	<i>Index</i>	<i>Bereich</i>	<i>Intervall</i>	<i>Autor</i>
1	W/M	Analytismus	(0; 1 >	G./mod. Krupa
2	A/J	Agglutination	< 0; 1 >	Greenberg
3	W/R	Komposition	(0; 1 >	G./mod. Krupa
4	D/M	Derivation	< 0; 1)	G./mod. Krupa
5	I/M	Flexion	< 0; 1)	G./mod. Krupa
6	P/M	Präfixation	< 0; 1)	G./mod. Krupa
7	S/M	Suffixation	< 0; 1)	G./mod. Krupa
8	O/N	Morphosyntax	< 0; 1 >	Greenberg
9	Pi/N	Morphosyntax	< 0; 1 >	Greenberg
10	Co/N	Morphosyntax	< 0; 1 >	Greenberg

W	=	Anzahl der Wörter
M	=	Anzahl der Morpheme
A	=	Anzahl der agglutinierenden Konstruktionen
J	=	Anzahl der morphematischen Junktoren
R	=	Anzahl der Wurzelmorpheme
D	=	Anzahl der Derivationsmorpheme
I	=	Anzahl der Flexionsmorpheme
P	=	Anzahl der Präfixe
S	=	Anzahl der Suffixe
O	=	Anzahl der Orders
N	=	Anzahl der Nexus
Pi	=	Anzahl der durch Flexion signalisierten syntaktischen Konnexionen
Co	=	Anzahl der Kongruenzen

Tab. 117.1.: Morphologische Indizes der Sprachtypologie

	W/M	A/J	W/R	D/M	I/M	P/M	S/M	O/N	Pi/N	Co/N
Sanskrit	0.39	0.09	0.88	0.24	0.32	0.06	0.46	0.16	0.46	0.38
Bengali	0.53	0.46	0.92	0.15	0.28	0.01	0.42	0.57	0.29	0.14
Altpersisch	0.41	0.20	0.98	0.17	0.41	0.08	0.50	0.23	0.39	0.38
Neupersisch	0.66	0.34	0.97	0.07	0.26	0.01	0.32	0.52	0.29	0.19
Altgriechisch (Homer)	0.48	0.10	0.99	0.10	0.41	0.03	0.48	0.48	0.27	0.26
Neugriechisch	0.55	0.40	0.98	0.07	0.37	0.02	0.42	0.53	0.21	0.26
Altenglisch	0.47	0.11	1.00	0.09	0.42	0.03	0.48	0.15	0.47	0.38
Neuenglisch	0.60	0.30	1.00	0.09	0.32	0.02	0.38	0.75	0.14	0.11
Jakutisch	0.46	0.51	0.98	0.16	0.38	0.00	0.53	0.29	0.59	0.12
Suaheli	0.39	0.67	1.00	0.03	0.31	0.45	0.16	0.40	0.19	0.41
Vietnamesisch	0.94	–	0.93	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Eskimo	0.27	0.03	1.00	0.34	0.47	0.00	0.73	0.02	0.46	0.38
Türkisch (Schrift)	0.43	0.60	1.00	0.11	0.43	0.00	0.54	0.43	0.67	0.20
Türkisch (Mündl.)	0.57	0.67	0.96	0.06	0.38	0.00	0.44	0.69	0.16	0.03
Gotisch	0.43	0.19	0.97	0.13	0.42	0.04	0.52	0.37	0.34	0.29
Altkirchenslaw.	0.44	0.20	1.00	0.15	0.41	0.05	0.51	0.41	0.33	0.26
Hethitisch	0.51	0.42	1.00	0.12	0.36	0.01	0.48	0.35	0.32	0.33
Rigveda	0.39	0.08	0.91	0.19	0.38	0.07	0.48	0.26	0.48	0.27
Asoka	0.40	0.26	0.82	0.17	0.34	0.03	0.49	0.40	0.18	0.42
Griechisch (N.T.)	0.41	0.12	0.97	0.11	0.47	0.07	0.51	0.34	0.32	0.34

Tab. 117.2: Typologische Indizes für 20 Sprachen

Den Schritt hin zu einer relationalen Klassifikation unternehmen V. Krupa und G. Altman in einer 1966 erschienenen Arbeit. ‘Relationale’ Klassifikation bedeutet, daß eine Sprache nicht aufgrund des absoluten Wertes eines Indexes hinsichtlich der durch diesen Index quantifizierten Eigenschaft eingestuft wird, sondern unter Berücksichtigung (mindestens) des Wertes eines anderen Indexes, von dem gezeigt worden ist, daß er mit dem fraglichen Index logisch verknüpft ist. In dieser Weise korrelieren beispielsweise die Indizes W/M und D/M, und zwar negativ: Wenn W/M den Wert 1 annimmt, dann muß D/M den Wert 0 aufweisen. Betrachten wir die Werte von D/M isoliert, dann sind, wenn wir dem Vorschlag von Krupa (1965) folgen, fast alle von Greenberg untersuchten Sprachen “highly non-derivative”, weil diese Werte fast immer in das Intervall $< 0.00; 0.25 >$ fallen. Berücksichtigt man jedoch die Korrelation von D/M zu W/M – je analytischer eine Sprache, um so geringer ihr ‘Derivations-

grad' –, sind ganz ähnliche absolute Werte von D/M u.U. verschieden zu bewerten. Im Swahili ist $D/M = 0.03$, im Vietnamesischen $D/M = 0.00$. Der Wert 0.03 bedeutet für das Swahili, daß es eine “extremely non-derivative language” ist, weil es nämlich eine “highly synthetic language” mit $W/M = 0.39$ ist. Ein solcher Wert würde eigentlich einen höheren ‘Derivationsgrad’ erwarten lassen. Im Vietnamesischen ist $W/M = 0.94$, d.h., diese Sprache ist “highly analytical”, so daß “the zero degree of derivation is a normal (average) event” (Krupa & Altmann 1966, 36).

3. Taxonomische morphologische Sprachklassifikation und Suche nach Zusammenhängen

In einem weiteren Schritt wurden von G. Altmann und W. Lehfeldt (1973) die Indizes von Greenberg–Krupa als Grundlage für eine – exemplarisch zu verstehende – morphologische Sprachklassifikation benutzt sowie, im Zusammenhang damit, für ein Verfahren zur systematischen Untersuchung der Beziehungen zwischen den von Greenberg–Krupa gemessenen morphologischen Merkmalen.

Bei der Konstruktion der morphologischen Sprachklassifikation orientieren sich Altmann und Lehfeldt an den Verfahren der v.a. in der Biologie entwickelten Numerischen Taxonomie. Danach werden die Indizes von Greenberg–Krupa als Elemente eines Profilvektors betrachtet, die die Koordinaten einer Sprache in einem zehndimensionalen Raum darstellen. So können wir in Tabelle 2 jede Zeile als einen solchen Vektor interpretieren. Dieser Ausgangspunkt ermöglicht eine Klassifikation der so charakterisierten Sprachen, wobei man zuerst die Ähnlichkeit bzw. die ihr komplementäre ‘Entfernung’ der Sprachen hinsichtlich der Indexwerte berechnet und dann auf der Grundlage der Ähnlichkeits- bzw. Entfernungskoeffizienten die Klassen aufstellt (vgl. hierzu auch Haarmann 1976, 46 ff.).

Zur Messung der Entfernung je zweier Sprachen bedienen sich Altmann & Lehfeldt (1973, 39 ff.) der euklidischen Distanz, die die Differenz zwischen den Werten aller Indizes berücksichtigt. Dieses Maß darf allerdings, strenggenommen, zur Entfernungsmessung nur dann verwendet werden, wenn die berücksichtigten Koordinaten logisch voneinander unabhängig sind. Diese Voraussetzung trifft im Falle der Greenberg–Krupa-Indizes nicht allgemein zu, wie oben gezeigt wurde (vgl. dazu u.a. auch Mejlach 1973, 165 f.). Daher müssen alle Ergebnisse, die aus der Interpretation der auf der Grundlage der euklidischen Distanz gemessenen Entfernungen gewonnen werden, als nur vorläufig betrachtet werden.

Das von Altmann & Lehfeldt verwendete Distanzmaß lautet:

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

Hierbei bezeichnen A und B die jeweils betrachteten Sprachen, a und b die Werte des Indexes i in diesen beiden Sprachen. Beispielsweise errechnet sich die Distanz zwischen dem Swaheli und dem Eskimo als

$$d(\text{Swaheli}, \text{Eskimo}) = [(0.39 - 0.27)^2 + (0.67 - 0.03)^2 + \dots + (0.41 - 0.38)^2]^{1/2} = 1.14$$

Berechnet man die Distanzen für alle Sprachpaare aus Tab. 1, so erhält man eine Distanzmatrix, die die Grundlage für den eigentlichen Klassifikationsvorgang abgibt (vgl. Alt-

mann & Lehfelddt 1973, 42–43; eines solchen Verfahrens bedienen sich auch Sil'nickij et al. 1986 für die Klassifizierung von 31 Sprachen, wobei sie von den Korrelationskoeffizienten zwischen diesen Sprachen ausgehen. Vgl. dort S. 124 f. einen Vergleich mit der Klassifikation von Altmann & Lehfelddt und mit der traditionellen morphologischen Sprachklassifikation).

Für die Klassifikation ist wesentlich, daß die Anzahl der Klassen nicht von vornherein festgelegt wird. Altmann & Lehfelddt bedienen sich eines sog. Cluster-Algorithmus. Dieser fügt, beginnend mit der kleinsten Entfernung, alle Sprachen in einem Iterationsverfahren zu immer größeren Gruppen zusammen. Das Ergebnis läßt sich in Form eines Baumdiagrammes darstellen (vgl. Abbildung 1, nach Altmann & Lehfelddt 1973, 41).

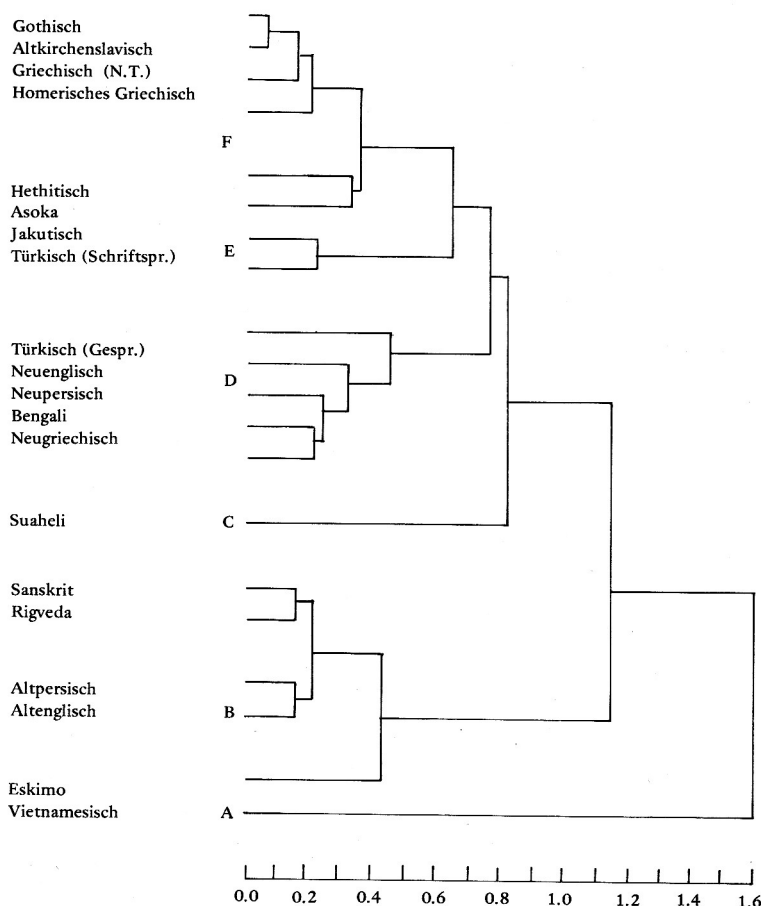


Abb. 117.1: Hierarchische Klassifikation von 20 Sprachen

Erst nach Abschluß des Klassifikationsvorganges wird die Anzahl der Klassen bestimmt. Benötigt man viele Klassen mit jeweils wenigen, dafür aber relativ ähnlichen Sprachen, so wird man den Schnitt etwa bei einer Distanz von 0.5 anlegen (6 Klassen). Interessiert dagegen eine globale Einteilung, so wäre ein Schnitt bei 1.0 denkbar, der nur noch drei Klassen ergibt (eine Klasse umfaßt jeweils alle Sprachen, die an einem der 'Zweige' hängen).

Eine weitere morphologische Sprachklassifikation, die auf morphologischen Indizes beruht, ist von G. Silnitsky (1993) vorgelegt worden. Obwohl diesem Autor Krupas Modifi-

kation von Greenbergs Ansatz bekannt ist, operiert er mit den ursprünglichen Indizes von Greenberg, wobei allerdings im Nenner der Indizes O/N, Pi/N und Co/N N durch W ersetzt wird (dazu s.o.). Hinzukommt der oben bereits erwähnte "Index der Analytizität". Silnitsky klassifiziert 31 Sprachen, indem er die Werte der Korrelationskoeffizienten auswertet, die für jedes Sprachenpaar auf der Grundlage der genannten insgesamt elf Indizes berechnet worden sind. Die Klassifikation, zu der der Autor auf diese Weise gelangt, weist "a relatively high degree of correspondence with the traditional genealogical and typological language classes" (1993, 159) auf (vgl. S. 156 einen Vergleich mit der Klassifikation von Altmann & Lehfeldt 1973).

Mit einer bloßen Sprachklassifizierung, auf wie vielen Merkmalen sie auch immer beruhen mag, kann sich die Sprachtypologie nicht zufriedengeben. Ihr geht es darüber hinaus 1) um die Aufdeckung der empirischen Zusammenhänge zwischen den bei der Klassifizierung berücksichtigten Merkmalen sowie 2) um die Frage, welche Indizes ausschlaggebend für die Klassenbildung gewesen sind. Auf diese Weise wird die interne Struktur der Klassen erhellt.

Was das erste Problem betrifft, so berechnen Altmann & Lehfeldt (1973, 44–46) die Korrelationskoeffizienten zwischen den gemessenen Merkmalen und stellen die höchsten positiven wie negativen Korrelationen eines jeden Indexes in Form eines Graphen dar (vgl. Abbildung 2):

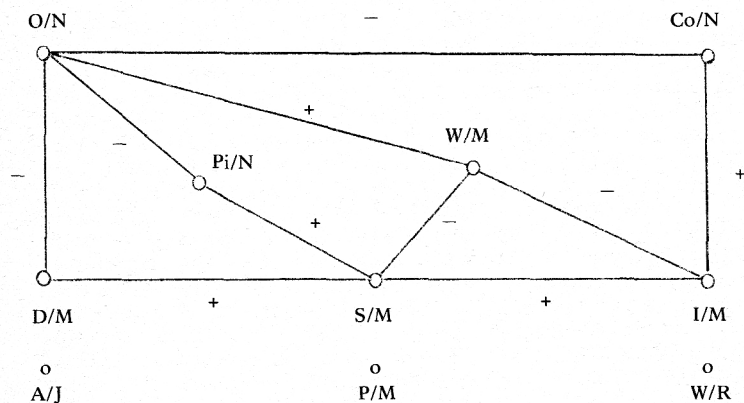


Abb. 177.2: Graph der Merkmalszusammenhänge

Hier sieht man, daß drei Merkmale vollkommen unabhängig voneinander sind: A/J, P/M und W/R. Bei den übrigen sind zwei Gruppen zu erkennen: O/N – W/M sowie D/M – S/M – I/M – Pi/N – Co/N. Die Merkmale dieser Gruppen sind untereinander positiv, die Gruppen negativ korreliert. Die positive Korrelation etwa zwischen S/M und Pi/N zeigt, daß Sprachen, deren Wörter viele Suffixe enthalten, dazu tendieren, syntaktische Konnexionen durch Flexion zu signalisieren (vgl. Sil'nickij et al. 1986, 109; diese Autoren haben für 31 Sprachen verschiedener Familien- und Klassenzugehörigkeit die Korrelationskoeffizienten zwischen den – z.T. modifizierten (vgl. Abschnitt 2) – Greenbergschen Indizes ausgewertet und sind dabei zu ähnlichen Ergebnissen wie Altmann & Lehfeldt gelangt; vgl. auch Silnitsky 1993).

Die Kenntnis der Regression eines Merkmals auf ein anderes erlaubt es, das Intervall abzuschätzen, in dem der Wert des zweiten Merkmals liegen wird, wenn der des ersten

gegeben ist. Die konkreten Werte für die Indizes aus Tab. 1 haben Krupa & Altmann (1966) berechnet.

Zum zweiten Problem: Um herauszufinden, welche Merkmale in den einzelnen Klassen ähnlich ausgeprägt sind, konstruiert man einen sog. ‘klassifikatorischen Schlüssel’ (vgl. Altmann & Lehfeldt 1973, 46–48), indem man ‘von oben her’ in das Baumdiagramm hineingeht und bei jeder Verzweigung in der Tabelle der Ausgangswerte (Tab. 1) nach einem oder mehreren Merkmalen sucht, das/die die jeweilige Zerlegung ergibt/ergeben. In unserem Falle etwa unterscheidet das Kriterium $W/M > 0.90$ die erste Klasse, die nur das Vietnamesische umfaßt, eindeutig von allen übrigen Klassen. Gesucht werden, mit anderen Worten, hinreichende, nicht notwendige Unterscheidungsmerkmale. Eine vollständige Analyse ergibt:

Klasse

A	$W/M > 0.90$				
B	$W/M \leq 0.90,$	$O/N < 0.27$			
C	$W/M \leq 0.90,$	$O/N \geq 0.27,$	$P/M > 0.10$		
D	$W/M \leq 0.90,$	$O/N \geq 0.27,$	$P/M \leq 0.10,$	$S/M < 0.45,$	
E	$W/M \leq 0.90,$	$O/N \geq 0.27,$	$P/M \leq 0.10,$	$S/M \geq 0.45,$	$Pi/N > 0.50$
F	$W/M \leq 0.90,$	$O/N \geq 0.27,$	$P/M \leq 0.10,$	$S/M \geq 0.45,$	$Pi/N \leq 0.50$

Schließlich kann man die numerischen Aussagen noch in ‘Normalsprache’ rückübersetzen, um zu anschaulichen, einprägsamen Charakterisierungen der Klassen zu gelangen. Sie könnten etwa folgendermaßen lauten:

A	hochgradig analytisch, maximal isolierend
B	gering isolierend, relativ freie Wortfolge
C	stark präfigierend, stark agglutinierend
D	schwach präfigierend, schwach suffigierend
E	stärker suffigierend, viele flexivische Kongruenzen, stark agglutinierend
F	wenige rein flexivische Kongruenzen

Es ist bezeichnend, daß sich die quantitative morphologische Typologie bisher weitgehend darauf beschränkt hat, einen einzigen Ansatz, den von Greenberg, weiterzuentwickeln. Bei allen Fortschritten, die hierbei erzielt worden sind, ist zweierlei nicht zu übersehen: 1. Zahlreiche Eigenschaften der Greenberg–Krupaschen Indizes, ihre Reliabilität und Validität, sind bisher gar nicht oder nur ungenügend untersucht worden. Auch sind uns die statistischen Eigenschaften der Indizes, etwa ihre Stichprobenverteilung, kaum bekannt. 2. Bei den Greenberg–Krupaschen Indizes geht es um die Quantifizierung und Messung einer kleinen Menge relativ einfach zu bestimmender morphologischer Merkmale. Es sollte danach gestrebt werden, komplexe morphologische Eigenschaften zu quantifizieren und zu messen. Einen solchen Ansatz wollen wir im folgenden abschließend vorstellen.

4. Quantifizierung der Eigenschaften grammatischer Subsysteme

Alle Indizes, die auf Greenberg zurückgehen, beruhen auf pragmatischen Stichproben, d.h. auf Zählungen an Texten, und erfassen Relationen, in denen die Elemente dieser Stichprobe zueinander stehen. In mehrfacher Hinsicht anders verfährt der Ansatz, der von Lehfeldt (1977; 1983; 1985; vgl. auch Raster 1980) entwickelt, ausgearbeitet und von ihm und seinen Mitarbeitern im Rahmen eines Projektes zur “Vergleichenden Morphologie der slawischen Sprachen” erprobt wurde. Dieser sog. “analytisch-synthetisch-funktionale” Ansatz beschreibt zunächst einen Ausschnitt der Formenbildung einer Sprache nach einer bestimmten Konzeption, um anschließend zu einer ganzheitlichen Charakterisierung des betreffenden Teilsystems zu kommen.

Der deskriptive Ansatz zerlegt die Formenbildung zunächst in drei Komponenten – **Endungsreihen** (“Flexionsparadigmen”), **morphonologisches Alternationsverhalten** (“morphologische Paradigmen”) und – soweit in der fraglichen Sprache vorhanden – **Akzentuierungsmuster** (“Akzentparadigmen”) – und ermittelt die Zahl der sprachlichen Mittel auf jeder dieser drei Ebenen. So findet man im Präsens des Ukrainischen beispielsweise drei Reihen verbaler Endungen, fünf Möglichkeiten der Stammalternationen (keine Veränderung, Verkürzung um den vokalischen Auslaut, Alternation des Auslautkonsonanten usw.) und drei Betonungsmuster (Stambbetonung, Endungsbetonung, Wechselbetonung). Dies ist der analytische Teil der Beschreibung. Jedes konkrete Verb realisiert (mindestens) eine spezifische Kombination dieser zur Verfügung stehenden Teile, wobei nach Zusammenhängen zwischen den verschiedenen Klassen von Ausdrucksmitteln sowie verschiedenen Stammklassen und den jeweils realisierten Kombinationen der grammatischen Mittel gesucht wird. Die Beschreibung dieser Zusammenhänge ist der synthetische Teil des Ansatzes. Im abschließenden funktionalen Teil wird der Informationsgehalt untersucht, der jedem konkreten Formenbildungsmittel zukommt, also seine funktionale Belastung, und zwar bezogen auf das jeweilige Teilsystem der Formenbildung.

Die Quantifizierung der Eigenschaften des auf diese Weise beschriebenen Systems erfaßt mehrere Aspekte. Unmittelbar gegeben ist die Zahl der Formenbildungsmittel für die genannten drei Ebenen, im ukrainischen Präsens etwa 3 (Endungsreihen), 5 (Alternationsmuster) und 3 (Betonungsmuster). Theoretisch kann es in dieser Sprache demnach $3 \times 5 \times 3 = 45$ Möglichkeiten der Präsensbildung geben; tatsächlich kommen jedoch nur 22 verschiedene Kombinationen vor, wie die systemische Untersuchung aller Verben gezeigt hat. Das Minimum verschiedener Kombinationen liegt natürlich bei 5. Diese drei Werte, d.h. die minimale, die tatsächliche und die maximale Zahl verschiedener Kombinationen, setzt das sog. “Maß der Verbundenheit” zueinander in Beziehung, um dieses Subsystem der Grammatik ganzheitlich zu charakterisieren. Es hat folgende Form:

$$M_V = \frac{K_{\max} - K_{\text{real}}}{K_{\max} - K_{\min}}$$

Die Werte dieses Maßes liegen im Einheitsintervall $<0;1>$ und sind somit für vergleichende Zwecke nutzbar. Das *tertium comparationis* liegt dabei auf der Inhaltsseite (“Präsensinhaltsparadigma”). Niedrige Werte des Indexes bedeuten eine Beschränkung auf wenige Formenbildungsmuster und signalisieren zugleich einen hohen Grad an implikatorischen

Beziehungen zwischen den Elementen, hohe Werte zeigen eine freiere Kombinierbarkeit aller Mittel untereinander, so daß sich aus der Kenntnis eines Mittels (beispielsweise der Endung) nur schlecht Voraussagen über die anderen Mittel (etwa die Betonung) machen lassen.

Für das Ukrainische ergibt sich konkret der folgende Wert: $(45-22)/(45-5) = 0,575$. Damit liegt diese Sprache innerhalb der bisher bearbeiteten Sprachen im unteren Bereich, wie die Vergleichszahlen zeigen:

Russisch 0,75	Ukrainisch 0,575	Weißrussisch 0,60	Slovakisch 0,67	Slovenisch 0,69	Serbokroatisch 0,72
Obersorbisch 0,78	Tschechisch 0,93	Polnisch 0,92	Bulgarisch 1,00	Makedonisch 1,00	

Diese numerischen Daten lassen sich nun wieder sprachwissenschaftlich interpretieren. Die Tatsache, daß die drei ostslavischen Sprachen am unteren Ende der Werteskala eine komplette Gruppe bilden, läßt sich auf den Einfluß der Betonungsmuster zurückführen, wie sie in allen diesen Sprachen vorhanden sind. Das heißt, anders formuliert, daß in den ostslavischen Sprachen die Betonungsmuster durch ihr "Hinzutreten" zu den Endungsreihen und den morphologischen Alternationsmustern die Verbundenheit des Systems der Präsensformenbildung schwächen. Im Serbokroatischen und im Slovenischen hingegen, die gleichfalls verschiedene Betonungsmuster kennen, wird durch diese die Verbundenheit nicht nur nicht verringert, sondern sogar verstärkt (vgl. hierzu Lehfeldt 1998).

Einen weiteren Aspekt erfaßt das "Maß der Prädiktivität", das wir hier nicht im einzelnen behandeln wollen, da es bisher nur für wenige Sprachen berechnet worden ist. Es setzt die jeweils zur Auswahl stehenden Ausdrucksmittel in Beziehung zu den Wortstämmen und fragt danach, wie gut sich die Verwendung dieser Mittel aufgrund der Stammklassenzugehörigkeit des Lexems vorhersagen läßt. Selbst wenn sich beispielsweise alle einzelnen Komponenten miteinander kombinieren, das Maß der Verbundenheit also seinen Maximalwert annimmt, kann es ja durchaus sein, daß eine jede konkrete Stammklasse immer nur eine ganz bestimmte Kombination zuläßt, so daß in dieser Hinsicht dennoch vollkommene Eindeutigkeit gegeben ist. Dieser Index gibt damit implizit auch Auskunft über die Frage, wie leicht oder schwer das Formenbildungssystem dieser Sprache zu erlernen bzw. zu beherrschen ist: Bei einem hohen Grad an "lexikalischer Spezialisierung" beispielsweise muß zu jedem Lexem seine Formenbildung separat gelernt werden.

Es versteht sich, daß auch dieser Ansatz nur ein erster Schritt zu einer Quantifizierung der morphologischen Struktur einer Sprache sein kann, zumal auch hier noch nicht alle theoretischen Probleme geklärt sind (so wird der Index nachweislich von Entscheidungen auf der Phonemebene unmittelbar beeinflusst). Immerhin ist damit jedoch ein Weg vorgezeichnet, auf dem eine quantitative morphologische Sprachtypologie weiter ausgebaut werden kann.

5. Zitierte Literatur

- Altmann, Gabriel (1978), "Zur Verwendung der Quotiente in der Textanalyse". *Glottometrika* 1, 91–106
- Altmann, Gabriel & Grotjahn, Rüdiger (1988), "Linguistische Meßverfahren". In: Ammon, Ulrich & Dittmar, Nobert & Mattheier, Klaus J. (Hrsg.), *Soziolinguistik: Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft, Bd. II*. Berlin, New York: De Gruyter, 1026–1039
- Altmann, Gabriel & Lehfeldt, Werner (1973), *Allgemeine Sprachtypologie*. München: Fink (UTB 250)
- Altmann, Gabriel & Lehfeldt, Werner (1980), *Einführung in die quantitative Phonologie*. Bochum: Brockmeyer
- Contreras, Heles (1963), "Una clasificación morfo-sintáctica de las lenguas románicas". *Romance Philology* 16, 261–268.
- Cowgill, Warren (1976), "A Search for Universals in Indo-European Diachronic Morphology". In: Greenberg, Joseph H. (ed.), *Universals of Language*, Cambridge MA: The MIT Press [1963], 114–141
- Glazov, Jurij Ja. (1965), "K probleme tipologičeskogo schodstva dravidijskich i tjurskich jazykov" [On the problem of typological similarity between Dravidian and Turkish languages]. *Lingvističeskaja tipologija i vostočnye jazyki*, Moskva, 205–212
- Greenberg, Joseph H. (1957), "The Nature and Use of Linguistic Typologies". *International Journal of American Linguistics* 23, 68–77
- Greenberg, Joseph H. (1960), "A Quantitative Approach to the Morphological Typology of Language". *International Journal of American Linguistics* 26, 178–194
- Haarmann, Harald (1976), *Grundzüge der Sprachtypologie*. Stuttgart usw.: Kohlhammer (Urban-Taschenbücher 242)
- Hamp, Eric P. (1958), "The Calculation of Parameters of Morphological Complexity". *Proceedings of the Eighth International Congress of Linguists*, Oslo, 134–142
- Ineichen, Gustav (1979), *Allgemeine Sprachtypologie: Ansätze und Methoden*. Darmstadt: Wissenschaftliche Buchgesellschaft
- Kasevič, Vadim B. & Jachontov, Ssergej E. (eds.) (1982), *Kvantitativnaja tipologija jazykov Azii i Afriki* [A quantitative typology of Asian and African languages]. Leningrad
- Kelemen, Joseph (1970), "Sprachtypologie und Sprachstatistik". Dezső, L. & Hajdú, P. (Hrsg.), *Theoretical Problems of Typology and the Northern Eurasian Languages*. Amsterdam, 53–63
- Kroeber, Alfred L. (1960), "On Typological Indices I: Ranking of Languages". *International Journal of American Linguistics* 26, 171–177
- Krupa, Viktor (1965), "On Quantification of Typology". *Linguistics* 12, 31–36
- Krupa, Viktor & Altmann, Gabriel (1966), "Relations between Typological Indices". *Linguistics* 24, 29–37
- Lehfeldt, Werner (1985), *Sprjaženie ukrainskogo glagola* [Ukrainian verbal inflection]. München: Sagner
- Lehfeldt, Werner (1998), "Das System der Präsensformenbildung im Polnischen im Vergleich mit den anderen slavischen Sprachen". In: Rothe, Hans & Thiergen, Peter (Hrsg.), *Polen unter Nachbarn, Polonistische und komparatistische Beiträge zu Literatur und Sprache. XII Internationaler Slavistenkongreß in Krakau 1998*. Köln usw.: Böhlau, 29–53.
- Mejlach, M.: (1973), Indeksy morfoložičeskoj tipologii [Indices for a morphological typology]. *Problemy grammatičeskogo modelirovanija*, Moskva, 155–170
- Pierce, Joe E. (1962), "Possible Electronic Computation of Typological Indices for Linguistic Structures". *International Journal of American Linguistics* 28, 215–226
- Pierce, Joe E. (1966), "Sampling and Typological Indices in Language". *Linguistics* 24, 43–50
- Raster, Peter (1980), "Paradigmatische Strukturen von Verbalsystemen indogermanischer Sprachen". *Kwartalnik Neofilologiczny* 27, 259–286

- Robins, Robert H. (1965), "Some Typological Observations on Sundanese morphology". *Lingua* 15, 435–450
- Sapir, Edward (1949), *Language. An Introduction to the Study of Speech*. New York [¹1921]
- Silnitsky, George (1993), "Typological Indices and Language Classes: A Quantitative Study". *Glottometrika* 14, 139–160
- Silnitsky, George (1998), "Correlational Analysis of the Indo-European Morphological System". *Journal of Quantitative Linguistics* 5, 81–95.
- Sil'nickij, Georgij B. & Jachontov, K. S. & Jachontov, Sergej E. (1986), "Primenenie korreljacionnogo i faktor-nogo analiza k tipologii jazykov" [Correlation and factor analysis in language typology]. *Aktual'nye voprosy derivatologii i derivatografii*, Vladivostok, 105–125
- Skalička, Vladimír (1935), *Zur ungarischen Grammatik*. Praha: Facultas philosophica Universitatis Carolinae
- Slavičková, Eleonora (1968), "Toward a Typological Evaluation of Related Languages", *Travaux linguistiques de Prague* 3, 281–289
- Son, Jos van & Peters, Toine (1989), "Quantificerende taalclassificatie". *TABU, Bulletin voor Taalwetenschap* 19.1, 16–30
- Stepanov, Arthur V. (1995), "Automatic Typological Analysis of Semitic Morphology", *Journal of Quantitative Linguistics* 2, 141–150
- Thorndike, Edward L. (1943), "Derivation Ratios". *Language* 19, 27–37
- Voegelin, Charles F., Ramanujan, R. A. & Voegelin, Florence M. (1960), "Typology of Density Ranges I: Introduction". *International Journal of American Linguistics* 26, 198–205

Sebastian Kempgen, Bamberg (Deutschland)

Werner Lehfeldt, Göttingen (Deutschland)



Bibliographische Angaben / Bibliographical Entry:

Sebastian Kempgen, Werner Lehfeldt: Quantitative Typologie. In: G. Booij et al. (eds.), *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung*, 2. Halbband, Berlin—New York 2004, 1235–1246.

Copyright und Lizenz / Copyright and License:

© Prof. Dr. Sebastian Kempgen 2021; <http://orcid.org/0000-0002-2534-9423>
Bamberg University, Germany, Slavic Linguistics
<https://www.uni-bamberg.de/slavling/personal/prof-em-dr-sebastian-kempgen/>
<mailto:sebastian.kempgen@uni-bamberg.de>

License: by-nc-nd



February 2021, postprint