DOI: 10.20378/irb-92449

Akzent und Wortlänge: Überlegungen zu einem typologischen Zusammenhang[1]

Sebastian Kempgen, Konstanz

Abstract

Advances in synergetic linguistics have led to an integration of previously isolated laws of language into a kernel of a unified theory of language. From some basic assumptions it can be deduced that there must a reciprocal relationship between the number of phonemes in a language and the length of its words. From a functional viewpoint accent and tone serve the same purpose as phonemes, so the same relationship has been proposed für suprasegmentals. Based on an empirical investigation of word lenght in Bulgarian and Macedonian, the article comes to the conlusion that this is not necessarily the case: the presence of a moving accent alone cannot be expected to have an effect on word lenght, whereas tone clearly has.

Einleitung

Die Suche nach Gesetzen hat in der Geschichte der neueren Sprachwissenschaft eine lange Tradition. Ein wesentliches Ziel der *Junggrammatiker* war es beispielsweise, die von ihnen untersuchten Sprachveränderungen zu Gesetzen, Lautgesetzen nämlich, zu verallgemeinern. Wenn wir heute nicht mehr geneigt sind, hier wirklich von „Gesetzen" zu sprechen, so deshalb, weil diese Ansätze zwangsläufig auf einer vortheoretischen Stufe stehen bleiben mußten: Es gab noch keine Wissenschaftstheorie im heutigen Sinne, die eine Diskussion über grundlegende Begriffe wie ‚Gesetze', ‚Erklärungen', ‚Theorie' und andere ermöglicht hätte.[2]

Die *Universalienforschung* hat vieles zu unserem Verständnis von den Implikationen zwischen den verschiedenen Ebenen der Sprache und ihren Elementen beigetragen. Auch hier fehlt aber noch ein theoretischer Hintergrund, der es ermöglichen würde, von einer Beschreibung der Beziehung zwischen den Fakten zu ihrer Erklärung überzugehen.

Parallel zu dieser Richtung der Sprachwissenschaft ist in den letzten Jahrzehnten der Versuch unternommen worden, Gesetze, die diesen Namen nach einem strengen Verständnis dieses Begriffes auch verdienen, zu formulieren. Grundlegendes dazu hat ein Sprachwissenschaftler beigetragen, dessen Werke gerade heute wieder auf ein breiteres Interesse stoßen, nachdem sie lange weithin unberücksichtigt geblieben sind:

George Kingsley Zipf war es, der schon 1935 in einer Arbeit u. a. den Zusammenhang zwischen der Häufigkeit eines Wortes und seiner Länge zum erstenmal detailliert aufzeigte: *häufig* gebrauchte Wörter sind in der Regel *kurz*, *seltener* gebrauchte *länger*. Diese Tatsache, von deren Gültigkeit man sich mit einem Blick in ein beliebiges Häufigkeitswörterbuch überzeugen kann, versuchte er später mit dem „Prinzip der kleinsten Anstrengung" zu erklären (so der Titel seines bekannten Buches von 1949).[3] Er betrachtete dieses Prinzip als eine dem gesamten menschlichen Handeln unbewußt zugrundeliegende Strategie zur Ökonomie. Auf der Seite des Sprechers äußert sie sich z. B. in der Tendenz zur Minimierung des *Produktionsaufwandes*. Auf der Seite des Hörers entspricht ihr die Tendenz zur Minimierung des *Dekodierungsaufwands*. Zipf spricht hier von der ‚Unifikationskraft' und der ihr entgegenstehenden ‚Diversifikationskraft', zwischen denen jede Sprache ein bestimmtes Gleichgewicht suche. Auf der lexikalischen Ebene wäre es für den Sprecher z. B. am bequemsten, wenn sämtliche Bedeutungen in einem einzigen Wort zusammenfielen. Das aber stünde im Widerspruch zu den Interessen des Hörers, für den es mit der geringsten Anstrengung verbunden wäre, wenn jeder Bedeutung genau *ein* Wort entspräche. Zipfs Annahmen sind in der Folgezeit von anderen Autoren theoretisch weiterentwickelt und mehrfach empirisch überprüft worden und werden heute als das „Zipfsche Gesetz" bezeichnet.[4]

Andere bedeutende Beiträge zur quantitativen Linguistik stammen gerade auch von deutschen Autoren.[5]

Paul Menzerath fiel bei seiner Untersuchung der „Architektonik des deutschen Wortschatzes" (1954) ein weiteres Phänomen auf, nämlich der Zusammenhang zwischen der Wortlänge in Silben und der Länge dieser Silben: je *mehr* Silben ein Wort umfaßt, desto *kürzer* sind diese Silben in der Regel.[6] Vgl. z. B. das einsilbige *streifst* mit einer ausgesprochen langen Silbe mit dem sechsilbigen Wort *Gefahrenzulage*, das fast ausschließlich kurze offene Silben aufweist. Altmann (1980) konnte diesen Zusammenhang theoretisch ableiten und nannte ihn das „Menzerathische Gesetz", das inzwischen sowohl auf Morphem- und Satzebene wie auch für die Inhaltsebene bestätigt werden konnte.[7] Andere Autoren untersuchten die Wortlänge z. B. mit der Absicht, bestimmte Texte einem Autor zuordnen zu können, vgl. etwa Wilhelm Fuchs mit seiner Arbeit „Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen" (1955a, vgl. auch 1955b u. a.). Vor kurzem ist es Rüdiger Grotjahn (1984) gelungen zu zeigen, daß die Verteilung der Wortlänge in verschiedenen Sprachen einem einheitlichen allgemeinen Prinzip folgt.[8]

In seinem 1986 erschienenen Buch mit dem Titel „Zur linguistischen Synergetik: Struktur und Dynamik der Lexik" versucht Reinhard Köhler vor dem hier kurz um-

rissenen Hintergrund verschiedene allgemeine Anforderungen an die Sprache als ein natürliches Kommunikationsmittel in ein einheitliches System zu bringen und die genannten und andere in letzter Zeit formulierte Gesetze, die bisher noch unverbunden nebeneinander standen, systematisch miteinander zu verknüpfen. Dieser Ansatz gehört zu den Versuchen, Sprache unter synergetischen Gesichtspunkten zu betrachten[9], d.h. als ein *selbstregulierendes System* zu begreifen, dessen Elemente miteinander kooperieren und auf diese Weise die Anforderungen, die von außen an die Sprache herangetragen werden, nämlich vor allem die Sicherung der Effektivität der Kommunikation bei gleichzeitiger Wahrung einer genügenden Redundanz, zu erfüllen versuchen, der Entwicklung dieser Anforderungen, die sich ja gleichfalls ständig wandeln, jedoch immer hinterherhinkt.[10] Die Selbstregulation besteht dabei darin, daß die Sprache über Mechanismen verfügt, die ein Gleichgewicht zwischen unterschiedlichen Anforderungen herstellen und Störungen dieses Gleichgewichts abfangen, kompensieren. Diese Mechanismen sind es, die durch Sprachgesetze erfaßt werden können.

Die sprachwissenschaftliche Fruchtbarkeit dieses Ansatzes zeigt sich u. a. darin, daß es Köhlers Kernmodells einer Sprachtheorie erlaubt, neue Hypothesen zu formulieren und an geeigneten Sprachen gezielt empirisch zu testen. Dies ist auch ausdrücklich Köhlers Ziel.[11] Sprachen werden hier nicht mehr – oder nicht mehr in erster Linie – nur um ihrer selbst willen miteinander verglichen, oder um induktiv zu typologischen Verallgemeinerungen zu kommen, sondern dienen als Prüfinstanz für theoretisch abgeleitete Annahmen.

Mit *einer* solchen Folgerung, die aus diesem Modell abzuleiten ist, will ich mich in diesem Beitrag näher beschäftigen. Es geht um den Zusammenhang von *Akzent* und *Wortlänge*.

1 Herleitung der Hypothese

Es läßt sich – abgesehen von dem Zusammenhang mit der Häufigkeit – aus dem Zipfschen Gesetz deduzieren, daß die *durchschnittliche Länge der Wörter einer Sprache möglichst gering sein sollte*.[12] Diese Eigenschaft ist, wie man weiß, direkt von der Phonemebene abhängig.

Betrachten wir den hypothetischen Fall einer Sprache, die nur über ein einziges Phonem *a* verfügt. In einer solchen Sprache kann jedes neu zu bildende Wort nur durch Verlängerung des bisher längsten Wortes gebildet werden:

(1) a, aa, aaa, aaaa ...

Von jeder Länge kann es genau *ein* Wort geben.[13] In einer Sprache mit *zwei* Phonemen, beispielsweise dem Vokal *a* und dem Konsonanten *b*, könnten bei gleicher Wortlänge dagegen schon wesentlich mehr Wörter gebildet werden (vgl. 2); von der Länge 1 kann es zwar wieder nur *ein* Wort geben (wenn jedes Wort mindestens einen Vokal erhalten muß), von der Länge 2 aber schon drei, von der Länge 3 sechs, von der Länge 4 schon mindestens neun usw.

(2) *Länge*	*1*	*2*	*3*	*4*
	a	ab	aaa	aaaa
		ba	aab	aaab
		aa	aba	aabb
			baa	abab
			bab	abba
			abb	baab
			...	baba
				babb
				bbab
				...

Diese Beispiele zeigen: je mehr Phoneme eine Sprache hat, desto mehr Möglichkeiten hat sie, Wörter einer bestimmten Wortlänge zu bilden. Das heißt aber auch, daß mit anwachsendem Phoneminventar die Lexeme dieser Sprache im Durchschnitt theoretisch immer *kürzer* sein können.[14]

Man könnte nun für einen gegebenen Wortschatz von z. B. 100000 Lexemen und für ein gegebenes Phoneminventar die minimal notwendigen Wörter jeder Wortlänge einfach auszurechnen, wenn wir es nicht mit einer natürlichen Sprache zu tun hätten. Diese muß nämlich Rücksicht auf die Schwierigkeiten, bestimmte Konsonantenfolgen auszusprechen, nehmen, außerdem kennt jede Sprache typische und untypische Konsonanten- und Vokalkombinationen, z. B. am Wortanfang und am Wortende, deren Untersuchung schon Trubetzkoy zum Programm gemacht hat, typische Silbenstrukturen usw. Die Regeln der Phonemkombinatorik einer gegebenen Sprache bestimmen also mit, wieviele der theoretisch denkbaren Kombinationen einer gegebenen Länge tatsächlich auch ausgenutzt werden. Wenn es sich bei den hypothetischen Wörtern unter (2) z. B. um deutsche Wörter handeln würde, wäre überhaupt nur ein ganz geringer Teil verwirklicht. Die phonotaktischen Einschränkungen, die aus Sprecher-Sicht wie eine Vergeudung kurzer Lautfolgen aussehen, haben für den Hörer und damit für die Kommunikation insgesamt aber durchaus eine Funktion: sie stellen nämlich eine gute Sicherung des Kommunikationsprozesses gegen Störungen dar und erlauben es oft, verstümmelt übertragene Informationen ohne Nachfrage zu rekonstruieren.

Unter anderem macht nun Köhler in seiner Arbeit auch folgende Bemerkung: „Dieses einfache Modell [gemeint ist die Abhängigkeit der durchschnittlichen Wortlänge von der Phonemzahl] vernachlässigt jedoch u. a. den Einfluß von Tönen und Akzenten auf die Wortlänge: Tonsprachen kommen mit kürzeren Wörtern aus, da sie über ein zusätzliches Merkmal verfügen. In solchen Sprachen wirkt sich die Anzahl der Töne verringernd auf die Wortlänge aus. Sprachen mit beweglichem Akzent besitzen ebenfalls ein zusätzliches Mittel zur Differenzierung, das man z. B. häufig in der Flexion findet." (1986: 24)[15]

Unter einem *freien Akzent*, dies zur Erinnerung, versteht man gewöhnlich einen Akzent, der nicht an eine bestimmte Position im Wort, etwa an eine bestimmte Silbe, gebunden ist. Kann die Betonungsstelle dann auch noch in der Flexion variieren, spricht man zusätzlich von einem *beweglichen Akzent*. Köhler trifft diese Unterscheidung nicht, meint aber offenbar u. a. den freien Akzent.

Was die Wortlänge in Tonsprachen wie dem Chinesischen und dem Vietnamesischen oder in afrikanischen Sprachen wie dem Ewe und dem Tch(w)i betrifft, so ist in der Tat bekannt, daß hier einsilbige Wurzeln und kurze Wörter vorherrschen (vgl. Wendt 1961: 20 ff., 237, 342 ff.). Im Vietnamesischen beispielsweise gibt es sechs Töne, „zwei ebene Töne, bei denen die Stimme ungefähr auf der gleichen Höhe bleibt... zwei ‚melodische' Töne, bei denen die Stimme von einem gewissen Punkt an steigt... zwei ‚knarrende'Töne, die von einer Zusammenpressung der Stimmlippen im Kehlkopf begleitet sind" (Wendt: 342 f.).

(3)	ma	(hoch, eben)	„Teufel"
	mà	(tief, eben)	„was, damit, aber"
	má	(hoch, steigend)	„Wange"
	mâ	(tief, steigend)	„Grab"
	mã	(tief, knarrend)	„Pferd"
	mạ	(tief, knarrend)	„junge Reispflanze"

Auf der Ebene der Ausdrucksseite betrachtet, sind sechs Töne „so viel wert" wie sechs zusätzliche Phoneme oder wie Zweierkombinationen von drei zusätzlichen Phonemen usw. Die sechs Töne könnten die Zahl der Morpheme einer jeden Länge theoretisch versechsfachen.[16] (Sie tun es im übrigen nicht, da auch die Kombinierbarkeit dieser suprasegmentalen Elemente mit den Silben bzw. Morphemen Einschränkungen unterliegt.) Daß in solchen Sprachen irgendein Zusammenhang zwischen Tönen und Wortlänge vorliegt, scheint unmittelbar einleuchtend.[17] Im Sinne der Hypothese wäre weiterhin zu vermuten, daß in Sprachen mit melodischem Akzent – wie z. B. im Serbokroatischen – der Einfluß auf die Wortlänge größer ist als in Sprachen mit einfachem Druckakzent und in Tonsprachen größer als in Sprachen mit melodischem Akzent.[18] Etwas präziser formuliert, lautet die Hypothese also folgendermaßen: *Die Anzahl der suprasegmentalen Mittel wirkt umgekehrt proportional auf die Wortlänge.* Oder anders: *Je mehr suprasegmentale Mittel eine Sprache besitzt, desto geringer die durchschnittliche Wortlänge.*

Ein Zusammenhang der Wortlänge mit dem Vorhandensein von Tönen scheint, wie gesagt, plausibel; was ist aber von der Hypothese eines Einflusses auch des *freien Akzentes* auf die Wortlänge zu halten? Diese Hypothese, die von Köhler eher beiläufig formuliert, aber nicht in sein Sprachmodell eingebaut und empirisch nicht weiter überprüft wird, soll im folgenden etwas genauer untersucht und zum erstenmal empirisch getestet werden.

2 Test der Hypothese

2.1 Voraussetzungen

Bevor die Hypothese einem Test unterzogen werden kann, müssen die verwendeten theoretischen Begriffe in Beschreibungsbegriffe umgesetzt werden. Dies ist in diesem Falle recht einfach. Das *Wort* definieren wir als jede Folge von Phonemen (bzw. Graphemen) zwischen zwei Pausen (bzw. Leerstellen). Die *Wortlänge* wird gewöhnlich in Phonemen (bzw. Graphemen) gemessen, aber auch Silben kommen als sinnvolle Einheiten in Frage. Wir werden beide Arten der Messung verwenden, wobei man aber schon im voraus sagen kann, daß mögliche Auswirkungen auf die Zahl der Phoneme größer sein dürften als auf die Zahl der Silben.[19] Die Wortlänge ist ein quantitativer Begriff, d. h. ihre Messung erfolgt auf der Verhältnisskala.[20] Natürlich kann man nun nicht einfach zwei beliebige Sprachen heranziehen, ihre Wortlängen miteinander vergleichen und die möglicherweise ermittelten Unterschiede auf das Vorhandensein oder Nichtvorhandensein eines freien Akzentes als Ursache zurückführen. Dann wäre es vermutlich sehr einfach, eine ganze Reihe von Ergebnissen, die die Hypothese stützen, zu bekommen.[21] Damit die Resultate auch wirklich die gewünschte Aussagekraft besitzen, muß vielmehr sichergestellt sein, daß die *ceteris paribus*-Klausel beachtet wird. Das heißt: die zu untersuchenden beiden Sprachen müssen in allen relevanten Merkmalen möglichst ähnlich sein. Wir können nicht fordern, daß sie in allen Hinsichten *vollkommen* identisch sein müssen, denn dann könnte ja jede Sprache nur mit sich selbst verglichen werden. Es ist vielmehr plausibel darzulegen, daß die Bereiche, in denen sich die Sprachen unterscheiden, keinen oder höchstens einen zu vernachlässigenden Einfluß auf das untersuchte Merkmal haben. Gesucht sind, mit anderen Worten, zwei Sprachen, die sich im wesentlichen nur dadurch unterscheiden, daß die eine einen freien Akzent aufweist, die andere aber einen festen.

Welche Merkmale es sind, die einen unmittelbaren Einfluß auf die Wortlänge haben, ist von unseren Ausgangsüberlegungen her klar, nämlich die *Anzahl* der vorhandenen Phoneme (bzw. Grapheme) und die *Kombinatorik* dieser Segmente.[22]

a) Phonemzahl

Was die Phonemzahl betrifft, so ist es naheliegend, hier zu fordern, daß die beiden Sprachen in der Klassifikation von Lehfeldt (1975, vgl. auch Altmann/ Lehfeldt 1980: 87 – 101) zur gleichen Klasse gehören, d. h. zu den Sprachen mit *geringer* (<19), *normaler* (19 – 47) oder *hoher* Phonemzahl (>47). Diese Forderung dürfte aber nicht genügen. Es ist darüber hinaus zu fordern, daß die Sprachen eine möglichst ähnliche, wenn nicht identische Phonemzahl aufweisen und auch ihr konkreter Bestand sollte ähnlich sein: eine Sprache mit vielen Vokalen, aber wenigen Konsonanten baut ihre Lexeme bekanntlich anders als eine Sprache mit vielen Konsonanten, aber nur wenigen Vokalen: im ersten Fall haben wir gewöhnlich viele kurze offene Silben, im zweiten Fall dagegen mehr geschlossene Silben mit längeren Konsonantengruppen.

b) Phonemkombinatorik

Was die Phonemkombinatorik angeht, so wäre zu fordern, daß die spezifischen Muster der möglichen Phonemverbindungen in den beiden Sprachen einander möglichst ähnlich sind, wobei man sich auf Gruppen von Phonemen beziehen könnte, also z. B. auf die Verbindungen von Vokal und Konsonant oder von plosiven und frikativen Konsonanten miteinander. Dabei wäre ferner nach typischen Positionen zu differenzieren, d. h. die Kombinatorik im Wortanlaut wäre getrennt von der im Wortinnern und der im Wortauslaut zu betrachten.

c) Weitere Voraussetzungen

Neben den beiden genannten phonologischen Eigenschaften sollten einige weitere Merkmale erfüllt sein, um einen Vergleich linguistisch aussagekräftig durchführen zu können. So sollten beide Sprachen auch über vergleichbare *morphologische* Mittel verfügen, z. B. beide Präfigierung, Suffigierung etc. in ähnlicher Weise verwenden, also etwa beide flektieren, indem Suffixe und Endungen an einen Stamm gehängt werden. Außerdem sollten sie vor allem vergleichbare Verfahren zur Wortbildung besitzen, also z. B. die Präfigierung kennen, die Komposition, die Suffigierung und dergleichen. Desweiteren sollten die Sprachen über eine ähnliche *Sprachgeschichte* mit ähnlichen Sprachkontakten verfügen, um etwa auszusschließen, daß das Lexikon der einen Sprache in starkem Maße durch Fremdwörter und Lehnprägungen geprägt ist, das heißt durch Elemente, auf die die Regeln der eigenen Sprache nur einen mittelbaren Einfluß ausüben, das Lexikon der anderen Sprache aber nicht. Schließlich muß auch gewährleistet sein, daß das *Lexikon* der Sprachen etwa gleich groß ist – man kann z. B. eine große Kultursprache nicht gut mit einem Eingeborenendialekt vergleichen, weil die Notwendigkeit zur Benennung etwa von Gegenständen in beiden Sprachen viel zu unterschiedlich wäre.[23]

2.2 Bulgarisch und Makedonisch

Bei diesen Anforderungen ist die Zahl der Sprachpaare, an denen man die Hypothese sinnvoll überhaupt noch überprüfen kann, sicher nicht mehr groß. Gerade unter den slavischen Sprachen gibt es aber zwei, die die formulierten Anforderungen in nahezu idealer Weise erfüllen: das *Bulgarische* und das *Makedonische*. Diese beiden Sprachen sind nicht nur geographisch benachbart, sie sind auch eng miteinander verwandt, ja, das Makedonische, die Amtssprache der südlichsten jugoslawischen Republik also, galt bis vor kurzem einfach als ein (westlicher) Dialekt des Bulgarischen – für die Bulgaren ist es das heute übrigens immer noch. Beide Sprachen haben eine sehr ähnliche Geschichte – beide sind im lexikalischen Bereich stark durch das Griechische und Türkische beeinflußt worden –, beide gehören zum sogenannten Balkansprachbund und haben hier gleiche Eigenschaften angenommen bzw. entwickelt. Der wichtigste und wesentlichste Unterschied zwischen beiden Sprachen ist tatsächlich der, daß das Bulgarische einen *freien Akzent* aufweist, das Makedonische aber einen *festen* (und zwar auf der drittletzten Wort-

silbe).[24] Beispiele, die die unterschiedlichen Positionen zeigen, sind etwa die folgenden:

(4)					
	Bulg.	ednogodíšen	neopravdánost	zapraší	máslenica
	Mak.	ednogódišen	neoprávdanost	zápraša	maslénica

Das Bulgarische kann folglich mithilfe des Akzentes lexikalische Minimalpaare bilden:

(5) Bulg.				
petà	„Ferse"	–	pèta	„fünfte" (fem.);
pàra	„Dampf"	–	parà	„Münze"
kàlen	„schlammig"	–	kalèn	„gehärtet"

Wenn die von Köhler beiläufig formulierte Hypothese stimmt, so sollte das Bulgarische, das über ein Ausdrucksmittel *mehr* verfügt, im Schnitt folglich *kürzere* Wörter haben als das Makedonische.

Betrachten wir zunächst die oben genannten Voraussetzungen in bezug auf diese beiden Sprachen etwas genauer, wobei ich mich hier der graphematischen Ebene bediene, für die aber im Prinzip die gleichen Anforderungen gelten wie für die phonologische. Die nachfolgende Übersicht zeigt die *Grapheme* beider Sprachen in Transliteration.

(6) Grapheme:

M.:	a b v g d g' e ž z dz i j k l lj m n nj o p r s t k' u f x c č dž š	(31)
B.:	a b v g d e ž z i j k l m n o p r s t u f x c č š št " ' ju ja	(30)

(7) Unterschiede: št, " , ' , ju, ja — nicht im Makedonischen
lj, nj, dz, dž, g', k' — nicht im Bulgarischen

Der Unterschied in der Graphem*zahl* ist minimal, die Unterschiede im Graphem*bestand* sind von nur geringer Bedeutung. Die fehlenden Grapheme werden nämlich in der jeweils anderen Sprache meist durch Graphempaare wiedergegeben. Wichtig für uns ist vor allem, daß die meisten der hier genannten Grapheme nur eine geringe Frequenz im Lexikon aufweisen, so daß ihr Einfluß auf Lexembildung und Wortlänge als sehr gering betrachtet werden kann.

Auch die *Kombinatorik* der Phoneme bzw. Grapheme beider Sprachen ist sich sehr ähnlich, wie hier nur an einigen Beispielen gezeigt werden soll:

(8)		
	Bulg.	pr-, gl-, str-, mn-, zd-; -st, -jk, -jn, -rs, -rc
	Mak.	pr-, gl-, str-, mn-, zd-; -st, -jk, -jn, -rs, -rc

2.3 Empirische Ergebnisse

Um die Hypothese nun konkret zu überprüfen, wurde den Rückläufigen Wörterbüchern beider Sprachen (OBRATEN REČNIK 1967 bzw. OBRATEN REČNIK 1975) eine systematische Stichprobe[25] entnommen, und zwar derart, daß aus jeder Kolumne das erste und das mittlere Wort berücksichtigt wurde. Damit wurde ca. jedes 25. Wort aus dem Lexikon ausgewertet. Das Lexikon des Bulgarischen enthält insgesamt >70000 Wörter (S. 5), das des Makedonischen nur unwesentlich weniger, nämlich 64578 (vgl. 355), so daß die resultierenden Stichproben einen fast gleichen Umfang hatten (2582 bzw. 2540 Wörter).[26]

2.3.1 Silben

Für die Verteilung der Wortlänge in Silben in den beiden Sprachen ergaben sich die Werte der Tab. 1:

Tab. 1

	Makedonisch		Bulgarisch	
Silben x_i	f(i)	%	f(i)	%
0	0	0,00 %	0	0,00 %
1	43	1,69 %	67	2,59 %
2	448	17,64 %	452	17,51 %
3	865	34,06 %	935	36,21 %
4	750	29,53 %	709	27,46 %
5	352	13,86 %	309	11,97 %
6	64	2,52 %	82	3,18 %
7	15	0,59 %	22	0,85 %
8	3	0,12 %	5	0,19 %
9	0	0,00 %	1	0,04 %
10	0	0,00 %	0	0,00 %
Σ	2540	100,00 %	2582	100,00 %

Abbildung 1 zeigt beide Verteilungen im Vergleich (und zwar die relativen Häufigkeiten). Sie sind, wie man sieht, fast deckungsgleich. Das Bulgarische zeigt dabei eine geringe Verschiebung zu Wörtern mit einer etwas höheren Silbenzahl.

Dieser optische Eindruck bestätigt sich auch, wenn man die *mittlere Wortlänge* (in Silben) für beide Sprachen errechnet:

		Silben/Wort	
	Makedonisch	8807 / 2540 =	3,4673
(9)	Bulgarisch	8852 / 2582 =	3,4284
	Differenz		0,0390

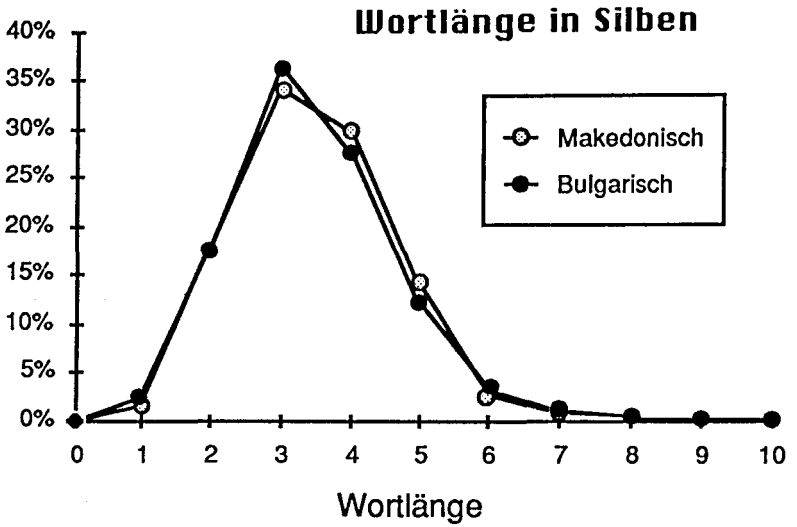


Abb. 1

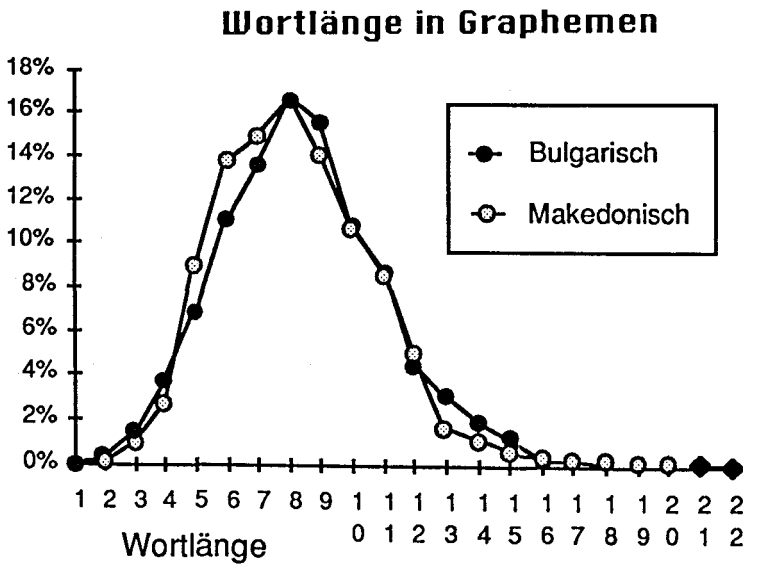


Abb. 2

Das heißt: in beiden Sprachen weisen die Lexeme im Durchschnitt praktisch die gleiche Länge in Silben auf; die festgestellte Differenz kann vernachlässigt werden, sie ist auf keinen Fall signifikant und nicht aussagekräftig. *Einen Einfluß des Akzentes auf die durchschnittliche Wortlänge in Silben gibt es also nicht.* [27]

2.3.2 Grapheme

Für die Verteilung der Wortlänge in Graphemen im Bulgarischen und im Makedonischen ergaben sich die Werte der Tab. 2:

Tab. 2

	Makedonisch		Bulgarisch	
Grapheme x_i	f(i)	%	f(i)	%
1	1	0,04 %	3	0,12 %
2	3	0,12 %	7	0,27 %
3	25	0,98 %	36	1,39 %
4	68	2,68 %	97	3,76 %
5	227	8,94 %	175	6,78 %
6	350	13,78 %	285	11,04 %
7	380	14,96 %	353	13,67 %
8	422	16,61 %	429	16,62 %
9	354	13,94 %	399	15,45 %
10	270	10,63 %	279	10,81 %
11	222	8,74 %	226	8,75 %
12	125	4,92 %	116	4,49 %
13	41	1,61 %	78	3,02 %
14	28	1,10 %	49	1,90 %
15	13	0,51 %	29	1,12 %
16	4	0,16 %	6	0,23 %
17	5	0,20 %	8	0,31 %
18	1	0,04 %	4	0,15 %
19	0	0,00 %	2	0,08 %
20	0	0,00 %	0	0,00 %
21	1	0,04 %	1	0,04 %
Σ	2540	100,00 %	2582	100,00 %

Auch hier zeigt wieder eine Graphik die relativen Anteile der einzelnen Wortlängen an der Stichprobe (Abb. 2).

Wie man sieht, weisen die Verteilungen zwar einen relativ ähnlichen Verlauf auf, einige Unterschiede sind aber dennoch nicht zu übersehen: Das Makedonische hat einen höheren Anteil unter den kürzeren Wörtern als das Bulgarische (Längen 5, 6, 7), dagegen einen niedrigeren Anteil unter längeren Wörtern (Längen 9, 13, 14, 15).

Diese Verhältnis zeigt auch die *durchschnittliche Wortlänge* (in Graphemen):

		Grapheme/Wort	
	Makedonisch	20637 / 2540 =	8,1248
(10)	Bulgarisch	21758 / 2582 =	8,4268
	Differenz		-0,3020

Das heißt: im Bulgarischen sind die Lexeme der Stichprobe durchschnittlich um (fast) ein Drittel Graphem länger als im Makedonischen und die Verhältnisse damit gerade andersherum, als es bei Gültigkeit der Hypothese zu erwarten wäre. Im Bulgarischen ist also, mit anderen Worten, im Durchschnitt jedes dritte Wort um einen Buchstaben länger. Ob dieses Ergebnis als *Nichtbestätigung* oder aber als *Widerlegung* der Hypothese zu werten und damit linguistisch zu interpretieren ist oder nicht, darf in diesem Falle nicht einfach intuitiv geschätzt, sondern muß exakt beurteilt werden. Dazu dient ein statistischer Test, auf den ich hier aber nicht im einzelnen eingehen will.[28]

Das Ergebnis dieses Tests ist jedenfalls: die beobachtete Differenz zwischen beiden Sprachen ist zu groß, als daß wir annehmen könnten, die beiden Verteilungen seien noch gleich. Oder anders: *das Bulgarische weist eine signifikant andere Verteilung der Wortlänge als das Makedonische auf*, und zwar in Richtung auf längere Wörter (im Lexikon). Dieses Ergebnis ist folglich keine Bestätigung der Ausgangshypothese; in den beiden Sprachen scheinen Wortlänge und Ausprägung des Akzentes von anderen Faktoren dominiert zu werden.

3 Interpretation der Ergebnisse

Wenn man nach Gründen dafür sucht, warum sich die Hypothese nicht bestätigte, so lassen sich verschiedene Fälle unterscheiden: Es könnten rein *formale* Gründe vorliegen, also z. B. Größe und Gewinnung der Stichprobe. Es können aber auch *inhaltliche* Gründe vorliegen, bei denen wieder zwei Möglichkeiten unterschieden werden können. Es könnten einerseits spezielle sprachliche Gründe vorliegen, die die Hypothese genau für das Bulgarische und das Makedonische nicht bestätigen. Solche Gründe führen nicht zu einer Modifikation der allgemeinen Hypothese. Zum zweiten sind Überlegungen generellerer Natur möglich, die die Plausibilität der Hypothese insgesamt betreffen. Alle drei Möglichkeiten sollen nacheinander kurz betrachtet werden.

3.1 Formale Fehlerquellen

Eine mögliche Ursache für Verzerrungen der Ergebnisse ist natürlich z. B. die Größe der Stichprobe, die erhoben wurde. Anders als für phonologische Fragestellungen existiert für unser Problem bisher keine exakte Empfehlung, wie groß die Stichprobe sein muß, um mit einer gewissen Sicherheit, z. B. einer Sicherheit von 95 %, sagen zu können, daß die ermittelten Stichprobenanteile der einzelnen Wortlängen höchstens um einen gewissen Wert von ihrem „wahren" Anteil am Lexikon verschieden sind.[29] Auf jeden Fall gilt für unsere Untersuchung aber, daß mögliche Fehler für beide Stichproben in gleicher Weise gelten, so daß von daher kein ernsthafter Einwand gegen die Ergebnisse möglich ist. Allenfalls die Frage, ob die Ergebnisse der Hypothese widersprechen oder sie nur nicht stützen, könnte von

dem Umfang der Stichprobe abhängen. Eine völlige Umkehrung des Trends ist aber auf keinen Fall zu erwarten. Mit anderen Worten: es sind inhaltliche, linguistische oder sprachliche Gründe für das Ergebnis verantwortlich zu machen.[30]

3.2 Spezielle Ursachen

Spezielle Gründe dafür, warum sich die Hypothese nicht bestätigte, könnten in folgendem liegen:

a) Das Makedonische ist noch nicht lange genug eine eigene Sprache. Das Makedonische wurde ja erst nach dem Zweiten Weltkrieg als Literatursprache kodifiziert und normiert, und es könnte sein, daß der Zeitraum von 40 Jahren noch zu kurz ist, als daß sich die in der Hypothese vorausgesagte Tendenz schon in spürbarem Umfang ausgewirkt haben könnte. Entwicklungsprozesse wie der fragliche können sich nämlich nur dann bemerkbar machen, wenn die Voraussetzungen über einen längeren Zeitraum gültig waren. Entgegenhalten könnte man diesem Einwand, daß die (zentralen) makedonischen Dialekte, aus denen die Literatursprache später entstanden ist, die gleiche Regel zur Festlegung des Akzentes schon seit Jahrhunderten kannten, so daß eigentlich Zeit genug gewesen sein müßte, um die fragliche Tendenz sichtbar werden zu lassen.[31] An einen Dialekt sind aber in kommunikativer Hinsicht, gerade was den notwendigen Umfang des Lexikons betrifft, natürlich geringere Anforderungen gestellt als an eine Literatursprache. Außerdem wurde die normierte makedonische Literatursprache nicht völlig neu geschaffen; es bestand vielmehr seit Anfang des 20. Jahrhunderts Einigkeit über die wesentlichen Merkmale der zu schaffenden Literatursprache.[32]

a) Der Akzent ist im Makedonischen nicht hundertprozentig fest, oder, mit anderen Worten: es gibt im Makedonischen doch einen mindestens teilweise freien Akzent. Das Rückläufige Wörterbuch verzeichnet in einem Anhang (S. 329 – 354) ca. 1400 Wörter, die eine Betonungsstelle haben, die von der allgemeinen Regel abweicht. Das sind 2,19 % des gesamten Lexikonumfanges, also eine doch relativ kleine Zahl. Unter diesen Wörtern sind sehr viele Fremdwörter (*reklàma, legènda*), aber, wenn auch in geringerer Zahl, auch echt makedonische Wörter, wie z. B. *novogràdba* „Neubau", *protivmèrka* „Gegenmaßnahme". In den genannten Fällen handelt es sich jedoch immer um Komposita (deren Kern sich im übrigen der allgemeinen Betonungsregel fügt). Die in diesem Anhang aufgeführten Wörter haben einen Akzent auf der vorletzten oder auf der letzten Silbe. Fast immer ist diese spezielle Akzentstelle an bestimmte Suffixe oder Morpheme gekoppelt, z. B.

(12)

(optim)-ìstka	(simboli)-izìra	(asimil)-ìra	(zoo)-lòg
(festiv)-àlski	(anat)-òmski	(atmosf)-èrski	(dekad)-èntski
(šovin)-ìzam	(kombin)-ìran	(progres)-ìven	(muskul)-òzen
(šocij)-àlen	(popul)-àren	(definit)-ìvno	(xrono)-mètar

(milicion)-èr (realiz)-àtor (sedim)-ènt (final)-ìst
(popul)-àrnost (purit)-ànec

Das heißt: die Abweichungen von der allgemeinen Betonungsregel sind morphologisch begründet und insofern erklärlich. Das Makedonische bedarf zwar aufgrund seiner kurzen Geschichte noch immer einer vergleichsweise stärkeren Normierung, doch scheint es mir abwegig, in diesem abweichenden Akzentverhalten einen Einfluß auf den Zusammenhang mit der Wortlänge zu vermuten. Die gleichen Wörter könnten ja ohne weiteres der allgemeinen Betonungsregel untergeordnet werden, ohne daß sich daraus irgendwelche Schwierigkeiten (etwa Homonymien) ergäben. Tatsächlich verläuft die Entwicklung des Makedonischen in dieser Richtung, wenn auch langsam.

Beide spezielle Ursachen sind also meiner Überzeugung nach *nicht* für die empirischen Ergebnisse verantwortlich zu machen.[33]

3.3 Allgemeine Ursachen

a) Wie Untersuchungen zum Russischen, in dem der Akzent *frei* ist, gezeigt haben, wird der freie Akzent in dieser Sprache in erster Linie nicht dazu verwendet, lexikalische Minimalpaare zu bilden, sondern dazu, verschiedene Wortformen eines und desselben Wortes voneinander zu unterscheiden, d. h. er wird v. a. als *beweglicher* Akzent eingesetzt; man könnte auch sagen: der freie Akzent ist hier primär ein lexematisch freier Akzent und nicht ein lexikalisch freier Akzent. Seine Funktion liegt also auf der Ebene der Grammatik. Tornow gibt in seiner Arbeit (1984: 59 f.) folgende Zahlen für den von ihm untersuchten Grundwortschatz von 3223 Lexemen an: in nur 53 Fällen lagen lexikalische Minimalpaare vor, vgl.

(13) muká „Mehl" – múka „Qual" 53 von 3223

aber in 432 Fällen grammatische Minimalpaare

(14) lésa Gen. Sg. – lesá Nom. Pl. „Wald" 432 von 3223

Naturgemäß kann nur die erste Funktion, der lexikalisch freie Akzent, ihren Niederschlag in einem Lexikon der üblichen Art finden, d. h. wenn Lexeme und nicht die einzelnen Wortformen verzeichnet werden. Für das Bulgarische und das Makedonische liegen vergleichbare Zahlen nicht vor, doch können wir annehmen, daß ähnliche Verhältnisse auch hier gelten. Darüber hinaus dürfen wir generell annehmen, daß in allen Sprachen mit ausgeprägter Derivationsmorphologie der freie Akzent vornehmlich grammatisch genutzt wird. Das heißt: die untersuchte Hypothese ist zweifellos in folgender Weise zu präzisieren: Nur dann, wenn für eine bestimmte Sprache gilt, daß der freie Akzent auch wirklich in größerem Umfang zur Bildung lexikalischer Minimalpaare benutzt wird, ist überhaupt denkbar, daß sich eine signifikante Auswirkung auf die durchschnittliche Wortlänge der Lexeme ergeben könnte. Mit anderen Worten: die Randbedingun-

gen, unter denen die Hypothese gelten soll, wären in dieser Richtung zu präzisieren. Den genannten Befund könnte man auch als Argument dafür nehmen, nicht Lexeme, sondern Wortformen in Texten zu untersuchen. Da sich die Flexionsendungen beider Sprachen hinsichtlich ihrer Länge so gut wie nicht unterscheiden, ist aber nicht zu erwarten, daß sich die empirischen Ergebnisse bei einem solchen Vorgehen wesentlich ändern würden.

b) Eingangs wurde dargelegt, daß die Zahl und die Kombinatorik der Phoneme einer Sprache einen unmittelbaren Einfluß auf die Möglichkeit, mit ihnen neue Wörter zu bilden, haben. Dieser Einfluß gilt jedoch nicht uneingeschränkt. Wie informationstheoretische Untersuchungen gezeigt haben, regelt die Phonemkombinatorik die Struktur von Wörtern nur bis zu einer bestimmten Länge und mit wachsender Wortlänge mit schwindendem Einfluß (vgl. Piotrovskij 1965: 164). Der Grund hierfür ist einleuchtend. Der Rahmen für die Phonemkombinatorik ist in erster Linie die Silbe, allenfalls noch das Morphem. Lange Wörter bestehen aber fast nie nur aus einem einzigen Morphem, d. h. werden nicht mehr unmittelbar aus Phonemen zusammengesetzt, sondern aus mehreren Morphemen. Für den Sprecher einer Sprache ist dies von großem Vorteil, denn so muß er nicht für jede neue Bedeutung eine vollständig neue Lautfolge lernen, es wird vielmehr eine große Anzahl von Lautfolgen mehrfach verwendet. Für die hier untersuchte Hypothese bedeutet dies aber, daß der mögliche Wirkungsgrad suprasegmentaler Mittel auf die durchschnittliche Wortlänge des Lexikons eine weitere Einschränkung erfährt.

Schließlich kann man sich noch folgendes überlegen – und das ist das gewichtigste Argument von allen –: das Vorhandensein eines freien Akzentes leistet zur Unterscheidung von Lexemen genau so viel wie *ein einziges zusätzliches* Phonem. Die Abhängigkeit der Wortlänge vom Phoneminventar ist aber nicht so direkt, als daß sich *ein* Phonem mehr oder weniger schon bemerkbar machen könnte.

3.4 Schlußfolgerungen

Nach den oben dargelegten empirischen Ergebnissen und den theoretischen Überlegungen läßt sich abschließend die folgende *Gegenhypothese* aufstellen: *das Vorhandensein eines freien Akzentes allein reicht noch nicht aus, um einen Effekt auf die Wortlänge auszuüben. Schon geringe Einflüsse anderer sprachlicher Ebenen, wie z. B. der morphologischen, sind offenbar für Wortstruktur und Wortlänge in viel stärkerem Maße verantwortlich, als es das bloße Vorhandensein des freien Akzentes sein könnte.* Bei einem melodischen Akzent scheint es dagegen schon eher möglich, daß es gelingen könnte, einen Einfluß auf die Wortlänge nachzuweisen. Bei Tonsprachen schließlich ist der empirische Zusammenhang zwischen dem Vorhandensein der suprasegmentalen Mittel und der Kürze der Wörter in den schon erwähnten Sprachen zunächst einmal evident; aber auch hier muß man in jedem Einzelfall genau prüfen, ob der empirischen Korrelation auch ein inhaltlicher Zusammenhang entspricht, also prüfen, was Ursache und was Wirkung ist, welche

Eigenschaften der betreffenden Sprache für das sichtbare – bzw. hörbare – Ergebnis verantwortlich sind.[34]

Anmerkungen

1 Diese Arbeit entstand im Rahmen des von der Stiftung Volkswagenwerk geförderten Projektes „Sprachliche Synergetik". Für kritische Anmerkungen danke ich G. Altmann und R. Köhler (Bochum).

2 Aus dem gleichen Grunde mußte die Auffassung von der Sprache als einem „Organismus", die im Grunde in einem synergetischen Modell (s. u.) ganz ähnlich vertreten wird, eine Metapher bleiben. Zum Begriff des Lautgesetzes vgl. etwa Schneider (1973).

3 Zur Bedeutung Zipfs vgl. genauer Altmann (1985).

4 Wie in den natürlichen Sprachen die Wörter kodiert sind, sind im Morsecode die einzelnen Buchstaben kodiert: dem häufigsten Laut, dem *e*, ist das kürzeste Signal zugeordnet usw. bis hinunter zum seltensten Laut, dem z, dem die längste Signalfolge entspricht. Auf diese Weise wird erreicht, daß die Länge einer gesendeten Botschaft – und damit auch die mechanische Anstrengung des Morsers – im Durchschnitt möglichst gering gehalten wird. Der Vorteil einer solchen variablen Codierung für die menschliche Kommunikation liegt sozusagen auf der Hand. Informationstheoretisch gesprochen, sind (natürliche) Sprache wie (künstlicher) Morsecode Codes mit *variabler Codewortlänge*. Unter anderen Bedingungen ist ein Code mit *konstanter* Codewortlänge oft praktischer als ein variabler. Die binäre Codierung des Alphabetes in Computern ist hierfür ein Beispiel: Jeder Buchstabe wird hier durch eine z. B. stets 7-stellige Folge von Nullen und Einsen charakterisiert, d. h. die Länge der Codierung jedes Buchstabens ist konstant.

5 Dies scheint kein Zufall zu sein. Zipf selber lehrte als Germanist in Harvard und hatte einige Zeit in Deutschland studiert. Für sein erstes Buch stellte ihm u. a. Paul Menzerath umfangreiches Material zur Verfügung.

6 So auch schon Menzerath/Meyer-Eppler (1950). Erste Ansätze im Bereich der Phonetik wurden schon viel füher entwickelt, vgl. Menzerath (1928), also gleichzeitig mit Zipf.

7 Vgl. Altmann, Best, Beöthy (1982), Gerlach (1982), Heups (1983), Köhler (1982, 1984), Rothe (1983), Sambor (1984), Fickermann/Markner/Jäger/Rothe (1984), Schwibbe (1984), Teupenhayn/Altmann (1984).

8 Das heißt: es konnte nachgewiesen werden, daß die jeweilige Anzahl der Wörter mit 1, 2, 3 und mehr Silben nicht zufällig zustande kommt, sondern daß ihr ein allgemeines Gesetz zugrunde liegt, das durch die negative Binomialverteilung am besten angenähert werden kann.

9 Eine leicht verständliche Einführung in Grundgedanken der Synergetik stammt von ihrem Begründer, Hermann Haken (vgl. 1981). Zur Übertragung auf sprachwissenschaftliche Probleme vgl. ausführlicher Köhler/Altmann (1986).

10 Für funktionale Erklärungen, die mit dem Prinzip der kleinsten Anstrengung verknüpft werden, hat Holenstein (1978: 45) den Begriff „Poronomie" geprägt (von gr. *poros* „Weg, Mittel").

11 Vgl. 1986: 16. Konkret beschäftigt er sich mit folgenden Zusammenhängen, die er an Material aus dem Deutschen überprüft (vgl. 98): 1) der Länge eines Wortes als Funktion seiner Frequenz (je häufiger ein Wort ist, desto kürzer ist es), 2) der Bedeutungsvielfalt eines Wortes als Funktion seiner Länge (je länger ein Wort ist, desto spezifischer ist seine Bedeutung), 3) der Kontextfreiheit eines Wortes als Funktion seiner Bedeutungsvielfalt (je spezifischer die Bedeutung eines Wortes, desto größer seine Kontextbeschränktheit) und 4) der Frequenz eines Wortes als Funktion seiner Kontextfreiheit (je größer die Kontextfreiheit eines Wortes, desto höher seine Frequenz).

12 Explizit gemacht, ist die genaue Herleitung folgendermaßen: 1) Für die Sprache gilt das Prinzip der kleinsten Anstrengung; 2) je größer die Phonemzahl, desto geringer die durchschnittliche Wortlänge; 3) der freie Akzent ist ein phonologisches Mittel; 4) das Vorhandensein eines freien

Akzentes verringert die Wortlänge. Man erkennt hier leicht das Hempel-Oppenheim-Schema der wissenschaftlichen Erklärung bzw. Deduktion. Es sind 1) und 2) die allgemein geltenden Gesetze, 3) ist die Beschreibung des konkreten Sachverhaltes und 4) ist die logische Folgerung aus 1) bis 3).

13 Genauer: es kann nur einen *Ausdruckskomplex* geben. Um des Argumentes willen soll die Möglichkeit zur Toleranz von Homonymen an dieser Stelle unberücksichtigt bleiben.

14 Diese Überlegungen gelten auf graphematischer Ebene natürlich nur für Sprachen, die im wesentlichen das Prinzip einer Lautschrift verfolgen. Auf die chinesische Schrift z. B. ließen sie sich nicht ohne weiteres übertragen.

15 Der gleiche Hinweis auf die angenommene wortverkürzende Funktion dieses Akzentes findet sich noch an zwei weiteren Stellen (vgl. 1986: 54 und 60).

16 Wendt irrt, wenn er sagt: „Durch die sechs Tonstufen besitzt das Vietnamesische theoretisch die Möglichkeit, sein Phoneminventar zu versechsfachen." (1961: 244). Anstelle von *‚Phonem*inventar' muß es natürlich richtig *‚Morphem*inventar' heißen.

17 Auf die Frage, warum Töne dann nicht öfter eingesetzt werden, als es tatsächlich zu beobachten ist, läßt sich mit Bolinger (1978: 479) z. B. folgendermaßen antworten: „Lexical tone is only one of the phonemes making up a word, and context can easily make it redundant. Accent is directly meaningful, tone is only distinctive."

18 Nur nebenbei sei bemerkt, daß sich aus diesen Überlegungen folgende Arbeitshypothese ableiten läßt: Da das Serbokroatische seinen melodischen Akzent gegenwärtig verliert, müßten auf lange Sicht die Wörter dieser Sprache länger werden, wenn nicht ersatzweise andere Mittel zur Unterscheidung herangezogen werden. Eine Möglichkeit besteht dabei natürlich einfach darin, daß die entstehenden Homonymie-Fälle noch toleriert werden, weil der Kontext das Gemeinte eindeutig erkennen läßt. Dieser Fall wird in Köhlers Modell durch die die Tendenz zur lexikalischen Unifikation, die einer Minimierung des Kodierungsaufwandes entspringt, erfaßt.

19 Eine Messung der Wortlänge *in Morphemen* ist zwar auch möglich, hier jedoch nicht gemeint: da es sich bei dem Akzent um ein phonologisches Mittel handelt, sind Auswirkungen auf die Länge der Lexeme in Einheiten der gleichen Ebene, also der phonologischen, gemeint.

20 Die Meßvorschriften, die für die verschiedenen Skalen bzw. Begriffe im einzelnen erfüllt sein müssen, erläutert Carnap (1974: 59 – 124) sehr gut.

21 Die hier verfolgte Möglichkeit, die Hypothese an einem Sprach*paar*, das bestimmten Bedingungen genügt, zu überprüfen, ist natürlich nur eine Möglichkeit. Eine andere bestünde darin, eine möglichst große Zahl von Sprachen zu untersuchen, wobei dann allerdings die „Randbedingungen" nicht gleich gehalten werden können.

22 Und natürlich die Frequenz der Lexeme, von der wir aber hier absehen wollen, da vergleichbares Material nicht vorliegt.

23 Vgl. ähnlich Köhler/Altmann (1983: 429): „So ist die Lexikongröße ein Bedürfnis, das weit mehr kulturspezifisch ist als die anderen drei hier behandelten, die sich vielleicht auf psychologische Universalien zurückführen lassen."

24 Das Makedonische verwirklicht damit ein allgemein eher seltenes Muster der Akzentpositionierung, vgl. die kurze Übersicht bei Bolinger (1978: 481 – 482).

25 Im Gegensatz zu einer zufälligen bzw. autoritativen Stichprobe, vgl. hierzu Altmann/Lehfeldt (1980: 120 ff). Unsere Untersuchung betrifft folglich die Lexem-Länge der Wörter, nicht die Text-Länge der Wortformen. Eine zusätzliche Überprüfung der Hypothese an Texten würde das hier gewonnene Bild vervollständigen (und evtl. korrigieren).

26 Wir nehmen also an, daß die Systembedürfnisse nach Kodierung einer bestimmten Anzahl von Bedeutungen (vgl. Köhler 1986: 52) und nach Redundanz (vgl. Köhler 1986: 55 f.) für beide Sprachen in gleicher Weise gelten.

27 Diese Verteilung der Wortlänge ist es, die Grotjahn (1984) untersucht hat.

28 Es sei nur kurz erläutert, wie vorgegangen wurde: Für beide Stichproben berechnet man die *kumulativen (summierten) Häufigkeitsverteilungen* der relativen Werte und ermittelt die größte Diffe-

renz, D, zwischen beiden Verteilungen, die auf ihre Signifikanz hin getestet wird. Dies ist der sogenannte „Kolmogorov-SmirnovTest", der für den Fall zweier unabhängiger Verteilungen das empfohlene Testverfahren ist (Siegel 1976: 123 ff., vgl. auch Kriz 1973: 187 – 189). Mit ihm prüft man, ob zwei Stichproben-Verteilungen aus der gleichen Population stammen bzw. aus Populationen mit gleichen Verteilungen (dies ist die sogenannte „Nullhypothese") oder nicht (dies ist die Alternativhypothese). Der Wert D ist in unserem Falle 0,044705 (zu beobachten bei der Wortlänge 7, nur knapp gefolgt von der Differenz bei Wortlänge 8). Für dieses D berechnet man leicht einen sog. „kritischen Wert", d. h. den Wert, den D nicht überschreiten darf, wenn man noch von zufälligen Abweichungen zwischen beiden Verteilungen sprechen will. Dazu werden die beiden Stichprobenumfänge n_1 und n_2 in eine Formel eingesetzt (vgl. Siegel 1976, Tafel M, S. 265). Bei einer Irrtumswahrscheinlichkeit von a = 0.05 berechnet sich D_{max} als 0,038007. Der „Kolmogorov-Smirnov"-Test ist zwar eigentlich für stetige Daten gedacht, das Resultat bleibt jedoch auch bei anderen Tests gleich.

29 Zu Empfehlungen für phonemstatistische Untersuchungen vgl. Altmann/Lehfeldt (1980: 122 ff.).

30 Eine andere Möglichkeit, die Hypothese zu testen, bestünde, wie schon gesagt, natürlich darin, *viele* Sprachen beider Gruppen anstelle eines Sprach*paares* zu vergleichen. – Einen Einfluß könnte auch die Tatsache haben, daß hier notwendigerweise von der *Häufigkeit* der Lexeme abstrahiert wurde, von der ja bekannt ist, daß sie die – individuelle – Wortlänge direkt beeinflußt.

31 Zur – bisher nicht vollständig geklärten – Frage der Entstehung des Antepaenultima-Akzentes in den makedonischen Mundarten vgl. Koneski (1966: 89 ff.). Offenbar sind innersprachliche Gründe ebenso verantwortlich wie der Einfluß benachbarter Balkansprachen (Griechisch, Albanisch). Vom synenergetischen Standpunkt ist besonders interessant, daß in den Dialekten mit Antepaenultima-Akzent *nicht* gleichzeitig auch – wie in anderen Gebieten – eine Reduktion unbetonter Vokale zu beobachten ist (vgl. Koneski 1966: 12).

32 Seit dem Erscheinen des Buches „Za makedonskite raboti" von Krste Misirkov, Sofija 1903 (auch engl. unter dem Titel „On Macedonian Affairs").

33 Der „wahre" Grund für die größere Länge der bulgarischen Wörter ist vermutlich in der ingesamt etymologischeren Orthographie und Orthoepie dieser Sprache zu suchen. So verzichtet das Makedonische z. B. auf Geminaten wie *–nn–* in Suffigierungen, vgl. Bulg. *zakónnost*, aber Mak. *bezzákonost*, *–ss–* in Präfigierungen, vgl. Bul. *razséanost*, aber Mak. *raséanost*; weitere Faktoren sind z. B. Unterschiede in häufigen Suffixen wie B. *-nie*, aber Mak. *-n'e*. Zusammengenommen dürfte dies einen größeren Effekt haben als die Tatsache, daß das Makedonische die Graphempaare *št* und *ja* verwendet, wo das Bulgarische nur *ein* Graphem benötigt. Das zeigt nur die Gültigkeit unserer allgemeinen Schlußfolgerung (s. u.). Zu dem Ergebnis dürfte weiterhin die Tatsache beitragen, daß im Bulgarischen die 1. Ps. Sg. Präsens als Nennform der Verben benutzt wird (*-am*), während im Makedonischen der endungslose Stamm (= 3. Ps. Sg. Präs.) die gleiche Funktion erfüllt (*-a*).

34 Es ist z. B. auffällig, daß die „typischen Tonsprachen" dem isolierenden, bzw. in der Terminologie von Skalička (1979: 340), dem polysynthetischen Sprachtypus zuzurechnen sind bzw. ihm nahe kommen.

Literatur

Altmann, G. (1980): „Prolegomena to Menzerath's Law." In: R. Grothjahn, Hrsg.: Glottometrika 2. Bochum, 1 – 10.

– (1985): „On the Dynamic Approach to Language." In: Th. T. Ballmer, ed.: Linguistic Dynamics. Discourses, Procedures and Evolution. Berlin-New York, 181 – 189.

Altmann, G., Best, K.-H. & E. Beöthy (1982): „Die Bedeutungsmenge und das Menzerathsche Gesetz." Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 537 – 543.

Altmann, G. & W. Lehfeldt (1980): Einführung in die Quantitative Phonologie. Bochum (Quantitative Linguistics Vol. 7).
Bolinger, D. (1978): „Intonation Across Languages." In: J. H. Greenberg, ed.: Universals of Language, Vol. 2, Phonology, Stanford, 471 – 524.
Carnap, R. (1974): Einführung in die Philosophie der Naturwissenschaften. Darmstadt.
Fickermann, I., Markner-Jäger, B. & U. Rothe (1984): „Wortlänge und Bedeutungskomplexität. In: J. Boy & R. Köhler, Hrsg.: Glottometrika 6. Bochum, 115 – 126.
Fuchs, W. (1955a): Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen. Köln-Opladen.
– (1955b): „Unterschied des Prosastils von Dichtern und anderen Schriftstellern. Ein Beispiel mathematischer Stilanalyse." Sprachforum I, 234 – 244.
Gerlach, R. (1982): „Zur Überprüfung des Menzerath'schen Gesetzes im Bereich der Morphologie." In: W. Lehfeldt, U. Strauss, Hrsg.: Glottometrika 4. Bochum, 95 – 102.
(1982) Gramatika na săvremennija bălgarski knižoven ezik. Tom I. Fonetika. Sofija.
Grotjahn, R. (1984): „Ein statistisches Modell für die Verteilung der Wortlänge." Zeitschrift für Sprachwissenschaft 1, 44 – 75.
Haken, H. (1981): Erfolgsgeheimnisse der Natur. Synergetik: Die Lehre vom Zusammenwirken. Stuttgart (3. Auflage).
Heups, G. (1983): „Untersuchungen zum Verhältnis von Satzlänge und Clauselänge am Beispiel deutscher Texte verschiedener Textklassen." In: J. Boy & R. Köhler, Hrsg.: Glottometrika 5. Bochum, 113 –133.
Holenstein, E. (1978):„Präliminarien zu einer Theorie der funktionalen Aspekte der Sprache. In: H. Seiler, Hrsg.: Language Universals. Tübingen, 33 – 52.
Köhler, R. (1982): „Das Menzerathsche Gesetz auf Satzebene." In: W. Lehfeldt & U. Strauß, Hrsg.: Glottometrika 4. Bochum, 103 – 113.
– (1984): Zur Interpretation des Menzerathschen Gesetzes. In: J. Boy & R. Köhler, Hrsg.: Glottometrika 6. Bochum, 117 – 183.
– (1986): Zur linguistischen Synergetik: Struktur und Dynamik der Lexik. Bochum (Quantitative Linguistics, Vol. 31).
Köhler, R. & G. Altmann (1983): „Systemtheorie und Semiotik." Zeitschrift für Semiotik 5, 424 – 431.
Koneski, Bl. (1966): Istorija makedonskog jezika. Beograd.
– (1967): Gramatika na makedonskiot literaturen jazik. Del I i II. Skopje.
Kriz, J. (1973): Statistik in den Sozialwissenschaften. Einführung in die kritische Diskussion. Reinbek.
Lehfeldt, W. (1975): „Die Verteilung der Phonemanzahl in den natürlichen Sprachen." Phonetica 31, 274 – 287.
Menzerath, P. (1928): „Über einige phonetische Probleme." In: Actes du premier congrès international de linguistes. Leiden, 104 – 105.
– (1954): Die Architektonik des deutschen Wortschatzes. Bonn.
(1967): Obraten rečnik na makedonskiot jazik. Sostavuvač VI. Miličik'. Skopje.
(1975): Obraten rečnik na săvremennija bălgarski ezik. Sofija.
Piotrovskij, R. G. (1965): „Teoretiko-informacionnaja struktura russkogo slova." Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 18, 149 – 172.
Rothe, U. (1983): „Wortlänge und Bedeutungsmenge: Eine Untersuchung zum Menzerathschen Gesetz an drei romanischen Sprachen." In: R. Köhler & J. Boy, Hrsg.: Glottometrika 5. Bochum, 101 – 112.
Sambor, J. (1984): „Menzerath's Law and the Polysemy of Words." In: J. Boy & R. Köhler, Hrsg.: Glottometrika 6. Bochum, 94 – 114.
Schneider, G. (1973): Zum Begriff des Lautgesetzes in der Sprachwissenschaft seit den Junggrammatikern. Tübingen (TBL 46).
Schwibbe, M. H. (1984): „Text- und wortstatistische Untersuchungen zur Validität der Menzerath'schen Regel." In: J. Boy & R. Köhler, Hrsg.: Glottometrika 6. Bochum, 152 – 176.
Siegel, S. (1976): Nichtparametrische statistische Methoden. Mit einem Vorwort und Flußdiagramm zur Deutschen Ausgabe von W. Schüle. Frankfurt.

Skalička, V. (1979): „Ein ‚typologisches Konstrukt'". In: ders., Typologische Studien. Braunschweig – Wiesbaden, 335 – 341.
Teupenhayn, R. & G. Altmann (1984): „Clause Length and Menzerath's Law." In: J. Boy & R. Köhler, Hrsg.: Glottometrika 6. Bochum, 127 – 138.
Wendt, H. F. (1961): Sprachen. Frankfurt.
Zipf, G. K. (1935): The Psycho-Biology of Language. An Introduction to Dynamic Philology. Cambridge, Mass. (²1965).
– (1949): Human Behaviour and the Principle of Least Effort. An Introduction to Human Ecology. Reading, Mass. (Reprint New York – London 1965).