

Silbenschnitt und Tonakzente

Herausgegeben von
Peter Auer, Peter Gilles und Helmut Spiekermann

Max Niemeyer Verlag
Tübingen 2002



Die Anschlusskorrelation des Deutschen im Horizont einer Typologie der Silbenstruktur

1. Einleitende Bemerkungen

Bis vor kurzem hatten Fragen des Silbenschnitts wie der Silbe allgemein in der theoretisch orientierten sprachwissenschaftlichen Diskussion einen *haut goût*. Bei Ladefoged/Maddieson (1996:282) heißt es anstelle einer Definition der Silbe "The best that we can do is to suggest that syllables are necessary units in the organization and production of utterances".¹ Jenseits konkreter empirischer Fragen besteht ein zentrales Problem der Silbentheorie darin, zwischen Strukturen, die mit der Sprachpraxis gelernt werden, und dabei genutzten physiologischen und kognitiven Ressourcen (und Beschränkungen) zu trennen. Die silbische Gliederung der Äußerung partizipiert an Strukturierungsprinzipien, die der Sprache vorgängig sind und die sich auch in der Kontrolle anderer motorischer Abläufe wie auch von Gestaltwahrnehmungen ausdrücken; aber diese sind überformt von spezifischen sprachlichen Strukturen, wie nicht zuletzt an deren typologischer Vielfalt abzulesen ist. Gerade auch die neuere 'nicht-lineare' Phonologie, die einen extensiven Gebrauch von der Silbenkategorie macht,² zeigt eine Tendenz, diese Differenzierung mit einer abstrakten Begrifflichkeit zu verdrängen, die Gefahr läuft, nur noch metaphorisch zu sein.³

In diesem Beitrag geht es um sprachspezifische (gelernte) Strukturen: die der Silbe und die des Silbenschnitts, vor allem im Deutschen. Selbstverständlich muss deren Analyse in einem allgemeinen begrifflichen Rahmen erfolgen, letztlich in Hinblick auf die Möglichkeit, die phonologische Ausgliederung und Strukturierung der Silbe in der Sprachproduktion und -wahrnehmung zu fundieren (Abschnitt 2-7). Da die Argumentation in einem typologischen Horizont erfolgt, soll die für das Deutsche in Abschnitt 8 ein Stück weit geklärte Kategorie des Silbenschnitts durch einen Vergleich mit den Silbenstrukturen des Marokkanischen verallgemeinert werden (Abschnitt 9-10). Abschnitt 11 deutet Perspektiven der weiteren Forschung an.

¹ Letztlich ein Zitat aus dem einflussreichen Handbuch von Ladefoged 1982.

² Für einen Überblick vgl. Blevin (1995).

³ Das ist z.B. dann der Fall, wenn ausdrücklich "phonologisch" genannte "Silbenstrukturen" auch anderen Zeichensystemen unterlegt werden. Hier kann die in Mode gekommene Erforschung von Strukturierungsfähigkeiten geborener Taubstummer weiterführen, wenn sie dieser Neigung zur Metaphorisierung nicht unterliegt.

2. Das Silbenproblem aus vortheoretischer Sicht

Auf einer vortheoretischen Ebene lässt sich die Silbe empirisch als die kleinste selbständig zu äußernde Lauteinheit bestimmen: Subsyllabische Einheiten erschließen sich erst mithilfe analytischer Operationen – in der Regel in Reaktion auf den Umgang mit einer Alphabetschrift. Dem entspricht die phonographische Fundierung der Schriftsysteme. Sowohl in der Schriftgeschichte wie in der beobachtbaren Stabilisierung von Schriftsystemen zeigt sich die Silbengliederung der Wortformen als kritische Schwelle, deren Überwindung Schwierigkeiten bereitet: Auch bei der Übernahme von alphabetischen Schriftsystemen spielt sich oft genug deren silbenschriftliche Umnutzung wieder ein (vgl. Daniels/Bright 1996). In der Forschung herrscht inzwischen Konsens, dass die Alphabetschrift als eine genuine Erfindung bezeichnet werden kann, bei der diese Schwelle nur aufgrund einer typologischen Besonderheit der semitischen Sprachen überwunden wurde: die bei diesen gegebene Isolierbarkeit subsyllabischer Konstituenten durch ihre grammatischen bzw. lexikalischen Funktionen.

Dieser Zusammenhang zwischen Schrift- und Lautanalyse ist nun nicht nur eine kontingente Angelegenheit der Fachentwicklung, die theoretische Fragen nicht berührt; vielmehr spiegelt sie sich in den dominanten phonologischen Konzepten, worauf H. Lüdtkke (1969) nachdrücklich hingewiesen hat. Traditionell hatte die Lautbetrachtung ohnehin eine subsidiäre Funktion für die Vermittlung der Schrift: Priscian (6. Jhd.) fasste den distinkten Laut als funktionale Nutzung einer lautlichen Diskriminierung. Insofern grenzte er das Phonem von einem 'naturgeschichtlichen', physikalisch-physiologischen Ereignis, dem *sonus* ('Lautgeräusch'), ab als *vox qui scribi potest* ('Laut, den man schreiben kann', I, 1,1, bei Keil 1864:5). In dieser Tradition ergaben sich die phonetischen Grundbegriffe (die insofern immer phonologische waren) gewissermaßen durch die Rückprojektion der graphischen Repräsentation der Wortformen in Buchstabenketten auf die Lautstruktur der Äußerung – eben als segmentale Kette von Phonemen. Diese Begrifflichkeit bestimmt den fachlichen *common sense* bis heute – nicht zuletzt in der fest etablierten Unterscheidung von segmentaler und prosodischer (=suprasegmentaler?) Phonologie, die auf den antiken Schreib-/Leseunterricht zurückgeht: Alles, was dort von Belang, aber nicht mit den Mitteln der Alphabetschrift (segmental) repräsentiert war, wurde als Begleiterscheinung betrachtet, die nur im rhetorischen Unterricht systematisch behandelt wurde, und zwar auf der Grundlange von Annotationen an den geschriebenen Text.⁴

In dieser alphabetschriftlich auf den Kopf gestellten Lautbetrachtung war und ist die Silbe eine solche prosodische Begleiterscheinung⁵ – in der antiken schriftfinalisierten Lautbetrachtung nicht anders bei den heutigen Syllabierungsalgorithmen. Lüdtkke ist zuzustim-

⁴ Von daher auch die entsprechende Terminologie: Die Bezeichnung der Diakritika (und des damit Bezeichneten) als *Akzent*, lat. *ad-cantus* ('das, was dazugesungen wird') eine Lehnübersetzung des griech. *pros-odia*.

⁵ Das erklärt auch den Terminus: *Silbe* aus griech. *syllabää* zu *syn-lambanoo*, 'zusammenfassen'. In der strikten Progression des Anfangsunterrichts wurde das 'syllabierende' Zusammenziehen der Buchstaben in einem Folgekurs zum Anfängerkurs der 'Alphabetisten' betrieben: Diese beschäftigten sich mit den einzelnen Buchstaben und ihrer hypostasierenden Lautierung, deren 'Zusammenfassen' zu Einheiten den Anfängern damals die gleichen Schwierigkeiten bereitete, wie heute bei einer synthetischen Vorgehensweise.

men, dass viele der Unklarheiten der neueren Phonologie durch die unreflektierte Fortschreibung dieser Tradition bedingt sind. Dass diese Probleme nicht deutlicher geworden sind, hat nun nicht zuletzt sprachtypologische Gründe: Tatsächlich ist die Lautstruktur vieler (der weitaus meisten?) Sprachen so beschaffen, dass eine solche Sichtweise durchaus zu plausiblen Analysen führt, insbesondere in der abendländischen Schulsprache schlechthin, dem Lateinischen, aus dessen Metrik (dem Kernstück der rhetorischen Sprachbetrachtung) auch die Grundbegriffe der neueren Silbenphonologie stammen (s.u.). Angelehnt an die landläufigen typologischen Sortierbegriffe werde ich hier von einem aggregativen Silbenbau sprechen, also einer Silbenstruktur, die sich als Zusammenfügung ihrer Konstituenten fassen und insofern auch in den üblichen Konstituentenmodellen darstellen lässt. Daneben gibt es aber auch andere Bauformen mit nicht aggregativen Strukturen, wie schon Trubetzkoy (1939) in seinem typologischen Überblick festgestellt hat: Für ihn war der "Silbenschnitt" des Hochdeutschen mit der Anschlusskorrelation (fester und loser Anschluss) ein Beispiel für einen solchen nicht-aggregativen Silbenbau. Das ist der Ausgangspunkt für das Folgende.

3. Silbe und Silbenschnitt

Zunächst sind einige Vorklärungen als Korrektiv gegenüber der verbreiteten Engführung der phonetischen Betrachtung erforderlich. In der apparativen Phonetik ist es bisher nicht gelungen, die Messverfahren auf die Silbenproblematik zu kalibrieren. Aber auch die 'naivere' impressionistische Herangehensweise an phonetische Fragen, die in der Regel auch die neueren phonologischen Arbeiten bestimmt, erfordert einige Klarstellungen, insbesondere in Hinblick auf die unterschiedlichen Implikationen der Fremd- gegenüber der Selbstwahrnehmung lautlicher Phänomene.

Bei der Selbstwahrnehmung dominiert die Registrierung artikulatorischer Bewegungsmomente, supraglottal vor allem von Kontakterscheinungen der Artikulationsorgane. Prägnant sind hier also konsonantische Artikulationen (Kontakt- bzw. Engebildungen der Artikulatoren). Anders ist es bei der (auditiven) Fremdwahrnehmung: Prägnante Eindrücke entsprechen hier dem, was man in der Phonetik mit großer Sonorität bezeichnet, vor allem also vokalische Artikulationen.

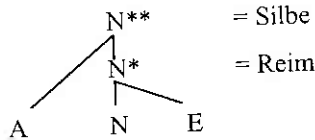
Die Dissoziierbarkeit dieser beiden Perspektiven spiegelt sich deutlich in der Anfangsphase des Schrifterwerbes, in der die Anfänger sich bemühen, ein wahrnehmbares Gegenstück zu den Schriftzeichen zu finden. Über deren Hyperlautierung, die seit über 2000 Jahren konstitutiv für den 'alphabetistischen' Unterricht ist,⁶ finden sie zuerst Anhaltspunkte bei den Konsonanten. Bei eigenen Schreibversuchen resultieren daraus dann die in der

⁶ Die *potestas* der Buchstaben, die in dem traditionellen Dreischritt des alphabetistischen Lehrgangs seit der Antike nach den Buchstabennamen (*nomen*) und ihrer Gestalt (*figura*) zu lernen war (wie es in gewisser Weise auch heute noch der Fall ist).

Sprachdidaktik sog. Skelettschreibungen: $KD\langle in \rangle D\langle er \rangle W\langle a \rangle G\langle e \rangle N$.⁷ Wie auch später Studierenden im Phonetikunterricht bereitet ihnen die Diskriminierung von Vokalen, also Lauten ohne eindeutige Kontaktwahrnehmungen, die größten Probleme.

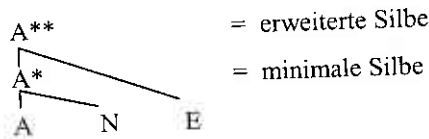
Im Folgenden spreche ich von der Bewegungssilbe, wenn damit die (supraglottalen) artikulatorischen Bewegungsabläufe in der Silbe gefasst werden sollen, und von der Sonoritätssilbe, wenn es um die Silbe als auditive Wahrnehmungseinheit mit einer spezifischen Sonoritätskontur geht. Diese beiden Betrachtungsweisen des Aufbaus der Silbenstruktur sind komplementär und lassen sich mithilfe der üblichen Konstituentenmodelle darstellen. Für die Sonoritätssilbe ist es inzwischen analog zu syntaktischen Modellen üblich, diese als 'Projektionen' ihrer Köpfe aufzubauen, mit dem Vokal als Kopf (Kern bzw. Nukleus).⁸ Bezeichnet man eine Projektion mit einem abgeleiteten Symbol für ihren Kern (im Folgenden notiert mit einem *), so ist die Erweiterung des Nukleus um einen Endrand also ein N^* , die Erweiterung des N^* um einen Anfangsrand ein N^{**} . Die Übersetzung in die sonst verwendeten Kategorien liegt auf der Hand: N^{**} ist die Silbe; N^* ist der Reim :

(1)



Die Bewegungssilbe kann in ähnlicher Weise als Projektion des Ausgangspunktes ihres Bewegungsablaufs, also ihres Anfangsrandes A , dargestellt werden: Eine offene Silbe ist dann die Erweiterung des Anfangsrandes um einen diese Bewegung (sonor) füllenden Vokal, also A^* , eine Schließung mit einem Endrand eine weitere Projektion A^{**} :

(2)



Die Sonoritätssilbe bestimmt entsprechend ihrem Gegenstand die vorsprachwissenschaftliche Phonetik in den klassischen Schulsprachen, und daran anschließend auch die neuere (westliche) Sprachwissenschaft. Das Modell der Bewegungssilbe entspricht dagegen dem Bau der semitischen Sprachen und leitet das Schriftsystem in der arabischen Nationalgrammatik seit über tausend Jahren, wo mit dem Konzept der Bewegung (ar. *harakah*) die Silbenstrukturen von den konsonantischen Rändern her gefasst werden. Konsonanten haben in diesem Sinne ein explosives bzw. Öffnungspotential, das vokalisch 'gefüllt' werden

⁷ Im Gegensatz zur üblichen Meinung handelt es sich hier aber nicht einfach um Konsonantenschreibungen, sondern um Abstraktionen von der silbischen Gliederung mit der Dominanz der Anfangsränder, s.u.

⁸ Wo wie hier Darstellungen und Argumentation einigermaßen verbreitet sind, werden sie nicht mit Literaturverweisen belegt, ausführlicher dazu Maas (1999). Zu den verwendeten Abkürzungen: S = Silbe, R = Reim, A = Anfangsrand, N = Nukleus, E = Endrand.

kann; es wird nicht realisiert, wenn der Konsonant implisiv bleibt. Eine direkte Umsetzung erfährt dieses Modell in der arabischen (und auch der hebräischen) Schrift: Geschrieben werden nur die konsonantischen Ränder der Bewegungsilbe, was dieses Schriftsystem im Zusammenspiel von extrem beschränkten phonotaktischen Filtern (keine komplexen Ränder, insbesondere keine komplexen Anfangsränder⁹) und morphosyntaktisch gesteuerter Qualität der vokalisierenden 'Füllung' durchaus funktional macht(e).¹⁰ Paradoxaerweise kommt die Tragweite dieses Modells aber erst bei der Verschriftung neuarabischer Varietäten zur Geltung, die anders als das klassische bzw. Hocharabische auch nicht vokalischesch erweiterte Öffnungssilben aufweisen, s.u. Abschn. 10.

Beide Konzepte, das der Bewegungs- wie das der Sonoritätssilbe, finden sich nun auch in der sprachwissenschaftlichen Tradition, allerdings in einer Weise verquickt, die die komplementären Perspektiven verdeckt hat. Bei den einflussreichen neueren Silbentheoretikern, bei Saussure nicht anders als bei Pike, wird das Konzept der Bewegung mit dem der Sonorität amalgamiert: die Silbe wird verstanden als kombinierte Öffnungs- und Schließungsbewegung, eingebunden in eine Kontur mit einem Gipfel beim vokalischem Kern.¹¹

Gewissermaßen quer zu dieser komplementären Modellierung der Silbengliederung als Bewegungs- und/oder Sonoritätssilbe liegt eine weitere Differenzierung, die an die angesprochene Dissoziierung von segmentaler und prosodischer Betrachtung gebunden ist. In der Tradition der antiken Prosodie-Lehre werden prosodische Strukturen von der Syllabierung bis zur Akzentuierung aus Eigenschaften der segmentalen Verkettung sozusagen hochgerechnet, so wie dies auch die neueren metrischen Algorithmen explizit unternehmen. Das ist auch möglich bei Sprachen mit aggregativem Silbenbau wie dem Lateinischen, in denen die prosodische Struktur gewissermaßen segmental kodiert wird. Bei anders gebauten Sprachen ist eine solche Reduktion nicht möglich, wie schon Sievers (1876) zeigt, der aus diesem Grund die Kategorie des Silbenschnitts (bei ihm "schwach und energisch geschnittener Akzent") in die phonologische Analyse einführt.¹²

⁹ Nur sekundär entstehen über Apokopierungen komplexe Endränder am Wortrand in *pausa*.

¹⁰ Diese Sichtweise spiegelt sich auch in der arabischen Terminologie, bei der im übrigen die Bezeichnung der Laute und der ihnen korrespondierenden Schriftzeichen systematisch unterschieden werden: Die maximale Öffnung, die Vokalisierung mit [a], wird *fatha* genannt (wörtlich 'Öffnung'), die nicht realisierte Öffnung *sukuun* 'die Ruhe', beides in der Regel nur als Diakritika an den konsonantischen Schriftzeichen markiert. Zur Silbentheorie der arabischen Tradition vgl. Fischer 1967.

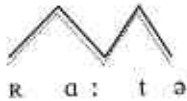
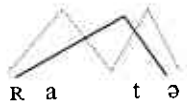
¹¹ Vgl. Saussure (1916:79-91), Pike (1943:116-119).

¹² Gegenüber der Euphorie der damals jungen apparativen Phonetik, die sich selbst in den "phonetischen Ateliers" der Weltausstellungen des 19. Jhd. feierte und im Sinne des neuen positivistischen Wissenschaftsverständnisses die sprachwissenschaftlichen Grundbegriffe auf 'Messbares' zu reduzieren versuchte, verwies Sievers auf die Grenzen des apparativ Kontrollierbaren und forderte, dass zunächst einmal der Gegenstandsbereich geklärt werden müsse, bevor in diesem Horizont eine partielle Modellierung für eine apparative Operationalisierung unternommen werden könne. In diesem Sinne entwickelte Sievers eine perzeptiv Phonetik, für die er Operationalisierungen suchte, die uns heute z.T. reichlich befremdlich vorkommen, etwa seine "Schallanalyse". In der sprachwissenschaftlichen Phonetik muss für ihn (im Gegensatz etwa zur medizinischen oder auch physikalischen Phonetik) die funktionale Betrachtung grundlegend sein, die auf lautliche Unterscheidungen abstellt, mit der in einer Sprache Formen unterschieden werden. In diesem Sinne gehören die Sieverschen Überlegungen, die sich in unterschiedlicher Akzentuierung bei einer ganzen Reihe der Phonetiker seiner Generation finden lassen, in die Tradition der Phonologie; im Folgen-

Die Sieversche Modellierung des Silbenschnitts beruht darauf, dass er unterschiedliche Dimensionen der prosodischen Gliederung der lautlichen Formen ansetzt, die mehr oder weniger kongruent sein können. Die an der segmentalen Struktur orientierte Sonoritätssilbe, für die schon in den 70er Jahren des vorigen Jahrhunderts Sonoritätsskalen auf der Basis von Wahrnehmungsexperimenten vorlagen (etwa von dem Hals-Nasen-Ohrenarzt Wolf, 1871), fasst er unter dem Begriff der "Schallsilbe".¹³ Ihr stellt er holistische Gliederungsmerkmale gegenüber, die entsprechend dem oben angesprochenen 'naiven' Silbenverständnis die Silbe als unanalytische Einheit verstehen, mit der komplexere Äußerungen gegliedert werden können; hier spricht er von der "Drucksilbe", orientiert am "Druckakzent" etwa im Deutschen. Für das Deutsche gilt nun im Gegensatz zum Lateinischen, dass diese beiden Silbengliederungen inkongruent sein können.

Das kann an Wörtern wie *Ratte* im Gegensatz zu *Rate* verdeutlicht werden. Auf der Ebene der Sonoritätsgliederung zeigen beide Wörter jeweils zwei abgrenzbare Einheiten, die beide als offene (Schall-) Silben zu bestimmen sind, strukturiert durch die Explosion eines initialen konsonantischen Elementes in ein sonores vokalisches Maximum. Dem stehen aber unterschiedliche Gliederungen in Hinblick auf die Druckdynamik gegenüber, wo sich bei *Rate* eine mit der Sonorität kongruente Druckgliederung in ein erstes syllabisches Maximum und eine zweite Reduktionssilbe zeigt, während Sievers bei *Ratte* nur eine Drucksilbe registriert:

(3)

*Rate**Ratte*

Das messbare Korrelat zu Sievers' Drucksilbe ist nicht geklärt. Fortgeführt wurden seine Überlegungen in gewisser Weise in Stetsons "Motorthorie" der Silbe (1945), der diese artikulatorisch als rhythmische Einheit der Bewegungsabläufe im Atemapparat verstanden wissen wollte, gewissermaßen als infraglottale Bewegungssilbe. Der Akzeptanz dieses Konzeptes standen und stehen weniger einzelne Messergebnisse entgegen als vielmehr die von Stetson (im Gegensatz zu Sievers) propagierte Verabsolutierung dieses Aspektes.¹⁴ Dass er dominant sein kann, zeigen die 'naiven' Anfänger auf dem Gebiet der Lautanalyse: Durch die Orthographie noch nicht beeinflusste Kindergartenkinder oder Schulanfänger,

den spreche ich in diesem Sinne auch von einer "Protofonologie" (wie man z.B. von der "Proto-industrialisierung" des 18. und frühen 19. Jhd. spricht).

¹³ Worauf wohl die heute übliche Redeweise von der Schallfülle der Segmente zurückgeht. Im Folgenden verwende ich den Sieversschen Terminus der Schallsilbe mit dem der Sonoritätssilbe synonym.

¹⁴ Die ausführlichste und in vielen Punkten auch klärende Diskussion dazu findet sich bei Pike (1967:365 – 392); vgl. jetzt auch Spiekermann (2000) für eine signalphonetische Operationalisierung dieser Unterscheidung.

die man zur Silbengliederung anhält, behandeln Formen wie *Ratte* häufig als einsilbig – zum Leidwesen einer Didaktik, die, wenn sie überhaupt mit lautstrukturellen Konzepten wie dem der Silbe umgeht, dann allein am Konzept der Sonoritätssilbe orientiert ist.

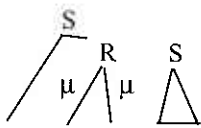
Offensichtlich ist die Möglichkeit einer inkongruenten Gliederung in Druck- und Schallsilben eine Begleiterscheinung von 'akzentzählenden' Sprachen, bei denen die Integration von (im Vergleich zur Sonoritätssilbe) komplexeren Strukturen in eine Drucksilbe eine Folge der Subsumption größerer Äußerungsteile unter einen (Druck-) Akzent ist. Das ist der Fall bei Sprachen wie Deutsch, Niederländisch oder Englisch, die eben auch die Silbenschnittkorrelation aufweisen. Entsprechend ist sie in sog. silbenzählenden Sprachen nicht zu erwarten.¹⁵

4. Silbe und Silbenmessung (das Morenmodell)

Schließlich ist eine weitere Vorklärung erforderlich, die weniger auf die phonetische Wahrnehmung als vielmehr auf begriffliche Implikationen bestimmter Modellierungen zielt. Orientiert an der Anschaulichkeit des alphabetischen Schriftbildes mit der räumlichen Anordnung der Buchstabenzeichen, sind auch die meisten Silbenkonzepte in einer räumlichen Metaphorik gefasst. Dabei wird zumeist, wie bei den verbreiteten Konstituentendarstellungen, mit einer sehr einfachen Topologie der Beziehungen zwischen Elementen operiert, die von einander abgrenzbar sind, wie es eben auch die Buchstaben in der Kette sind.¹⁶ Insofern ist die traditionelle Silbentheorie, wie sie aus der antiken Metrik übernommen ist, auch mit einem quantitativen (räumlichen) Verrechnungsmodus verbunden, der Morenmessung.

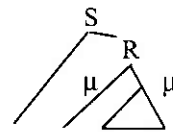
Die antike Metrik (wie die ihr heute darin folgende metrische Phonologie) bewertet das Gewicht einer Silbe nach der Struktur ihres Reims: Ist der Reim einfach, ist er (bzw. von daher geerbt: die Silbe) leicht, verzweigt er, ist er bzw. die Silbe schwer. Etwa (μ = eine More):

(4)



r e : . gis

regis 'König,
Gen.Sg.'

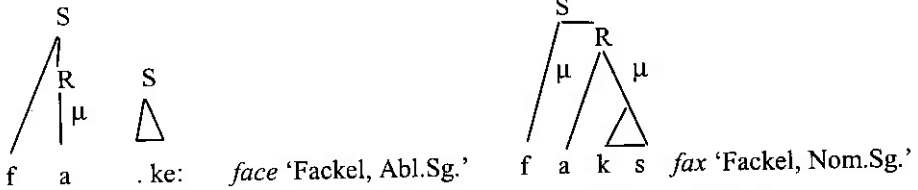


r e : k s *rex* 'König,
Nom.Sg.'

¹⁵ Vgl. zu dieser letztlich auf Pike zurückgehenden 'prosodischen' Typologie etwa Pike (1967).

¹⁶ Jedenfalls wenn man sich an der Buchschrift orientiert und von den in der älteren Schrifttradition häufigen Ligaturen absieht. Bei der Orientierung an der Schreibrift (vor allem bei routinierten Schreibern) sähe das ganz anders aus.

(5)



Mit einer solchen Modellierung lässt sich, wie es in der jüngeren, sog. autosegmentalen Phonologie üblich ist, die qualitative Artikulation gegenüber der rein segmentalen Gliederung (gewissermaßen in Zeittakten, die räumlich veranschaulicht werden) isolieren. Das erlaubt auch die Trennung von quantitativen gegenüber silbenstrukturellen (im Sinne der auch hier benutzten Konstituentenmodelle: konfigurationellen) Kriterien. Mehrere aufeinander folgende Zeiteinheiten können qualitativ gleich artikuliert sein – dann spricht man von Langsegmenten, Langvokalen wie bei lat. [re:ks] *rex* 'König' oder Langkonsonanten wie bei lat. [mit.te.re] ([mit:ere]) *mittere* 'schicken'.

Die Nutzung quantitativer Unterscheidungen ist perceptiv offensichtlich erheblich schwieriger als die qualitativer. Wo Sprachen phonologisch distinktive Quantitätenunterschiede machen, sind diese in der Regel mehrfach abgesichert: Eine stabile Quantitätensprache wie das Finnische markiert die Quantitätenunterschiede der Segmente recht robust in einem Zeitverhältnis von 1 : 3 (vgl. Lehtonen 1970). Im Sprachwandel werden Quantitätensysteme häufig von qualitativen Differenzierungen abgelöst, was u.U. zu Komplikationen wie der Silbenschnittkorrelation in einigen germanischen Sprachen führen kann.¹⁷ Zur eingeschränkten Nutzbarkeit von Quantitäten gehört auch ihre Bindung an silbenstrukturelle Restriktionen: Vokalische Quantitäten treten anscheinend nur tautosyllabisch auf.¹⁸ Umgekehrt sind konsonantische Quantitäten nur heterosyllabisch häufig – in der Lautentwicklung oft nur als Übergangsphase beim Abbau des Endrands. Dabei lassen sie gewissermaßen auf dem Weg zur offenen Silbe mit nur noch sonor (bzw. vokalisch) artikulierten verzweigenden Reimen als Vorstufe eine nicht-sonore konsonantische Artikulation als Kopie des folgenden Anfangsrandes zu. Ein klassisches Beispiel für diese Entwicklung liefert das Altgriechische. In *pausa* bzw. im absoluten Auslaut werden nicht-sonore Konsonanten eines Stamms nicht artikuliert, vgl. zum Stamm *paid-* 'Kind', Genetiv *paid-os* [pai.dos], Vokativ *pai*; im internen Sandhi erfolgen weitgehende Neutralisierungen, bei denen in phonologischer Hinsicht u.U. nur noch eine konsonantische Artikulation (notiert *K*) bewahrt wird, die durch eine Kopie des folgenden Silbenanlauts realisiert wird, z.B. zum Verbalstamm *leip-* 'lass:' wie in *leipoo* [lei.po:] 'ich lasse' die Perfekt Passiv-Bildung (1.Sg.) *le-leim-mai* [le.leim.mai] (= /le.leiK.mai/). Als Spielart dieser Entwicklung kann es

¹⁷ Die Entwicklung der romanischen Sprachen vom Quantitätensystem des Lateinischen aus und die der westlichen neuarabischen Varietäten vom Altarabischen aus verläuft ebenso.

¹⁸ Heterosyllabische ('ambisyllabische') Langvokale setzten vermutlich eine in ihrem Ablauf erfolgende Konturierung der Silbe auf einer anderen Ebene voraus: Tonverläufe, Glottalisierung o. dgl. Die mir bekannten Belege für derartige Erscheinungen (im Hiat) weisen immer auch qualitative Differenzierungen auf.

vorkommen, dass in einer Sprache Wortformen in *pausa* offene finale Silben aufweisen, im Satzzusammenhang aber durch Sandhi-Gemination geschlossene.¹⁹

Sprachen mit einer unbeschränkten phonologischen Nutzung konsonantischer Quantität sind demgegenüber ausgesprochen selten. Ein Beispiel dafür ist das marokkanische Arabische (s.u.), das Langkonsonanten als Variante komplexer Silbenränder aufweist: *dda* [d:ɑ] 'er hat genommen', *fedd* [fæd:] 'er hat geschlossen' usw. Auch heterosyllabische Langkonsonanten unterliegen dort entsprechend keiner besonderen Beschränkung. Im Folgenden werde ich von Geminaten nur bei einer solchen heterosyllabischen (ambisyllabischen) Artikulation von Langkonsonanten sprechen.

5. Silbe und Silbengrenzen (das Problem der Ambisyllabizität)

Ebenfalls auf der konzeptuellen Ebene liegt eine weitere Implikation der metaphorisch räumlichen Modellierung der Silbenstruktur: Die Vorstellung von einer abgrenzbaren Einheit, letztlich wieder abgelesen an der buchstäblichen Repräsentation der Lautstruktur in Alphabetschriften. Die Abgrenzbarkeit kann sich dabei auf die subsyllabischen Elemente wie auf die eventuelle Mehrsilbigkeit in komplexen Wortformen beziehen. Bei den klassischen Schulsprachen war die Abgrenzung der Silben gewissermaßen anschaulich gegeben: Hier wurden im Schreib-/Leseunterricht Silbengrenzen nach dem Modell von Wortgrenzen (also Pausabedingungen) eingeübt – nicht nur als Begleiterscheinung des Metrikunterrichts, sondern vor allem zur Lehre von der Wortbrechung beim Schreiben (vgl. Anm. 18), vielleicht auch beim syllabierenden Diktieren für einen Schreiber. Darauf geht nicht zuletzt die Verwirrung stiftende Bezeichnung der orthographischen Wortbrechung als 'Silbentrennung' zurück. Dass die Silbengliederung einer Wortform zu klaren Grenzen führt, ist jedoch eine typologische Besonderheit – wenn auch vermutlich statistisch der häufigste Fall. Er ist immer da gegeben, wo Sprachen nur offene Silben aufweisen; auch da, wo zwar komplexere Silbenstrukturen möglich sind, aber i.S. einer aggregierenden Syllabierung auch wordmedial die Unterscheidung von Anfangs- und Endrand unproblematisch ist. Dies ist bei den klassischen Schulsprachen nicht anders als beim (Hoch-)Arabischen der Fall, ist aber eben nicht notwendig, wie das Deutsche zeigt.

¹⁹ Etwa im nordwestlichen Okzitanischen /'bla/ 'Getreide', /'ne.gre/ 'schwarz', aber /'blan.'ne.gre/ 'Hirse', vgl. Maas (1970). Bekannter (und ausgiebig analysiert) ist diese Erscheinung als *raddoppiamento sintattico* im Italienischen.

Vennemann (1988) hat versucht, die Grenzwerte einer solchen Dynamik in einem Feld konkurrierender Beschränkungen zu fassen. Das altgriechische Beispiel macht derartige widersprüchliche Spannungen deutlich: da agr. *piptoo* 'ich falle' metrisch mit einer schweren erste Silbe gemessen wird, setzt man üblicherweise eine silbische Struktur [pip.to:] an – im Gegensatz zur zeitgenössischen Schreiberpraxis bei der Wortbrechung, die eine 'Silbentrennung' in *pi=ptoo* vornahm, nach dem Modell des Wortanfangs wie in *ptuoo* 'ich falle' (und wegen des -p#). In der weiteren Entwicklung des Griechischen wurde die Silbenstruktur von alt [pip.to:] optimiert, vgl. ngr. *pefto*, [pɛf.to] – aber nicht so weit, wie die Optimierung ähnlicher Strukturen in dem in dieser Hinsicht restriktiveren Italienischen gegangen ist: vgl. lat. *optimus* [op.ti.mus] mit it. *ottimo* [ot.ti.mo].

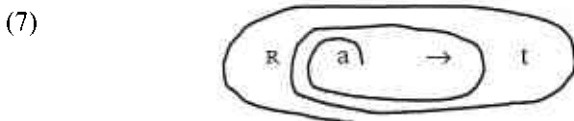
Betrachtet man den deutschen Silbenschnitt aus der Blickrichtung einer aggregierenden Syllabierung, so zeigt sich, dass Silben *fusionieren* können,²⁰ ihre Grenzen also nicht auf der Zeitachse abzutragen sind. In der autosegmentalen Tradition ist es üblich geworden, diese Fusionierung auf ein als *ambisyllabisch* bezeichnetes Segment zu projizieren. Das führt nicht nur in didaktischen Kontexten zu Verwirrungen;²¹ auch in phonologischen Darstellungen findet sich die Vorstellung von einer Verlagerung der Silbengrenze in das ambisyllabische Segment hinein und damit die Suggestion einer nur maßstäblichen Verfeinerung der Gliederung in Zeittakte, wie sie bei geminierten ambisyllabischen Segmenten tatsächlich möglich ist: Bei dem ambisyllabischen /t/ von *it. fatto* 'gemacht' ergibt sich eine zeitliche Zerlegung in ein implosives und ein explosives Segment ['fat.to], die sich auch durch die unterschiedliche Koartikulation in den jeweiligen Silben unterscheiden lassen – bei dem /t/ von *dt. Ratte* ist das aber nicht der Fall (zumindest in der norddeutsch geprägten Hochlautungsvariante unterscheiden sich die beiden /t/-Segmente von *Ratte* und *Rate* nicht).

Vor diesem Hintergrund ziehe ich es vor, im Deutschen, das selbst an der Morphemgrenze keine Geminaten kennt, nicht von ambisyllabischen Segmenten zu sprechen, sondern in der Sievers-Tradition und nach der Vorgabe von Jespersen (1913) von unterschiedlichen Anschlussformen dieser Konsonanten an den vorausgehenden (betonten) Vokal:²² von losem Anschluss (→) als einer schwachen Bindung beim Übergang zwischen problemlos segmentierbaren Elementen und festem Anschluss (⌋) als einer Übergangsform zwischen Elementen, die nicht oder nur problematisch segmentierbar sind.

Im Hochschulunterricht experimentiere ich schon länger mit Darstellungsformen, die diese Besonderheit einer fusionierenden gegenüber einer aggregierenden Syllabierung verdeutlichen sollen. Als probat haben sich dazu Spiral-Symbolisierung erwiesen, die ich auch in Maas (1999) benutze. Damit lässt sich die Silbenschnittkorrelation plastisch als eine der unterschiedlichen Bindung zwischen formal isolierbaren Elementen einer Äußerung darstellen. Mit einer Spirale wird eine abgeschlossene Silbe gegliedert, deren Kern eine Bindung nach rechts hat: mit (mindestens) einer Schlaufe wird die Silbe nach rechts abgeschlossen. Bei losem Anschluss (→) ist ein entsprechender Abschluss ohne angebundnen Konsonanten möglich wie bei *Rate*:



Es kann aber auch ein Konsonant (lose) angeschlossen werden wie bei *Rat, Rad* [Rɑ:t]:



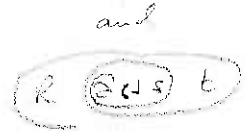
²⁰ Diese Begrifflichkeit benutzt auch Pike bei seiner Diskussion dieser Fragen (1967:380 – 382).

²¹ Hyperlautierungen wie *Ratte* [Ratʰ.tʰə].

²² Wenn man Freude an formalen Regelapparaten hat, lassen sich selbstverständlich zunächst ambisyllabische Konsonanten als Geminaten ableiten, mit denen die silbenstrukturellen Restriktionen des Deutschen definiert werden, um sie dann in einem späteren 'Regel-Zyklus' durch einen Filter wieder zu tilgen, der alle Geminaten in *Simplicia* überführt.

Bei festem Anschluss (┘) wird der Vokal dagegen an das Folgesegment angebunden. Tautosyllabisch steht dieses dann in der Bindungsschleife des Kerns wie bei *matt* oder auch *Rad* in der norddeutschen Aussprache [ˈr at]:

(8)



Steht das Folgesegment nicht im Endrand der gleichen Silbe, sondern fungiert als Anfangsrand der Folgesilbe, wird diese insgesamt in die Kontur der prominenten Silbe eingeschlossen wie bei *Ratte* [ˈr atə]:

(9)



Das Bindungspotential des Kerns ist in diesem Fall durch den Vokal nicht 'gesättigt'; die 'Sättigung' geschieht vielmehr durch die Inkorporation der Reduktionssilbe in die Schleife des Kerns.²³

6. Die (alte) Entdeckung des Silbenschnitts

Diese Differenzierung von Anschlussverhältnissen ist nun keineswegs eine Entdeckung der impressionistischen frühen Phonetik. Vielmehr findet sich die Beobachtung dieser Erscheinung schon in den Hauptwerken der humanistischen Reform des Lateinunterrichts, bei der die einzelsprachspezifisch ausgeprägten, unterschiedlichen Varietäten der Aussprache des Lateinischen in den Blick geraten waren. So hat sich z.B. Erasmus von Rotterdam explizit mit den unterschiedlichen Silbenstrukturen in der romanischen gegenüber germanischen Aussprache des Lateinischen beschäftigt, als er beobachtete, dass die Formen des festen Anschlusses medialer Konsonanten in der "teutonischen" Aussprache kein Gegenstück in der französischen Aussprache haben, die nur lose angeschlossene Konsonanten kennt. In seinem Lehr-Dialog von 1528/29 erläutert er die Konsequenzen für die Unterscheidung von intrinsisch ("von Natur aus") vs. positionslangen Vokalen im Lateinischen: "Et hoc te docebit uulgata Gallorum pronuntiatio, qui omnes fere sonant ut natura longas, duas consonantes aut unam geminam uoce separantes a uocali praecedenti" (S. 150) – Kramer übersetzt: "Auch darüber wird dir die Volkssprache der Franzosen Auskunft geben, wo man fast alle Silben ausspricht, als wären sie von Natur aus lang, weil man mit der Stimme zwei Konsonanten oder einen Doppelkonsonanten vom vorhergehenden Vokal trennt" (S. 151).

²³ Das ist einer der Gründe für diese Darstellungsform statt eines Konstituentenmodells, bei dem eine solche Inkorporation nur mit einer 'rekursiven' Kategorie Silbe darstellbar wäre.

Diese Unterscheidung war damals fester Bestandteil der humanistischen Lateinschule, deren Unterricht kontrastiv zu den muttersprachlichen Vorgaben der Schüler angelegt war. Insofern lagen für die Intellektuellen der damaligen Zeit Überlegungen bereit, um auch Handbücher für den Muttersprachunterricht zu verfassen, wie es etwa bei Valentin Ickelsamer nachzulesen ist. Und so ziehen sich diese Konzepte durch die phonetischen Diskussionen der Frühen Neuzeit bis hin zu den Altvorderen der neueren Sprachwissenschaft wie Eduard Sievers. Als so bei der Debatte um das beste Modell für das neue Hochdeutsch im 18. Jhd. die unterschiedlichen regionalen Ausprägungen in den Blick geraten waren, verstand es sich von selbst, dass die Untersuchung der unterschiedlichen 'Akzente' bzw. Silbenschnitte einen großen Raum einnahm. Daran schloss die spätere Dialektologie an, und zwar schon früh mit Blick auf die unterschiedlichen Sprechregister, also auf den Unterschied von Lento- gegenüber Allegro-Formen, von dessen Analyse sich schon der Sievers-Schüler Winteler Aufschlüsse für den Sprachwandel versprach.

Dass derartige Fragestellungen in der zweiten Hälfte des 20. Jhd. weitgehend aus der Diskussion verschwunden sind, lag an der verkürzten Reduktion der Forschung auf das apparativ Messbare – wovor Sievers u.a. seinerzeit schon gewarnt hatten (vgl. Anm. 12). Die erneute Beschäftigung mit dem Silbenschnitt, die vor allem ein Verdienst von Theo Vennemann ist, kam insofern fast einer Neuentdeckung gleich (vgl. Vennemann 1990, auch Maas/Tophinke 1993).

7. Aggregierende und fusionierende Syllabierung

Eine analytische Klärung des Silbenbegriffs und daran hängender weiterer Differenzierungen wie des Silbenschnitts muss also in einem mehrdimensionalen Betrachtungsraum erfolgen. Die im Vorausgehenden unterschiedenen Dimensionen markieren nur eine erste Näherung als Unterscheidung

- bei der perzeptiven Orientierung
 - nach Bewegungs- oder Sonoritätssilbe,
 - nach segmental durchgegliederter oder holistisch-prosodischer Silbe,
- bei der Modellierung in eine solche mit abgrenzbaren und folglich auch quantitativ messbaren Bestandteilen gegenüber einer mit zugelassenen fusionierten Strukturen.

Hier sind u.U. noch weitere Unterscheidungen anzubringen. Diese Komplexität des Problems steht hinter der eingangs zitierten Ladefogedschen Bemerkung, die in der ursprünglichen Version vollständiger lautete: "Syllables may be considered to be abstract units that exist at some higher level in the mental activity of a speaker. They may be necessary units in the organization and production of utterances" (1975:248). Vereinfacht gesagt läuft das darauf hinaus, dass die verschiedenen konkurrierenden Erklärungsansätze nicht unbedingt falsch sein müssen; allerdings sind sie nicht richtig, wenn sie ihre jeweilige Partialsicht verabsolutieren.

Dieses Problem wird bei einer typologischen Betrachtungsweise deutlich, die nicht nur auf eine deskriptiv angemessene Modellierung abzielt, sondern die Unterschiede im Bau der Sprachen fassen will, wie es Trubetzkoy in seinen "Grundzügen" (1939) mit einer Aufbereitung der phonologischen Systeme von nicht weniger als 200 Sprachen unternommen hat. Ausdrücklich an die Sieverssche 'Protophonologie' anknüpfend, setzt er wie dieser bei der Dissoziierbarkeit der silbenstrukturellen Eigenschaften an, die den traditionellen Silbenbegriff aufspannen (S. 166 ff.). Dabei entwickelt er das herkömmliche, in der Grundbegrifflichkeit der Lateinschule festgeschriebene Silbenmodell der Metrik als einen typologisch besonderen Fall, der daran gebunden ist, dass Quantitätenverhältnisse die Strukturierung bestimmen. Er tut dies mittels des oben dargestellten Morenkonzepts, das für Quantitätensprachen wie die klassischen Schulsprachen oder das Finnische analytisch Sinn macht (S. 169ff.). Dem stellt er nun Sprachen gegenüber, in denen zwar selbstverständlich auch die zeitliche Gliederung der Äußerung prosodisch genutzt wird, diese aber nicht das alleinige oder auch nur das dominante Kriterium für die funktionalen Unterscheidungen ist; im Anschluss an Sievers und Jespersen ist das für ihn beim Deutschen (in der Hochlautung) der Fall, wo er eine Silbenschnittkorrelation ansetzt.

Trubetzkoy entwickelt seine Argumentation dichotomisch: Er setzt einem quantitativen, von ihm arithmetisch genannten Silbenmodell (S. 174) ein holistisches gegenüber, das er silbenzählend nennt. Das erste ist adäquat räumlich zu modellieren, indem die distinktiven Differenzen in einem zeitlichen Nacheinander bzw. einem räumlichen Nebeneinander repräsentiert werden, wie es in einem Konstituentenmodell darstellbar ist (vgl. Abschn. 4). Dieses Modell legt er im Übrigen nicht nur Quantitätensprachen im engeren Sinne, sondern auch Tonsprachen zugrunde, sowohl bei Registertönen wie Konturtönen. Es handelt sich also um den angesprochenen aggregativen Typ der Syllabierung, bei dem die prosodische Struktur aus einer Kombination der Elemente der segmentalen Struktur hochzurechnen ist. Bei dem fusionierenden Typ ist das nicht der Fall: Hier kann die Analyse nicht von unten nach oben erfolgen, sondern muss umgekehrt von oben, jedenfalls von der Silbe zu den sie artikulierenden lokalen Strukturen verlaufen. Dass dabei in der zeitlichen Dauer der Artikulation auch quantitative Verhältnisse hereinspielen, ist an die Materialität der Äußerung gebunden, aber eben nicht distinktiv. In seinen exemplifizierenden Hinweisen findet Trubetzkoy diese Struktur bei den westgermanischen Sprachen Deutsch, Niederländisch und Englisch, in Kombination mit anderen Faktoren auch bei mitteleuropäischen Sprachen in einem arealen Konnex mit dem Deutschen wie Tschechisch und Ungarisch; er diskutiert aber derartige Strukturen auch in ganz anderen Sprachen wie z.B. im Hopi (S. 176).

8. Die Silbenschnittkorrelation im Deutschen

Im Folgenden werde ich an zwei Beispielen die Problematik eines typologisch flexibleren Silbenbegriffs und die damit definierte Kategorie des Silbenschnitts verdeutlichen. Dabei benutze ich Silbenschnitt als allgemeine Kategorie des Silbenbaus, so wie man vom Schnitt eines Kleidungsstücks spricht (oder vom Schnittmuster ...). Typologische Besonderheiten

beruhen im Trubetzkoy'schen Sinne auf der Phonologisierung von Silbenschnittendifferenzen in einer Sprache (und damit auf der von ihm sog. Silbenschnittkorrelation).

Für das erste Beispiel, das Deutsche, kann ich für eine ausführliche Analyse auf Maas (1999) verweisen. Die primäre Ebene der Artikulation der Wortformen ist hier die Akzentkontur, mit der unterschiedlich gebaute Silbentypen korrelieren – aus denen aber nun nicht umgekehrt die Akzentkontur hochgerechnet werden kann.²⁴ Im Vokalismus lassen sich qualitative Differenzierungen ausmachen, die in einer gewissen Annäherung mit Bauprinzipien der Silbe korrespondieren, wie sie auch in Sprachen mit Quantitäten zu finden sind. Das gilt so insbesondere für die nicht-prominente und nicht-reduzierte Silbe, also die Silbe, die nicht durch den primären Akzentkontrast von prominenter und Reduktionssilbe artikuliert ist. Hier finden wir eine Korrelation der Qualitäten im Vokalsystem mit den Bauprinzipien einer offenen und geschlossenen Silbe, also in Korrelation mit dem, was in einem Morenkonzept eine leichte und schwere Struktur des Reims wäre. Auch hier sind die Strukturen allerdings nicht rein lokal definiert, da am Wortrand, also einer höheren morpho-prosodischen Ebene, die gespanntere vokalischer Qualität auch in geschlossenen Silben vorkommt,²⁵ vgl.

- [i] vs. [ɪ]: [ˈtaksis] Plural oder Genetiv zu *Taxi*, [ˈtaksis] *Taxis* (Anordnung)
- [u] vs. [ʊ]: [ˈʃampus] Plural oder Genetiv zu *Shampoo*, [ˈʃampus] *Schampus* (Champagner),
- [e] vs. [ɛ]: [ˈpɔres] Plural oder Genetiv zu *Porree*, [doˈlo:RES] *Dolores* (Vorname).
- [o] vs. [ɔ]: [toˈre:ROS] Plural oder Genetiv zu *Torero*, [riˈno:tse:ROS] *Rhinozeros*.

Die primäre Akzentkontur in der Spannung von prominenter und Reduktionssilbe, die in der Sprachentwicklung der Dynamik des Umbaus von den germanischen Sprachen zum Deutschen entspricht, überformt nun diese Struktur. In der prominenten Silbe korreliert die Vokalqualität mit bestimmten Anschlussstypen: Die gespannten Vokale, die unter diesen Akzentbedingungen als Langvokale im losen Anschluss realisiert werden, sind hier in ihrem Vorkommen frei; sie kommen sowohl in offenen Silben wie in geschlossenen Silben vor, wie regelmäßige Kontraste in Flexionsparadigmen mit wechselnder Syllabierung des Stammes zeigen: *mahne* [ˈma:nə] vs. *mahnte* [ˈma:n.tə]; ebenso für die anderen Vokale: *lehn(t)e*, *höhn(t)e*, *schien(t)e*, *sühn(t)e*, *wohn(t)e*, *buch(t)e*.

Für die ungespannten Vokale gilt, dass sie in der prominenten Silbe nur im festen Anschluss vorkommen, also nicht im absoluten Auslaut; insofern bilden sie den markierten Part der Anschlusskorrelation, wie schon Trubetzkoy gesehen hat (1939:176). Der interessante Fall, der bei der Annahme eines einheitlichen aggregierenden Silbenmodells Schwie-

²⁴ Das unterscheidet meine Analyse von der Vennemanns, der sie ansonsten viel verdankt. Bei Vennemann wird die Akzentuierung aus der Silbenschwere hochgerechnet, für die die Bewertung der Anschlussverhältnisse konstitutiv ist. In dem hier vorgeschlagenen Sinne artikulieren die Anschlussformen dagegen die silbenstrukturellen Potentiale des Lexikons, wenn sie prominent (akzentuiert) artikuliert werden: gespannte Vokale durch losen Anschluss, ungespannte durch festen. In diesem Sinne kommen diese spezifischen Anschlussformen nur als Erscheinung der prominenten Silbe vor.

²⁵ Eine sehr detaillierte Untersuchung dieser Verhältnisse findet sich bei Becker (1999). Bei den Beispielen in phonetischer Notation bezeichnet das Längezeichen den losen Anschluss; der feste Anschluss ist nicht bezeichnet.