2 Phonologie und Phonetik der Intonation

In den folgenden Kapiteln werden die phonetischen und phonologischen Konzepte der Intonationsforschung diskutiert. Die Phonologie (Kap. 2.1 bis 2.4) wird mit einer doppelten Begründung vor der Phonetik (Kap. 2.5) behandelt: Einerseits werden perzipierte phonetische Merkmale vom Sprachteilhaber direkt und unreflektiert funktional interpretiert. Andererseits finden phonetische Untersuchungen im Rahmen der Sprachwissenschaft immer vor dem Hintergrund einer bestimmten phonologischen Theorie oder zur Überprüfung einer phonologischen Hypothese statt. Beides spricht dafür, die phonologische Aspekte vor den phonetischen zu behandeln. In Kap. 2.6 werden die in Auseinandersetzung mit den phonetischen und phonologischen Aspekten entwickelten Beschreibungskategorien erläutert: das Toninventar und die Art und Weise, wie die Zuweisung der Töne durch phonetische Messwerte gestützt wird.

2.1 Prosodische Merkmale und Einheiten

> Prosodie wird verstanden als Oberbegriff für diejenigen suprasegmentalen Aspekte der Rede, die sich aus dem Zusammenspiel der akustischen Parameter Grundfrequenz (F_0), Intensität und Dauer in silbengroßen oder größeren Domänen ergeben.[1]

Dieses Zitat gibt die in der Forschung weithin akzeptierte Definition für sprachliche Phänomene an, die oberhalb der Ebene der Lautsegmente angesiedelt sind. Mit dem Terminus *Prosodie* wird heute wieder auf all jene Phänomene referiert, die man in den 70er und 80er Jahren des 20. Jahrhunderts als *Suprasegmentalia* bezeichnet hat.[2] Bezugseinheit für die in dieser Arbeit behandelten prosodischen Phänomene ist die *Silbe*. Zur Systematisierung der zahlreichen prosodischen Phänomene wird hier die auf Kohler und Schmidt zurückgehende Unterscheidung von prosodischen bzw. suprasegmentellen *Merkmalen* und prosodischen bzw. suprasegmentellen *Einheiten* aufgegriffen.[3]

---

[1] Selting (1995), S. 1.

[2] Vgl. Schmidt (1986), S. 17, Fn. 4. *Prosodie* wird vorgezogen unter anderem von Zifonun et al. (1997), Bd. 1, S. 189 in der Grammatikschreibung, Wiese (1996), S. 26 in der generativen Phonologie, Cruttenden (^{2}1997), S. 1 in der Phonologie der Britischen Schule, Bertinetto/Magno Caldognetto (1993), S. 143 in der italienischen Terminologie und Selting (1995), S. 1 wie zitiert in der Konversationsanalyse. In der generativen Phonologie ist diese Präferenz mit der Einführung der autosegmentalen Phonologie begründet, siehe S. 22.

[3] Kohler (^{1}1977), S. 118f. oder (^{2}1995), S. 110 spricht nur von *prosodischen Merkmalen* und verzichtet auf einen eigenen Terminus für prosodische Merkmale mit sprachlicher Funktion. Schmidt (1986), S. 16-38 schließt an Kohlers Überlegungen an und führt den Terminus *suprasegmentelle Einheiten* für die suprasegmentellen bzw. prosodischen Merkmale mit sprachlicher Funktion ein. *Prosodische Einheit* entspricht in etwa *Prosodem*, vgl. dazu z.B. Hammarström (1963).

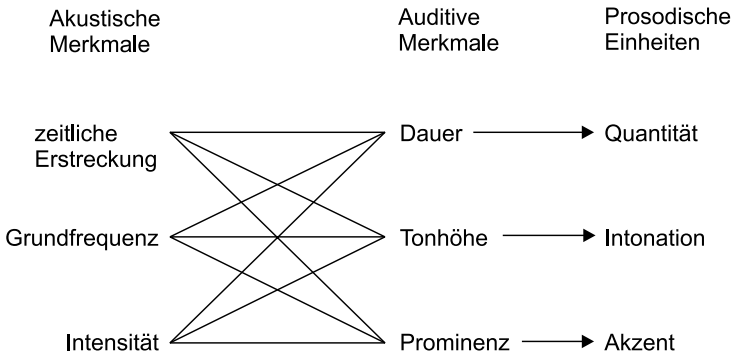

Abb. 2.1: Der artikulatorische Zusammenhang

Prosodische Merkmale sind phonetische Phänomene, die in einem Dreiebenenmodell als auditive, akustische und artikulatorische Merkmale beschrieben werden können. Dazu zählen:

> 1. die auditiven Merkmale Dauer, Tonhöhe und Prominenz, 2. die akustischen Merkmale zeitliche Erstreckung, Grundfrequenz und Intensität und 3. die artikulatorischen Merkmale zeitliche Steuerung der Artikulationsbewegung, Schwingungsverhalten der Stimmlippen, Erzeugung und Zustand des Ausatmungsluftstroms, soweit sie nicht intrinsische Eigenschaften kleinster syntagmatischer Segmente sind.[4]

Prosodische Einheiten sind dagegen phonologische Phänomene. Von prosodischen Einheiten spricht man dann, wenn die Phänomene das Kriterium der phonologischen Distinktivität erfüllen, das heißt, wenn sie sprachliche Funktionen distinguieren. Prosodische Einheiten unterscheiden eine Fülle von syntaktischen, semantischen und pragmatischen Funktionen. Kohler unterscheidet ausgehend von den auditiven prosodischen Merkmale drei prosodische Einheiten, nämlich *Quantität*, *Akzent* und *Intonation*:

> Neben die phonetische D a u e r tritt die phonologische Q u a n t i t ä t, neben die phonetische P r o m i n e n z der phonologische W o r t - und S a t z a k z e n t und neben die phonetische T o n h ö h e der phonologische T o n und die phonologische I n t o n a t i o n.[5]

Prosodische Einheiten sind auditive prosodische Merkmale mit sprachlicher Funktion. Zwischen prosodischen Einheiten und akustischen Merkmalen gibt es dagegen keine eineindeutige Beziehung. Die akustischen Merkmale stehen in einem komplexen artikulatorischen Zusammenhang. Das bedeutet, dass mit einer prosodischen Erscheinung (und einem auditiven Merkmal) jeweils ein ganzes Bündel akustischer Merkmale

---

[4] Schmidt (1986), S. 17f.

[5] Kohler (^{2}1995), S. 110. Gibbon (1995), S. 458 unterscheidet in seinem Modell die linguistischen Domänen *word level* mit den Einheiten *tone*, *stress* und *length* und *supra-word level* mit *intonation*, *accent* und *rhythm*.

korreliert.[6] Abb. 2.1 verdeutlicht den artikulatorischen Zusammenhang. Besonders augenfällig ist der artikulatorische Zusammenhang beim Akzent. Es ist umstritten, welches der akustischen Merkmale Intensität, Grundfrequenz und zeitlicher Erstreckung primär ist bzw. ob es überhaupt ein primäres Merkmal gibt.[7] Aber auch für die Intonation besteht ein artikulatorischer Zusammenhang.

In der vorliegenden Arbeit werden alle prosodischen Merkmale auf die Kategorien Quantität, Akzent und Intonation bezogen. Auf eine ihnen nebengeordnete „Restkategorie" für Phänomene wie Sprechtempo, Rhythmus, Stimmqualität und Pausen, wie sie Möbius (1993, S. 9) vorschlägt, wird verzichtet. Sprechtempo, Rhythmus und Pausen gehen partiell in die Analyse der metrischen Struktur ein. Die Funktionen von Stimmqualität und Klangfarbe werden hier nicht untersucht.

Die prosodische Einheit *Intonation* steht im Mittelpunkt der vorliegenden Untersuchung. In der Forschungsliteratur kommen unterschiedliche Konzeptualisierungen von *Intonation/intonazione* vor.[8] *Intonation* wird mit dem auditiven Merkmal 'Tonhöhe' gleichgesetzt,[9] als 'Intonation im weiteren Sinn' und damit als Synonym zu *Prosodie* verstanden[10] oder als prosodische Erscheinung wie oben ausgeführt neben Akzent und Quantität gestellt.[11] In der vorliegenden Untersuchung wird Intonation in der letztgenannten Konzeptualisierung verwendet: Intonation ist der „Gebrauch des Melodieverlaufs" und korreliert mit der Tonhöhe (auditiv) und dem Schwingungsverhalten der Stimmlippen am Kehlkopf (artikulatorisch).[12] Unter den korrelierenden akustischen Merkmalen kommt der Grundfrequenz besondere Bedeutung zu. Intonation meint den Melodieverlauf der Gesamtäußerungen und beschränkt sich nicht auf die letzte Tonhöhenbewegung.

Intonation wird in der autosegmentalen Phonologie durch Akzent-, Phrasen- und Grenztöne auf der Ton-Ebene repräsentiert. Die Intonationskontur wird als Summe lokaler Töne aufgefasst. Die in Kap. 3 referierten Studien zeigen, welche syntaktischen, diskurssemantischen und einstellungsbezogenen Funktionen durch Töne unterschieden werden können. Die in dieser Arbeit untersuchten Funktionen aus dem *System des Verhaltens in Gesprächen* werden allerdings nicht durch einzelne Töne oder Intonationskonturen ausgdrückt, sondern durch *intonatorische Verfahren*.[13] „Gebrauch des Tonhöhenverlaufs" heißt deshalb auch: 'Verwendung eines bestimmten intonatorischen Verfahrens'.

Lexikalische Töne oder Tonakzente kommen in den untersuchten Varietäten des Deutschen und Italienischen nicht vor.[14]

---

[6] Die Ergebnisse des Greifswalder Projekts zeigen, dass für die Perzeption von Funktionen Merkmalskombinationen signifikanter sind als Einzelmerkmale, vgl. Bandt et al. (2001).
[7] Siehe unten, S. 8f.
[8] Zur Begrifflichkeit in der englischsprachigen Debatte vgl. Ladd (1996), 6ff.
[9] Z.B. von Selting (1995), S. 1.
[10] Z.B. von Altmann et al. (1989), S. 2; Canepari (1985), S. 31; Pheby (1981), S. 839.
[11] Z.B. von Kohler (^{2}1995), S. 121 und Dominicis (1992), S. VIII.
[12] Vgl. im Detail Kap. 2.5.
[13] Siehe dazu Kap. 4.2 und 4.3.
[14] Zu Tonakzenten in deutschen Dialekten vgl. Schmidt (1986).

Auch *Akzent* wird in der Forschung unterschiedlich konzeptualisiert. Die wichtigste Unterscheidung ist die zwischen *Wortakzent* und *Satzakzent*.[15] In der vorliegenden Studie wird Akzent als Satzakzent oder, zutreffender, als *Äußerungsakzent* verstanden und folgendermaßen definiert: Akzent ist die Hervorhebung einer oder mehrerer sprachlicher Einheiten (Silbe, Wort, Konstituente) durch phonetische Mittel. Primäres auditives Merkmal des Akzents ist die Prominenz. Für die akustischen Merkmale des Akzents gilt das oben zum artikulatorischen Zusammenhang Ausgeführte ganz besonders: Zur Prominenzsignalisierung können alle phonetischen Merkmale beitragen. Der (Äußerungs-)Akzent ist durch kommunikative Erfordernisse (zum Beispiel Markierung der Informationsstruktur) gesteuert.[16]

Auch der *Wortakzent* zeichnet sich durch auditive Prominenz aus. Weil der Wortakzent aber weitgehend unabhängig von kommunikativen Erfordernissen ist und stattdessen nach abstrakten Regeln zugewiesen wird,[17] sind akustische Korrelate nicht in jedem Fall nachweisbar. Für die Perzeption von Wortakzenten mag in manchen Fällen das Sprachwissen ausreichend sein.[18] Wortakzente werden in der vorliegenden Untersuchung nicht empirisch untersucht.[19]

Dem Deutschen wird traditionell ein auf Grundfrequenzvariation gegründeter melodischer Akzent zugeschrieben.[20] Neuere Untersuchungen sehen dagegen in der zeitlichen Erstreckung der prominenten Silbe das primäre akustische Merkmal des Akzents.[21] Unter bestimmten Bedingungen (zum Beispiel beim Flüstern) ist Prominenzsignalisierung durch Grundfrequenzvariation sogar unmöglich.[22] Auch im Italienischen wird der Akzent primär durch zeitliche Erstreckung realisiert.[23] Unabhängig davon, welchen Stellenwert die Grundfrequenzvariation für die Prominenzwahrnehmung hat, lassen meine Analysen die Feststellung zu, dass auf Akzenten in den allermeisten Fällen Grundfrequenzvariation stattfindet.

In der autosegmentalen Phonologie wird die Akzentstruktur auf der metrischen Ebene repräsentiert. Die metrische Analyse[24] ist notwendige Voraussetzung für die Analyse der Intonationsstruktur, weil nur Akzenten (und Grenzstellen) Töne zugewiesen werden.

---

[15] Vgl. Kohler (^{2}1995), S. 114-120.
[16] Vgl. Kap. 2.4.
[17] Schmidt (1986), S. 27 spricht deshalb vom „normativen Wortakzent".
[18] In der englischsprachigen Literatur wird deshalb im Anschluß an Bolinger (1972b), S. 22 zwischen realisiertem *accent* und abstraktem *stress* unterschieden, wobei *stress* notwendige Bedingung für *accent* ist. Vgl. Ladd (1996), S. 48f.; Uhmann (1991), S. 21f.; Bertinetto (1981), S. 50-53. Anders konzeptualisiert Cruttenden (^{2}1997), S. 13.
[19] Vgl. aber die Ausführungen zu den Wortakzentregeln in Kap. 2.3.2, S. 26ff.
[20] Vgl. Isačenko/Schädlich (^{2}1971), S. 20ff.
[21] Vgl. Dogil (1999), S. 291-299, bes. S. 292f. und Jessen et al. (1995), passim. Kontrovers dazu Möbius (1993), S. 10-16. Kohler (1991), S. 298-305 unterscheidet im „Kiel Intonation Model" einen auf zeitlicher Erstreckung basierenden von einem auf Grundfrequenzvariation basierenden Akzent. Siehe Kap. 3.2.3, S. 75ff.
[22] Vgl. Kohler (^{2}1995), S. 114.
[23] Vgl. Voghera (1992), S. 96 und den umfassenden Überblick in Bertinetto (1981), S. 41-90.
[24] Vgl. Kap. 2.3.2, S. 24ff.

Quantität ist eine prosodische Einheit, deren primäres auditives Merkmal die Dauer ist. Im Standarddeutschen können Daueruntersch iede als redundante Merkmale von Vokalqualitäten aufgefasst werden können.[25] Im Standaritalienischen und in verschiedenen deutschen Dialekten sind Quantitäten dagegen prosodische Einheiten, weil sie nicht nur die Vokale, sondern die gesamte Silbe betreffen.[26]

Quantitäten spielen in der vorliegenden Untersuchung nur eine untergeordnete Rolle in der metrischen Analyse. Sie werden auf der CV-Ebene wiedergegeben, die in den Analysen nicht systematisch berücksichtigt wird. Die Intonation wird direkt an die Silbenstruktur angebunden.

2.2 Silbe

2.2.1 Die Silbe als zentrale Analysekategorie

Die Bezugseinheit prosodischer Merkmale und Einheiten ist in dieser Arbeit die *Silbe*. Der Vorschlag der (linearen) generativen Phonologie der „Sound Pattern of English", die Silbe als Trägerin prosodischer Eigenschaften abzuschaffen, indem in die Merkmalskomplexe von Lautsegmenten auch abstrakte Akzentstärken (*stress*) integriert werden, ist nicht zuletzt von der späteren generativen Phonologie selbst zurückgewiesen worden.[27] Zahlreiche phonotaktische Regeln beziehen sich auf die Silbe. Auch die Regeln zu Akzentverteilung und Tonassoziation lassen sich mit Bezug auf die Silbe einfacher formulieren als mit Bezug nur auf Lautsegmente und morphologische Grenzsymbole.[28] Außerdem besitzt die Silbe für die meisten Muttersprachler mentale Realität, was sich darin ausdrückt, dass den meisten Muttersprachlern die Zerlegung von Äußerungen in Silben auch ohne linguistische Kenntnisse keine Schwierigkeiten bereitet.[29] Die Silbe ist also eine perzeptiv relevante Einheit und eine wichtige phonologische Kategorie. Im generativen Modell von Nespor und Vogel (1986, S. 11) ist die Silbe eine von sieben prosodischen Konstituenten in hierarchischer Stufung:

---

[25] Vgl. Schmidt (1986), S. 22-24.

[26] Bannert (1976), S. 25 belegt, dass die Domäne der Quantität im Mittelbairischen „aus der Sequenz von betontem Vokal und dem folgenden Konsonanten besteht." Es sind nur die kontrastierenden Sequenzen 'Langvokal + Kurzkonsonant' und 'Kurzvokal + Langkonsonant' möglich. Im Mittelfränkischen sind vokalische und konsonantische Dauer unter Tonakzenten komplementär, vgl. Schmidt (1986), S. 185-191. Zum Standarditalienischen vgl. Solari (1997), S. 225.

[27] Der Verzicht auf die Silbe ist auch in den „Sound Pattern of English" nicht vollkommen, weil Chomsky und Halle (1968), S. 354 für Vokale das Merkmal [+syllabic] anstelle von [+vocalic] ansetzen. Siehe dazu in der vorliegenden Arbeit auch S. 21f.

[28] Vgl. Auer (1994), S. 56.

[29] Vgl. Wiese (1996), S. 33. Selbst in der exotischen Sprache Fijian kommt „syllabic oral spelling" vor, vgl. Blevins (1995), S. 209f. Interessant ist auch das Ergebnis einer Untersuchung von Burani und Cafiero (1991), die für die Silbe im Italienischen eine Rolle bei der Erkennung graphisch repräsentierter Wörter nachweist.

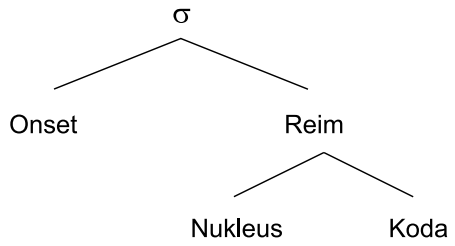

Abb. 2.2: Konstituentenstruktur der Silbe

> These seven units, from large to small, are: the phonological utterance (U), the intonational phrase (I), the phonological phrase (ϕ), the clitic group (C), the phonological word (ω), the foot (Σ) und the syllable (σ).[30]

Als phonetische Einheit ist die Silbe problematisch, weil Silbengrenzen exakte Korrelate weder in akustischer noch in artikulatorischer Hinsicht haben.[31]

Für die Silbe (symbolisiert durch $\gg\sigma\ll$) wird in der generativen Phonologie eine Konstituentenstruktur angenommen. Für Englisch, Deutsch und Italienisch gilt das in Abb. 2.2 dargestellte Modell.[32] Wiese (1996, S. 44) ergänzt dieses Modell um eine für das Deutsche gültige Skelett-Ebene. Die Annahme eines Silbenmodells mit einer

---

[30] Vgl. dazu auch Nespor (1999), S. 117-126. Die Konstituente *Fuß* lässt sich im Deutschen mit Gesetzmäßigkeiten in der Wortbildung und der Distribution von [ʔ] begründen, vgl. Wiese (1996), S. 56-61. Darüber hinaus ist der Fuß eine wichtige metrische Kategorie in den sog. akzentzählenden Sprache, siehe Kap. 2.2.2, S. 13. Das *phonologische Wort* – nach Auer (1994), S. 71 zusammengesetzt „aus dem Stamm, den Präfixen, den meisten Suffixen (ausgenommen *-heit*, *-keit*, *-bar*, *-lich* etc.) und den enklitischen Erweiterungen“ – wird als Domäne der Silbifizierung, zahlreicher phonologischer Regeln (z.B. der Assimilation) und des Wortakzents angesehen, vgl. Wiese (1996), S. 65ff. Auer (1994), S. 76 hält das phonologische Wort für die prosodische Hauptkategorie des Standarddeutschen. Die Bedeutung der *klitischen Gruppe* ist umstritten, die der *phonologischen Phrase* im Deutschen als Domäne für bestimmte Akzentverschiebungen sehr begrenzt. Vgl. Wiese (1996), S. 74-77 zur phonologischen Phrase im Deutschen, und Nespor (1985), passim zur phonologischen Phrase im Italienischen. Die *More* (μ) ist nach einer traditionellen Auffassung – vgl. Auer (1991), S. 10 – für das Standard-Neuhochdeutsche empirisch nicht gerechtfertigt und wird auch für das Italienische selten operationalisiert. In modernen autosegmentalen Ansätzen zeichnet aber eine Änderung dieser Auffassung ab. Van der Hulst (1999), S. 10-14 verwendet die More als universales Konzept zur Bestimmung des Silbengewichts. Peters (in Vorb.) unterscheidet durch Zerlegung von Silben in Moren unterschiedliche Realisierung der Fokusintonation in verschiedenen regionalen Varietäten des Deutschen.

[31] Vgl. dazu den Forschungsüberblick von Bertinetto (1981), S. 148-156. Heike (1992), S. 26 schreibt: „Zusammenfassend kann man sagen, daß die Silbe sich im (natürlichen oder simulierten) lautsprachlichen Produktionsprozess mit dem zeitlichen Erstreckungsbereich (ko-)artikulativer Steuerungsprozesse deckt [...], daß sie aber entgegen einer diskret-segmentalen Vorstellung keine scharfen Grenzen hat.“

[32] Graphik nach Eisenberg/Ramers/Vater (1992), S. V. Zum Englischen vgl. Blevins (1995), S. 213, zum Italienischen Nespor (1993), S. 156 und Mioni (1993), S. 128, zum Deutschen Vater (1992), S. 100.

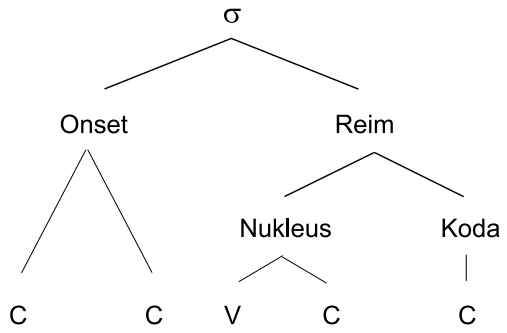

Abb. 2.3: Konstituentenstruktur der Silbe und CV-Struktur

Konstituentenstruktur – gegenüber einem „flachen“ Modell ohne Konstituenten – ist sinnvoll, weil sich viele phonotaktische und phonologische Regeln auf genau eine dieser Konstituenten beziehen (Abb. 2.3).[33] So sind zum Beispiel homorgane Obstruent-Sonorant-Cluster wie [pm], [fm], [kŋ] im Onset ausgeschlossen, während sie in der Koda als [mp], [mf], [ŋk] auftreten.[34] Nur im Reim findet dagegen die Vokalisierung von [ʀ] statt.[35] Regressive Nasalassimilationen scheinen auf die Koda beschränkt zu sein.[36]

Dagegen ist die Annahme einer Skala abnehmender Schallfülle oder *Sonorität*[37] vom Zentrum der Silbe (≫V≪ des Nukleus) zu ihren Rändern sowohl theoretisch als auch empirisch problematisch. Heike (1992, S. 9) sieht in der Konstruktion der Sonoritätshierarchie ein typisches Beispiel für zirkuläre Argumentation, weil einerseits behauptet wird, „die Sonoritätshierarchie steuere die phonotaktischen Gegebenheiten, und andererseits [...] die Beoachtungen an der Phonotaktik als Basis für die Formulierung des Sonoritätskonzepts angenommen“ werden. Heike weist auch darauf hin, dass der Vorwurf zirkulärer Argumentation bei der Bestimmung der Sonoritätshierarchie bereits von de Saussure (1972, S. 88-90) gegen Sievers ([2]1881) erhoben wurde.

Die Silbe ist eine auditive Einheit. In der Intonationsanalyse – im unten vorgeschlagenen autosegmentalen Modell ist die Silben-Ebene die zentrale Achse, mit der die Einträge auf der metrischen, der Ton- und der Segment-Ebene assoziiert werden – muss die Zahl der Silben einer Äußerung sicher bestimmt werden. Für das Italienische sieht Bertinetto (1981, S. 162) weder große Probleme bei der Bestimmung der Silbenzahl noch bei derjenigen der Silbengrenzen. Für die Bestimmung der Silbenzahl

---

[33] Die Symbole ≫C≪ und ≫V≪ bezeichnen Strukturstellen und sind nicht als Abkürzungen für ‘Konsonant’ und ‘Vokal’ misszuverstehen. Z.B. wird in diesem Modell ein langer Vokal oder Diphthong mit der Folge VC assoziiert, ein Beispiel findet sich unten, S. 13. Vgl. dazu Wiese (1996), S. 37-43.

[34] Vgl. Vater (1992), S. 105f., außerdem Wiese (1996), S. 44 und S. 234ff.

[35] Wiese (1996), S. 252-258.

[36] Vgl. Wiese (1996), S. 218-224; Vater (1992), S. 111f.

[37] Kohler ([2]1995), S. 73-75 nennt folgende Klassen von Lauten mit abnehmender Sonorität: offene Vokoide, geschlossene Vokoide, Liquide, Nasale, stimmhafte Frikative, stimmlose Frikative, Plosive. Vgl. zur Sonoritätshierarchie auch Wiese (1996), S. 258-261.

seien lediglich Laute, die bei schnellem Sprechtempo zwischen [i] und [j] bzw. zwischen [u] und [w] oszillieren, problematisch.[38] Für die Bestimmung der Silbengrenzen zeigt ein Perzeptionsexperiment differierende Urteile von Informanten nur bei Clustern aus [s] oder [z] und Konsonant bzw. bei [tm] und [tl].[39] Auch im Deutschen ist die Bestimmung der Silbenzahl nur dort problematisch, wo [r]- oder [i]-Glides auftreten, außerdem in Silben, die [ə] als Kern haben und in der fließenden Rede oft zu den silbischen Konsonanten [m̩], [n̩], [l̩] und [ʀ̩] reduziert werden.[40] Die Silbengrenzen sind im Deutschen mit seinen komplizierten Konsonantenclustern dagegen in vielen Fällen nur sehr schwer zu bestimmen.[41] In der phonologischen Theorie lassen sich Bestimmungsschwierigkeiten jedoch überwinden: Beispielsweise wird das Problem der Festlegung der Silbengrenze nach akzentuiertem Kurzvokal durch die Annahme *ambisilbischer* Konsonanten gelöst.[42]

Die unterschiedlichen Schwierigkeitsgrade in der Bestimmung von Silbenzahl und Silbengrenzen im Deutschen und Italienischen werden üblicherweise mit der unterschiedlichen typologischen Einordnung der Sprachen als silbenzählend (Italienisch) und akzentzählend (Deutsch) erklärt.[43] Die italienische Silbe hat in der Skelett-Ebene eine starke Tendenz zu CV.[44] Untersuchungen zeigen, dass 55% aller italienischen Silben die CV-Skelettstruktur aufweisen.[45] Die maximal möglichen dreifachen Konsonantencluster wie CCCV in [stra] sind im Onset sehr selten, in der Koda kommen Zwei- und Dreifachcluster – CVCC in /ˈgɔlf/ oder CVCCC in /ˈfɪlms/ – ausschließlich in Fremdwörtern und Neologismen vor.[46] Assimilations- und Reduktionsprozesse verstärken im Italienischen diese Tendenz. Der potentiell viersilbigen Form *lo ha detto* mit der Skelettstruktur CVVCVCCV entspricht gesprochen und geschrieben die dreisilbige Standardrealisierung *l'ha detto* mit CVCVCCV. Das rhythmisch ungünstige Zusammentreffen zweier V wird vermieden. Dagegen haben Assimilations- und Reduktionsprozesse im Deutschen oft gegenläufige Tendenz: Sie verstärken die ohnehin starke Neigung zu komplizierten Konsonantenclustern. In der Skelett-Struktur sind im Onset Dreifachcluster wie CCCVCC in /ˈʃtroːm/ und in der Koda sogar Fünffachcluster vom Typ CVCCCCC in /ˈhɛʀpsts/ möglich.[47] Die Tilgung von [ə], die in der gesprochenen Sprache häufig auftritt, aber auch im Konjugationsparadigma bestimmter Verben vorgesehen ist, führt dazu, dass zum Beispiel das Wort *Barren* mit der

[38] Vgl. Bertinetto (1981), S. 158.
[39] Vgl. Bertinetto (1981), S. 161-163.
[40] Vgl. Kohler (1996), passim; Wiese (1996), S. 49-51; Auer (1994), S. 68; Vater (1992), S. 106-110; Bertinetto (1981), S. 157f.
[41] Zur Phonotaktik des Deutschen vgl. Kohler (^{2}1995), S. 175-186.
[42] Vgl. dazu Ramers (1992). Wiese (1996), S. 51 ist deshalb der Meinung, „that very few clear cases for non-predictable syllabification are available" und skizziert deshalb ein (vorläufiges) Regelwerk der Silbifizierung im Deutschen (S. 51-56).
[43] Siehe Kap. 2.2.2.
[44] Was nach Prince/Smolensky (1993), S. 89 die Struktur von „universally optimal syllables" und in allen Sprachen der Welt möglich ist.
[45] Bortolini (1976), S. 12. Vgl. auch Auer/Uhmann (1988), S. 244-249.
[46] Vgl. Bortolini (1976), S. 8ff.
[47] Wieses Modell sieht einen solchen Fünffachcluster nicht vor. Das Wort /ˈhɛʀpst/ wird in eine Silbe mit der Struktur CVCC und zwei von einem extrasilbischen C abhängende Konsonanten zerlegt. Vgl. Wiese (1996), S. 47-49.

potentiellen Skelett-Struktur CVCVC in der Regel als /ˈbaʀn̩/ mit CVCC realisiert wird.

Die Fehlerforschung spiegelt die Relevanz der Unterschiede in der Skelett-Ebene von Deutsch und Italienisch wider. Italienische Deutschlerner tendieren zum „Segmentieren“ langer oder unbekannter Konsonantencluster durch Sprossvokalbildung. Deutsch /ˈpfanə/ mit der im Italienischen unbekannten Affrikate [pf] und der Struktur CCVCV wird zu /pəˈfannə/ mit CVCVCCV. Der im Italienischen seltene konsonantische Wortauslaut wird oft durch Tilgung des finalen C oder Ergänzung von [ə] beseitigt. Aus dem Infinitiv /ˈvapnən/ mit CVCCVC wird /ˈvapnə/ mit CVCCV; das Substantiv /ˈboːt/ mit CVCC wird zu /ˈboːtə/ mit CVCCV.[48] In beiden Fällen wird die Skelettstruktur der im Italienischen idealen CV-Alternation angenähert.

2.2.2 Die Isochronie-Hypothese

In der *Isochronie-Hypothese* wird die Beobachtung formuliert, dass der Mensch dazu tendiert, akustische Eindrücke in rhythmischen Gruppen zusammenzufassen. Die Tendenz ist relativ unabhängig von der Genauigkeit der Folge der akustischen Stimuli (beispielsweise in der musikalischen Aufführung), und sie kann sogar dort beobachtet werden, wo akustisch keine rhythmische Gliederung nachweisbar ist (zum Beispiel kann auch das gleichmäßige Ticken der Uhr als rhythmisch gegliedert perzipiert werden). Obwohl die Isochronie-Hypothese nicht bewiesen ist, kann man mit ihr unter bestimmten Bedingungen sinnvoll operieren.

In der sprachwissenschaftlichen Tradition werden zwei Typen von Isochronie und entsprechend zwei Typen von Sprachen unterschieden: silbenzählende Sprachen (Sprachen mit *isocronia sillabica*) und akzentzählende Sprachen (Sprachen mit *isocronia accentuale*).[49] Standarditalienisch wird weithin als Prototyp einer silbenzählenden Sprache genannt, während Deutsch und Englisch als akzentzählende Sprachen gelten. Konkret lässt sich die Isochronie-Hypothese wie folgt formulieren: Silbenzählende Sprachen erhalten die Isochronie ihrer Silben, das heißt, die Silbendauer ist unabhängig von der Akzentverteilung. Akzentzählende Sprachen erhalten dagegen die Isochronie ihrer Füße,[50] das heißt, die Silbendauer nimmt mit zunehmender Silbenanzahl zwischen zwei Iktus ab.[51] Gemäß dieser These müsste eine zweisilbige italienische Äußerung exakt die doppelte Dauer einer einsilbigen haben, eine dreisilbige exakt die dreifache. Im Deutschen wären solche Vorhersagen nicht auf der Ebene der Silben, aber auf derjenigen der Füße möglich: Der Abstand zwischen zwei Iktus müsste konstant bleiben, unabhängig davon, ob zwischen ihnen eine, zwei oder drei unbetonte Silben liegen. Auer und Uhmann (1988, S. 220-237) geben einen Überblick über Überprüfungen der Hypothese hinsichtlich verschiedener, auch nicht-indogermanischer Sprachen und stel-

---

[48] Beispiele aus Zuanelli Sonino (1975), S. 93 und S. 99. Die Notierung des langen Vokals [oː] durch VC auf der Skelett-Ebene folgt dem Vorschlag von Wiese (1996), S. 38f.

[49] Vgl. Bertinetto/Magno Caldognetto (1993), S. 147ff.

[50] Wiese (1996), S. 56 definiert *Fuß* wie folgt: „[...] the foot consists of the string of syllables starting from one stressed syllable up to (but not including) the next one.“ Damit ist *Fuß* synonym mit *Takt* in der Definition von Pheby (1981), S. 852.

[51] Vgl. Auer/Uhmann (1988), S. 217. Zu *Iktus* vgl. Pheby (1981), S. 852ff.

len fest, dass die Isochronie-Hypothese phonetisch nicht haltbar ist. Durch Messungen lassen sich die postulierte *isocronia sillabica* des Italienischen und *isocronia accentuale* des Deutschen nicht beweisen.[52] Exakte Messungen der Silben- und Fußdauer scheitern am Fehlen fester Messpunkte. Artikulatorisch oder akustisch sind Silbengrenzen als potentielle Messpunkte in der fließenden Rede nicht zu bestimmen. In Fällen, wo Testsilben unter experimentellen Verhältnissen deutlich voneinander abgegrenzt sind, hat man eine nicht zu quantifizierende Differenz zwischen der akustischen und der auditiven Silbengrenze festgestellt.[53]

Die Isochronie-Hypothese darf dennoch nicht einfach verworfen werden: Es gibt unbestreitbare Unterschiede in der Wahrnehmung der rhythmischen Struktur von Deutsch und Italienisch. Oben (S. 11f.) wurde ausgeführt, dass sich Silbengrenzen im Italienischen leichter bestimmen lassen als im Deutschen. Die komplizierte Konsonantenclusterung und die starke Tendenz zur Vokalreduktion sorgen im Deutschen dafür, die Silbengrenzen zu verwischen und den Eindruck zu erwecken, dass sich die Äußerung allein zwischen den Iktus strukturiert. Dieses Intervall wird aufgrund der Disposition des Menschen zur rhythmischen Gliederung seiner Wahrnehmungseindrücke als isochron perzipiert. Im Italienischen wird die Deutlichkeit der Silbengrenzen von Assimilations- und Reduktionsprozessen nicht verringert. Diese Prozesse unterstützen noch die Tendenz der Silben zur CV-Struktur. Auf dieser Grundlage fällt es leicht, alle Silben einer italienischen Äußerung als isochron wahrzunehmen.[54]

Eine neuere Untersuchung von Vékás und Bertinetto (1991) zieht die Konsequenzen aus den wenig erfolgreichen Versuchen, die Isochronie-Hypothese durch Zeitmessungen zu stützen. Die Autoren schlagen vor, die auf die gemessene Zeit aufbauende Dichotomie von *isocronia accentuale* (IA) und *isocronia sillabica* (IS) aufzugeben. Stattdessen werden Sprachen, die wie das Italienische versuchen, auch in fließender Rede die phonetischen Merkmale der einzelnen Laute möglichst intakt zu halten (Strategie des *controllo locale*), von solchen unterschieden, die wie das Deutsche stark zur Koartikulations-, Assimilations- und Reduktionsprozessen neigen (Strategie der *compensazione*):

> È infatti ipotizzabile che le lingue ad IA siano quelle che attuano, a tutti i livelli della struttura prosodica, una strategia di 'compensazione', imputabile ad una maggior flessibilità articolatoria. Per converso, le lingue ad IS sarebbero quelle in cui la produzione dei singoli elementi fonematici tende ad avvenire nel quadro di una strategia di 'controllo locale', mirante a conservare il più possibile intatte le caratteristiche acustiche e prosodiche delle entità che compongono la catena fonica.[55]

Dass die Wahrnehmung der Silbe als Grundbaustein der Äußerung tatsächlich von der Kontrolle der phonetischen Merkmale abhängt, versuchen Vékás und Bertinetto durch Messungen der Frequenz des zweiten Formanten der Laute [i] und [l] in Non-

[52] Auer und Uhmann (1988), S. 229 interpretieren dabei hinsichtlich des Italienischen die Daten eines Experiments von Bertinetto (1981), S. 266, mit denen dieser allerdings die Tendenz des Italienischen zur *isocronia sillabica* zu beweisen versucht. Vgl. Bertinetto (1981), S. 183ff.

[53] Zum Problem des sog. *P-Centers* siehe Auer/Uhmann (1988), S. 241-243.

[54] Vgl. Nespor (1993), S. 260.

[55] Vékás/Bertinetto (1991), S. 155.

senswörtern verschiedener Länge zu belegen. Dieselben Nonsenswörter werden von deutschen, englischen und italienischen Informanten produziert. Es zeigt sich, dass der zweite Formant bei den italienischen Informanten relativ konstant bleibt, während er sich bei den deutschen und englischen Informanten in Abhängigkeit von der lautlichen Umgebung verändert.[56] Dadurch scheint zumindest die unterschiedliche Tendenz der Sprachen – Italienisch tendiert zur lokalen Kontrolle, Deutsch zur Kompensation – belegt zu sein.

2.3 Intonationsmodelle

Intonation ist die prosodische Erscheinung, die primär auf dem auditiven Merkmal Tonhöhe beruht. Es hat seit Helmholtz (1870) oder Sievers (1876) viele Versuche gegeben, den Tonhöhenverlauf entsprechend der durch ihn ausgedrückten Funktionen[57] zu schematisieren. Die heutige Forschungssituation ist im Wesentlichen von der Konkurrenz zweier großer Richtungen geprägt: Die sog. „Britische Schule“ modelliert Tonhöhenverläufe als *Tonhöhenbewegungen* (*configurations*), die eher amerikanisch geprägte autosegmentale Phonologie dagegen als Folge von *Tonhöhenstufen* (*levels*). Bolinger (1951) hatte mit seiner Kritik an den Intonationsmodellen des amerikanischen Strukturalismus – Pike (1945), Wells (1945), Trager/Smith (1951) –, deren vier Tonhöhenstufen empirischen Distinktivitätstests nicht standhielten, die „levels-vs.-configurations controversy“ angestoßen.[58] Ich werde in Kap. 2.3.3 ausführen, dass dieser Gegensatz beim heutigen Stand der Modellbildung kaum noch eine Rolle spielt, dass sich die Modelle hinsichtlich der Bewertung der Akzenttöne aber substantiell unterscheiden.

Am niederländischen Instituut voor Perceptie Onderzoek (IP0) wurde ein Modell zur Schematisierung von Tonhöhenverläufen entwickelt, das sich von den in den beiden nächsten Unterkapiteln dargestellten dadurch unterscheidet, dass es keinen phonologischen Charakter hat. Die Arbeiten der „Holländischen Schule“ werden ausführlich in Kap. 2.5 vorgestellt.[59] Das stark experimentalphonetisch ausgerichtete „Kiel Intonation Model“ wird in Kap. 3.2.3 behandelt.[60]

Die Beschreibungskategorien der vorliegenden Studie (Kap. 2.6.1) werden aus dem Modell der autosegmentalen Phonologie und dem Notationssystem ToBI entwickelt. Ausschlaggebend für diese Entscheidung war neben den modellimmanenten Vorzügen, die im Folgenden deutlich werden, die Tatsache, dass ToBI die für eine kontrastive Studie unverzichtbare internationale Verständlichkeit sicherstellt – im Gegensatz zu den in den Einzelsprachen erheblich differierenden Modellen nach dem Vorbild der Britischen Schule.

---

[56] Vgl. Vékás/Bertinetto (1991), S. 158.
[57] *Funktion* und *Bedeutung* werden synonym verwendet.
[58] Vgl. Ladd (1996), S. 60f.
[59] Zum IPO-Intonationsmodell siehe Kap. 2.5.4, S. 49f.
[60] Zu Modellen für die Synthese von Intonationskonturen vgl. Avesani (1999).