

Klein, W. (1980). Der Stand der Forschung zur deutschen Satzintonation.
Linguistische Berichte 68/80, 3-33.

Der Stand der Forschung zur deutschen Satzintonation

Wolfgang Klein

Der Stand der Forschung zur deutschen Satzintonation gibt zum Rühmen keinen Anlaß. Sieht man von kleineren Untersuchungen zu Einzelproblemen ab, so stützen sich die meisten Beschreibungen, etwa jene in den gängigen Grammatiken, auf von Essen (1956, 2. Aufl. 1964) - soweit sie nicht überhaupt völlig impressionistisch sind. Davon gibt es zwei Ausnahmen, die allerdings den Charakter erster Studien haben, nämlich ISAČENKO-SCHÄDLICH (1966, 1970) und KOHLER (1977). Eine dritte Ausnahme ist BIERWISCH (1966), der sich zwar auch auf von Essens Material stützt, es aber in einen anderen Rahmen integriert. ISAČENKO und SCHÄDLICH haben demgegenüber einen völligen Neuansatz, der jedoch in der Folge weder von ihnen noch von anderen weitergeführt wurde; KOHLER überträgt HALLIDAYS System zur Beschreibung der englischen Intonation aufs Deutsche.

Was in den allgemein verbreiteten deutschen Grammatiken steht - soweit sie überhaupt die Intonation behandeln - sind weithin Paraphrasen, Modifikationen und impressionistische Ergänzungen der Arbeit von Essen. Dies ist allein schon darum bemerkenswert, weil dieses Büchlein vorwiegend didaktischen Zwecken diene und keineswegs als eine systematische wissenschaftliche Grundlagenstudie gedacht war. Daher kann von Essen zu Recht vieles beiseitelassen und mehr auf praktische Brauchbarkeit als auf theoretische wie empirische Fundierung achten; in diesem Verständnis ist seine Darstellung auch wertvoll. Eigentümlicherweise ist sie dann aber selbst zur theoretischen und empirischen Grundlage für andere geworden.

Im folgenden will ich versuchen, einen kritischen Überblick über die wichtigsten Arbeiten zu geben. In Abschnitt 1 wird von Essen diskutiert; in den Abschnitten 2 und 3 werden zwei dazu parallele Arbeiten besprochen, nämlich die Darstellung in der „Dudengrammatik“ (von Christian WINKLER), die der von Essens verwandt ist, und die in der Kleinen Enzyklopädie „Die deutsche Sprache“ (von Eberhard STOCK), die sich weiterhin auf von Essen stützte. Im vierten Abschnitt gehe ich auf die knappe Beschreibung in KOHLER (1977) ein; im fünften wird der Ansatz von ISAČENKO und SCHÄDLICH besprochen, und der sechste ist der Arbeit von Manfred BIERWISCH gewidmet. Der siebte sollte eigentlich die Klage zur Lage enthalten, aber es schien mir dann ermutigender, ihn zu einer kleinen Aufgabe für den intelligenten Leser zu nutzen.

Noch zwei Vorbemerkungen: Das Wort „Intonation“ wird verschieden gebraucht, und es ist unmöglich, die unterschiedlichen Verwendungsweisen zu erläutern, ohne die jeweiligen Theorien zu erklären; dies gilt auch für viele andere Wörter wie „Akzent“, „Melodie“, „betont“ u.a.; da ich nicht noch eine eigene terminologische Regelung hinzufügen will, verwende ich diese Wörter hier — sofern nicht auf eine spezielle Betrachtungsweise Bezug genommen wird — in einem laxen, umgangssprachlichen Sinn. In diesem Sinn befaßt sich die Intonationsforschung mit phonetischen Eigenschaften wie Tonhöhe, Lautstärke, Dauer von Segmenten (einschließlich Pausen), soweit sie sprachlich relevant sind.

Es ist in diesem Rahmen unmöglich, eine Einführung in die phonetischen Grundlagen der Intonation zu geben; ich verweise auf die vielen Einführungen in die Phonetik und speziell zu den prosodischen Eigenschaften auf FRY (1968), CRYSTAL (1969), Kap. 1-3, LEHISTE (1970) und LASS (1976), Kap. 1 und 7.

1 von Essen (1956, 2. Auflage 1964¹)

Von Essens Buch hat drei Teile, nämlich die Darstellung der Intonationsregeln selbst (S. 11—66), eine kleine Sammlung von Texten, deren Intonationsverlauf er transkribiert und interpretiert (S. 67—89) sowie Übungen (S. 91 -118).

Es ist nicht klar, woher von Essen seine Regeln nimmt. Ältere systematische Darstellung gibt es nicht; er selbst führt KUHLMANN (1931) — über den Aussagesatz — und HERMANN (1942) — über den Fragesatz — an. Über seine empirische Grundlage sagt es bloß: „Die folgenden rhythmisch-melodischen Aufzeichnungen wurden nach mitgeschnittenen Wortsendungen des Rundfunks (NDR) und nach eigenen Tonbandaufzeichnungen genommen. Aus diesen und vielen anderen Transkriptionen sind die im ersten Teil mitgeteilten Intonationsregeln abgeleitet.“ (S. 69) Wie diese Transkriptionen in von Essens einfachem Notationssystem zustande gekommen sind, wird nicht weiter ausgeführt² (was auch in einem didaktischen Werk durchaus am Platz sein mag).

Von Essen befaßt sich nur mit dem Tonhöhenverlauf (der *Melodie*) nicht mit der Dauer und nur indirekt mit der Intensität. Der Sprecher ist relativ frei in der Melodiegestaltung „zu einer Kundgabe der Stimmung, der Gefühle und Gemütsbewegungen“ (S. 12/13). Es gibt jedoch, so meint er, in jeder Sprechergemeinschaft verfestigte Muster des Tonhöhenverlaufs; ein solches Muster nennt er *Sprachmelodie* oder *Intonation*. Grundeinheit ihrer Beschreibung ist nicht der Satz (bzw. sein gesprochenes Gegenstück, der Ausspruch), sondern das *rhetorische Syntagma*. Leider ist dieser Begriff nicht eben klar definiert: „Längere Sätze gliedern sich in zwei oder noch mehr Sinnabschnitte auf. Diese Abschnitte entsprechen gewissen Vorstellungseinheiten oder gedanklichen Teilabläufen. Vielfach fallen sie mit den syntaktischen Einheiten des Satzes zusammen, oft aber umfassen sie auch mehr als einen Satzteil; es kommt in jedem Fall darauf an, was der Sprechende gerade als Sinneinheit zusammenfassen will.“ (S. 33). Diese Gliederung zeigt er durch Pausen,

vor allem aber durch die Melodie an: „Der Teil des Ausspruchs, der durch ein melodisches *Motiv* als eine Einheit gekennzeichnet wird, soll rhetorisches *Syntagma* genannt werden.“ (S. 33). Da nun aber über das Vorhandensein eines bestimmten melodischen Motives (z.B. die „progrediente Inonation“ - wir kommen noch darauf) nicht entschieden werden kann, wenn die Grenzen des rhetorischen Syntagmas nicht klar sind, wird die Bestimmung zirkulär; wir werden dies später noch sehen. Bei einfachen Sätzen fällt nach von Essen das (rhetorische) Syntagma mit dem Satz zusammen. Der Einfachheit halber will ich im folgenden annehmen, daß dieser Fall immer gegeben ist.

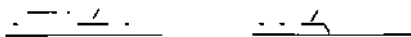
Das Syntagma besteht aus drei Teilen: dem *Vorlauf*, dem *rhythmischen Körper* und dem *Nachlauf*: „Was vor der ersten Hervorhebung steht, ist Vorlauf, was der letzten (...) folgt, ist Nachlauf; was zwischen Vor- und Nachlauf steht, können wir den rhythmischen Körper nennen.“ (S. 28). Was aber ist nun eine Hervorhebung (oder „hervorgehobene Silbe“)? Von Essen sagt dazu: „Jeder Satz enthält mindestens einen Begriff, auf den der Sprechende die Aufmerksamkeit des Angesprochenen besonders zu lenken beabsichtigt. Das Wort, das diesen Hauptbegriff bezeichnen soll, tritt dementsprechend in geeigneter Weise aus den übrigen Teilen der Rede hervor. Das betreffende Wort bildet eine *Hervorhebung*.“ (S. 13). Sie kann durch „Veränderung der Dauer, des Atemdrucks, der Tonhöhe, oder durch alles zusammen (erfolgen). Das wichtigste und wirksamste Mittel aber ist die Tonhöhenveränderung.“ (S. 14). Sie trifft lediglich die Hebungs- oder Akzentsilbe eines Wortes; was dies ist, wird nicht (in phonetischen Begriffen) definiert, aber gemeint sind offenbar druckstarke im Vergleich zu druckschwachen Silben. Sind mehrere Akzentsilben in einem Wort, so gilt: „Die druckstärkste ist die *Hauptakzentsilbe*, andere Hebungssilben sind *Nebenakzentsilben*,“ (S. 14). Auf sie fällt gegebenenfalls die Hervorhebung. Nicht alle Akzentsilben sind jedoch hervorgehoben, sondern nur die „sinnwichtigsten“. Leider sind nun auch nicht alle Akzentsilben druckstark: „Wie aus allen diesen Beispielen ersichtlich, ist nicht jede von der Sprache her als Akzentsilbe geforderte Silbe als druckstark verzeichnet. Natürlich haben die akzentuierten Silben immer mehr Druck als die nichtakzentuierten³); es kommt hier aber nur auf die aus Gründen des Sinns besonders hervorgehobenen Redeteile an. Welche Silben das im einzelnen Fall sind, liegt beim Sprechenden.“ (S. 29). Der Zusammenhang dieser verschiedenen Begriffe ist mir nicht völlig klar, zumal von Essen noch einen weiteren einführt: „Druckschwächere Silben sind immer nebensächlich behandelte Silben. Sie fallen auch melodisch nicht auf⁴). Das Ohr folgt im wesentlichen den Tönen der hervorgehobenen Silben. Aus ihnen gestaltet der Hörer die phrasierende Melodie, das melodische *Motiv*. Die Hebungstöne „führen“ den mitgestaltenden Hörer; sie wurden darum von Ed. SIEVERS (...) als *Führtöne* bezeichnet.“ (S. 28). Von Essen übernimmt diesen Ausdruck. Wir haben deshalb mindestens die folgenden Begriffe:

1. Akzentsilbe; sie bezieht sich auf das Wort und ist druckstärker als die anderen Silben des Wortes; es kann in einem Wort auch drei Druckstufen geben; man spricht dann von Hauptakzentsilbe, Nebenakzentsilbe und nichtakzentierter

- Silbe. Die Druckverteilung ist „von der Sprache gefordert“. Akustisch entspricht dem „Akzent“ also der Schalldruck, auditiv damit die Lautstärke.⁵⁾
2. Hebungssilbe; das ist jene Silbe, die von einer Hervorhebung getroffen werden kann, d.h. die Akzentsilbe oder - falls mehrere vorhanden sind — die Hauptakzentsilbe eines Wortes.
 3. Hervorhebung; damit ist entweder das Hervorheben oder aber die dadurch hervorgehobene Silbe gemeint. Hervorgehoben wird durch Veränderung des Drucks, der Dauer oder — und dies ist das wichtigste Mittel — der Tonhöhe. Wie dies geschieht, wird leider nicht gesagt — ob z.B. der Schalldruck verdoppelt oder verzehnfacht wird, wie hoch oder auch tief der Ton zu steigen oder zu fallen hat, ob die Veränderung in der Silbe erfolgt oder zwischen Silben, usw. Eindeutig ist jedoch, daß die Hervorhebung völlig im Ermessen des Sprechers steht (im Rahmen natürlich der Hebungssilben).⁶⁾
 4. Führtöne; das sind offenbar die tatsächlich hervorgehobenen Silben, auch wenn sie in dem zuletzt angeführten Zitat *auch* mit den „Hebungstönen“ gleichgesetzt werden; offenbar entspricht „Hebungston“ nicht „Hebungssilbe“, d.h. hervorhebbarer Silbe, sondern „hervorgehobener Silbe“.

Auf die Richtigkeit dieser Erläuterungen möchte ich nicht schwören, aber sie scheint mir am besten mit der Verwendung der verschiedenen Termini im Einklang zu stehen.

Bevor ich nun zu den Regeln der Intonation komme, noch kurz zu von Essens Transkriptionssystem. Er notiert die Tonhöhe relativ zu einem „Normal-Null“; sie ist die individuell verschiedene „völlig spannungslose Tiefe der Stimme. Sie wird in den Melodieaufzeichnungen durch eine Linie angedeutet. Höhere Töne⁷⁾ werden in kleineren oder größeren Abständen darüber notiert. Die Töne druckschwacher Silben werden durch Punkte, die durch Druck hervorgehobenen durch Striche dargestellt. Der Strich für die auffallendste der druckstarken Silben bekommt ein Akzentzeichen (Akzent). Ein Hinauf- oder Herunterschleifen des Stimmtones innerhalb einer Silbe wird durch aufwärts- bzw. abwärtsgeschwungenen Bogen angedeutet, z.B.



Vier Höhenstufen (tief, gehoben-tief, mittel, hoch) werden unterschieden, allerdings nicht weiter begründet und kaum in den Regeln verwendet. Vorlauf und Nachlauf bestehen also aus nicht hervorgehobenen Silben; dies können durchaus Akzentsilben sein. Es liegt völlig im Ermessen des Sprechenden, ob er in der Tat eine „von der Sprache her als Akzentsilbe geforderte Silbe als druckstark“ (S. 29) verwirklicht.

Die Vorlaufsilben fallen im allgemeinen leicht ab; das muß aber nicht sein: „Oft wird man sogar ansteigende Silbenfolgen bemerken; manchmal werden sämtliche Vorlaufsilben tief, manchmal hohtonig genommen (...), zuweilen aufsteigend, in anderen Fällen absteigend, je nach den emotionalen Triebkräften.“ (S. 24). Dies

scheint mir sehr verwirrend; wenn das wichtigste Mittel der Hervorhebung (nach S. 14) die Tonhöhenveränderung ist, wieso kann dann der Verlauf hohe Töne, tiefe Töne und Wechsel haben, ohne Hervorhebungen zu enthalten? Der einzige Ausweg scheint mir, daß Vorlauf, Nachlauf und (rhythmischer) Körper lediglich durch Druck-Hervorhebung (und deren Fehlen) gekennzeichnet sind; man muß dann Unabhängigkeit der Hervorhebung durch Druck und durch Tonhöhe annehmen; das ist — als Regelfall zumindest — wenig plausibel, aber nur so sind hohe Töne im Vorlauf möglich, ohne daß ein Widerspruch entsteht.

Der (rhythmische) Körper besteht aus einer beliebigen Anzahl von Hervorhebungen (mindestens einer) — und zwar Druck-Hervorhebungen: „Die Töne der hervorgehobenen Silben folgen einander in stufenweisem Abstieg.“ (S. 26)⁸) Der letzten Hervorhebung kommt besonderes Gewicht zu; unabhängig von ihrer Höhe gilt sie als *Schwerpunkt* des Satzes (bzw. des Syntagmas). Die zwischen den Hervorhebungen stehenden Silben sind unwichtig: „sie verfolgen (...) einen stufenweisen Abstieg, schmiegen sich entweder dem vorangehenden Führton enklitisch an oder streben proklitisch auf den folgenden zu. Oft fallen sie aber auch ein wenig nach unten, selten nach oben aus der Führtonrichtung heraus; aber wenn sie es tun, dann nur in kleinen Intervallen.“ (S. 28).

Nach der Schwerpunktsilbe folgt der Nachlauf; er besteht immer aus druckschwachen Silben (d.h. nicht hervorgehobenen oder gar nicht hervorhebbaren); von Essen unterscheidet zwei mögliche Verläufe:

1. Terminal: die Nachlaufsilben sind gleichmäßig tieftönig.
2. Interrogativ: die Nachlaufsilben steigen gleichmäßig an; die letzte hat gewöhnlich noch einmal eine „Aufwärtsschleife“.

Falls die Schwerpunktsilbe letzte Silbe überhaupt, der Nachlauf also leer ist, hat sie einen Abfall oder Aufstieg in sich selbst. Typ 1 „wird angewandt in Aussagen, Aufforderungen, Ausrufen, vorangestellten Anreden, Ergänzungsfragen, indirekten Reden, den zweiten Gliedern von Doppelfragen.“ (S. 65). Typ 2 „wird angewandt in Entscheidungsfragen, Nachfragen (...), ferner — als Variante — in Aussagen, Aufforderungen und Ergänzungsfragen, wenn sie Warnungen oder Drohungen darstellen sollen (...)“ (S. 66). Im Gegensatz zu von Essens eigenen Behauptungen entscheiden also gerade die druckschwachen Silben über den Intonationstyp; von den Hervorhebungen ist allenfalls die letzte wichtig, und auch dies nur, wenn eben keine druckschwachen Silben da sind.

Wir sind bis jetzt davon ausgegangen, daß einem Ausspruch nur *ein* rhetorisches Syntagma entspricht. Dies gilt aber nur in einfachen Sätzen. Kompliziertere (und auch einfache, wenn dies dem Gestaltungswillen des Sprechers entspricht) bestehen aus mehreren Syntagmen. In diesem Fall hat nur das letzte Syntagma einen Schwerpunkt, d.h. die letzte Hebung der anderen zählt nicht als Schwerpunkt, und nur der Nachlauf der letzten hat terminale oder interrogative Intonation; der Nachlauf der übrigen bleibt auf einer mittleren Höhe; diese rhetorischen Syntagmen haben eine *progređiente* Intonation. Ein ganzer Ausspruch besteht also aus einem

terminalen oder interrogativen Syntagma, dem beliebig viele progrediente vorangehen.⁹⁾ Die Grenze zwischen den Syntagmen wird als *Neuansatz* bezeichnet. Dies ist, von einigen Marginalien abgesehen, die gesamte Beschreibung der deutschen Intonation. Sie gibt zu einer großen Anzahl von Einwänden Anlaß. Ich will mit der Kritik nicht zu sehr ins Einzelne gehen, sondern einige Hauptpunkte herausgreifen.

1. Fällt die Melodie gegen Ende eines Ausspruchs, so handelt es sich um eine Aussage oder einen Ausruf oder eine vorangestellte Anrede oder eine Ergänzungsfrage oder indirekte Rede oder das zweite Glied einer Doppelfrage; steigt sie, so handelt es sich um eine Entscheidungsfrage oder eine Nachfrage oder ...; bleibt sie mittelhoch, so folgt noch etwas. Dies ist *alles*, was über Intonationsformen, soweit sie Bestandteil der deutschen Grammatik sind, gesagt wird.¹⁰⁾ Das ist nicht eben viel.
2. Alle sonstigen Gestaltungen der Satzmelodie hängen von „emotionellen Triebkräften“, dem Gestaltungswillen des Sprechers, der Art, wie er Sinnvorstellungen zusammenfaßt, usw. ab. Sieht man einmal davon ab, daß diese Begriffe noch klärungsbedürftig sind, so muß es auch dafür irgendwelche Regelmäßigkeiten geben, denn sonst könnte kein Hörer nachvollziehen, was der Sprechende mit seiner Gestaltung zum Ausdruck bringt. Was aber sind diese Regeln?
3. Es wird keinerlei Zusammenhang zwischen syntaktischem Aufbau der Äußerung und Intonation hergestellt. Von Essens Beispielanalysen in Teil II machen deutlich, daß er durchaus die Syntax berücksichtigt; aber dies kommt in seinen Regeln nicht zum Ausdruck. Deshalb werden zahlreiche wichtige Fakten der deutschen Intonation übersehen. Ich greife zwei heraus:

(a) Wenn man in dem Satz

(1) Fritz schrieb ein Buch (— · /)

bei „normaler“ Intonation (sie ist in Klammern angegeben)¹¹⁾ ein druckschwaches „auch“ vor das Objekt setzt, besagt der entstandene Satz

(2) Fritz schrieb auch ein Buch (- · · /),

daß Fritz nicht nur z.B. ein Kind gezeugt oder einen Baum gepflanzt, sondern eben auch ein Buch geschrieben hat. Hebt man nun aber „auch“ hervor, so besagt der Satz

(3) Fritz schrieb auch ein Buch (— · / ·),

daß nicht nur irgendjemand sonst, sondern eben auch Fritz ein Buch geschrieben hat. Nach der Theorie ist nun „auch“ Schwerpunkt und sinnwichtigstes Wort. Das will ich einmal so zugestehen, obwohl es mit etwas merkwürdig vorkommt. Denselben Sinn hat aber der folgende Satz:

(4) Auch Fritz schrieb ein Buch (· / · · ·).

In diesem sinn gleichen Satz ist nun plötzlich „Fritz“ Schwerpunkt und sinnwichtigstes Wort. Wie immer es sich mit der Festlegung des „Schwerpunktes“

Es ließe sich eine Anzahl weiterer schwerwiegender Kritikpunkte vorbringen. Ich glaube trotz allem, daß von Essens Buch seinen Nutzen hat und daß jemand, der als Lernender Schwierigkeiten mit der deutschen Intonation hat, einiges daraus lernen kann. Es ist aber sicher keine Grundlage für die Beschreibung der deutschen Satzintonation.

2 Dudengrammatik (1959, 1966, 1973)¹⁵

Abschnitt T der Dudengrammatik „Die Klanggestalt des Satzes“¹⁶ (S. 637-668), der von Christian WINKLER verfaßt wurde, ist vom Stoff her die umfangreichste Darstellung der deutschen Satzintonation überhaupt. Sie ist völlig impressionistisch. WINKLER bezieht sich weder auf eigene wissenschaftliche Grundlagenstudien noch auf die anderer, sieht man von zwei marginalen Hinweisen auf KUHLMANN (1931) und von ESSEN (1956) ab. Seine Darlegungen zeigen eine feinsinnige Beobachtungsgabe für die klangliche Gestaltung. „Klanggestalt“ ist für WINKLER die „Gesamtheit dessen, was am Sprachschall über die Reihe der Laute (Phoneme) hinaus als sinnträchtige akustische Gestalt gebildet und aufgefaßt wird“ (S. 637). Bestimmt wird sie durch dreierlei, den Tonfall, die Verteilung der Schweren und Leichten und die Gliederung (der Rede). Gliederungseinheiten sind der *Redeabschnitt*, der *Ausspruch*, die alle beide nicht weiter definiert werden und der *Sinnschritt*, der definiert wird: „Unter *Sinnschritt* verstehen wir eine Redeeinheit, die über ihre unmittelbare sprachliche Bedeutung hinausweist und das in der Wirklichkeit Gemeinte erkennen läßt.“ (S. 610). Die wichtigste Einheit für die Beschreibung der Klanggestalt ist der Ausspruch. Ihm entspricht jeweils ein „Spannbogen“: „Die Rede stellt die Einheit eines Ausspruchs in der Gestalt eines *Spannbogens* dar“. (S. 638). Es ist schwer zu sagen, was dieser Spannbogen nun genau ist; er ist Teil eines ganzen Metaphernfeldes, demzufolge in einer Äußerung eine Spannung aufgebaut und dann, jedenfalls in vielen Fällen, wieder gelöst wird. Definiert wird dieser Begriff nicht. Es heißt dazu aber:

„Die Stimme gestaltet den Spannbogen vornehmlich als Tonhöhenbewegung, mit der sich aber Gliederung und Schwereabstufung verbinden.“ (S. 653; mit „Gliederung“ ist hier offenbar die interne Gliederung des Satzes in Sinnschritte gemeint; ich komme darauf zurück). WINKLER unterscheidet nun die Grundformen der Tonhöhenbewegung oder, wie er sagt, des Tonfalls, die dem *Ausdrucksgehalt* der Rede entsprechen: Der Tonfall „macht die Gefühlslage des Sprechers kund (Kundgabe) und die Absicht seiner Änderung (Auslösung)“. (S. 639). Die drei Grundformen sind:

1. Entladender Ausruf: \searrow z.B. „Peinlich!“
2. Heischender Anruf (Hochschluß): \nearrow . Er tritt vorwiegend in „befehlenden Kurzaussprüchen“ und in Entscheidungsfragen auf.
3. Darstellende Aussage: $\nearrow \searrow$ oder \frown . Sie tritt vorwiegend in Aussagen und Ergänzungsfragen auf. Den ersten Teil dieser Tonhöhenbewegung nennt Winkler „Aufast“, den zweiten „Abast“.

Maßgeblich für die Tonhöhenbewegung sind lediglich „die Schwerpunkte der Silbenkämme, vornehmlich der betonten Silben" (S. 639). Ich weiß nicht, was die Schwerpunkte der Silbenkämme sind, noch, was mit „betonter Silbe" gemeint ist - im Vergleich zur gleich zu erörternden Schwereabstufung. Vielleicht bezieht sich dies auf den sogenannten „Wortton", aber es wird nichts weiter dazu gesagt. Sehr ausführlich geht WINKLER jedoch auf die „Schwereabstufung" ein.

„Unter Schwere verstehen wir das Gewicht, mit dem eine Silbe ins Ohr fällt." (S. 642). Sie ist relativ zu den Nachbarsilben und hängt von verschiedenen Faktoren ab: „Schwer wirkt eine Silbe vornehmlich, wenn sie entscheidende Stufen der Auf- und Abschwünge des Tonfalls übernimmt (musikalischer Akzent). Daneben hebt die Betonung durch Lautheit (Atemdruck) eine Silbe heraus (dynamischer Akzent). Auch Dehnung der Vokale kann eine Silbe beschweren (temporaler Akzent); anderes, z.B. Klangwechsel, kann mitspielen. Gewöhnlich verbinden sich diese Beschwerungsweisen. Für das Deutsche ist der starke Anteil der Betonung durch Lautheit bezeichnend, jedoch führt auch im Deutschen der musikalische Akzent." (S. 642). WINKLER unterscheidet vier Stufen der Schwere. „Sie werden mühelos unterschieden und erlauben, alles sprachlich Wesentliche darzustellen:

- Überschwere (.) gewöhnlich laute Steigspitze oder betonter Schlußfall
- Vollschwere (/) gewöhnlich geringe Tonerhöhung bei deutlich vermehrter Lautheit
- Kaumschwere (\) gewöhnlich fallend, aber noch laut (bleibt unbezeichnet) schwach
- Leichte und abfallend."

(S. 643; die Hinweise in Klammern beziehen sich auf die Notation). Über ihre Funktion wird gesagt, daß die Überschwere stets den Sinnkern des Satzes bezeichnet und entweder etwas Neues oder etwas Besonderes und Unterscheidendes „beschwert" (S. 643); Voll- und Kaumschweren - beide werden ihre Funktion nach nicht unterschieden — treffen „meist die adjektivischen Attribute und die ersten¹⁷⁾ Glieder von Wortpaaren, daneben die formal wirkenden Verneinungen und die Zeigschweren." (S. 648; mit letzteren sind „betonte" Demonstrativa und Anaphern bzw. Kataphern gemeint). Zu den Leichten wird nur gesagt, daß bestimmte Wörter, z.B. Relativpronomen, Artikel, vor allem häufig das finite Verb (!), „den Ton im allgemeinen nicht übernehmen" (nach S. 464). Im Aussagesatz steht die Überschwere am Ende¹⁸⁾; die erste Schwere im Aussagesatz wird als *Einschaltspitze* bezeichnet; zwischen beiden können noch *Binnenschweren* (S. 648) stehen, die entweder fallen oder steigen. Dies erinnert etwas an von Essens Unterscheidung von Vorlauf, rhythmischem Körper mit Schwerezentrum und Nachlauf. Entsprechend heißt es dann auch weiter: „Leichten, die der ersten Schwere vorausgehen, fallen aus mittlerer Höhe ab (*Vorlauf*)" (S. 648) und: „Folgen der Überschwere noch leichte Silben nach, so bleiben sie in der gleichen Tiefenlage (*Nachlauf*)... Bei hoch schließenden Fragen können die der letzten Schwere folgenden Leichten noch über diese hinaus schlagen." (S. 648). Allerdings folgt WINKLER nicht von ESSEN in seinen drei Intonationstypen.

Es bleibt als drittes bestimmendes Merkmal der Klanggestalt neben Tonfall und Verteilung der Schweren und Leichten die Gliederung: „Zur Gliederung der Rede verbinden sich gewöhnlich mehrere phonetische Mittel. Die tieferen Einschnitte (von der Fuge ... an aufwärts) erlauben Atempause und davor eine kennzeichnende Tonhöhenbewegung (Steig- oder Schwebekadenz, Halb- oder Vollschluß). Die flachen Einschnitte (vom Gelenk ... an abwärts) erlauben keine Atempause, sondern stauen den Strom der Laute und gliedern gegebenenfalls durch Neuanfang oder Wendung der Tonhöhenbewegung.“ (S. 649). Diese Erklärung gibt viele Rätsel auf. Was z.B. eine „Schwebekadenz“ ist, was „Steigkadenz“, wird nirgends klar erläutert („Steigkadenz“ ist auf S. 640 der „Hochschluß“ bei heischenden Anrufen, auf S. 650 das „Zeichen des Fortgangs“); der Begriff „Gelenk“ wird eher erläutert noch taucht er später wieder auf. Was ist mit „Stauung des Lautstroms“ gemeint? Man könnte am ehesten an Pausen denken, aber dies paßt einestils oft nicht zu den späteren Beispielen, und andernteils heißt es auf S. 650, daß der Fluß der Rede „evtl. nur durch Dehnung des Auslauts“ gestaut wird. Wie unterscheidet man Tonhöhenbewegungen, die einen Einschnitt anzeigen, von solchen, die eine Schwere kennzeichnen? Und wie verhalten sich beide zum Tonfall, der ja auch nichts als eine Tonhöhenbewegung ist?

Wie schon erwähnt, gibt es nach Winkler drei Gliederungseinheiten, den — weiter behandelten — Redeabschnitt, den Satz und den Sinnschritt; innerhalb eines Sinnschritts kann es noch weitere Untergliederungen durch „Stauung“ geben; sie werden aber nicht benannt. Im Rest seiner Darstellung erläutert WINKLER, wie die Gliederung für verschiedene Konstruktionen aussieht — z.B. bei „wachsenden Gliedern“, bei Anredenominativ, bei Aufzählungen, gereihten Sätzen usw. Dieser Teil enthält viele schöne Beobachtungen, ist allerdings recht unsystematisch und mehr Sprecherziehung als Element einer grammatischen Beschreibung.

3 Kleine Enzyklopädie „Die deutsche Sprache“ (1970)

Die „Kleine Enzyklopädie“ enthält eigentümlicherweise zwei Darstellungen der deutschen Satzintonation; die erste (S. 994-1008) stammt von Eberhard Stock⁹⁾ und liegt auf einer Linie mit von ESSEN und WINKLER; die zweite, kürzere, stammt von ISAČENKO und SCHÄDLICH und stellt in knapper Form deren Zweistufentheorie dar; wir besprechen diese Theorie in Abschnitt 5 unten.

Stock möchte darstellen, „was heute als gesichert gelten kann und für die Erziehung zum mündlichen Ausdruck von Nutzen ist“ (S. 994). Er gibt nicht weiter an, auf welche Untersuchung sich seine Beschreibung stützt; er verweist lediglich auf „Beobachtungen“, die „in zahlreichen Arbeiten veröffentlicht worden sind“ (S. 994); leider werden diese Arbeiten nicht zitiert. Wenden wir uns deshalb seiner Darstellung selber zu.

Zur Klanggestalt, so Stock, tragen die folgenden prosodischen Mittel bei: „der *Stimmklang*, die *Stimmhöhe*, die *Stimmstärke*, das *Sprechtempo* und die *Dauerverhältnisse* in der Wortgruppe sowie die Gliederung des Ausspruchs durch Pausen.“ (S. 994). Sie haben zwei Funktionen:

- "(1) Sie heben Sinnwichtiges hervor, gliedern den Satz in Sprechereinheiten und kennzeichnen ihn als Frage, Aussage oder Befehl, übernehmen also eine grammatische Funktion.
- (2) Sie bezeichnen das Gefühl des Sprechenden, seine emotionale Einstellung" (S. 995).

Für erstere — nur sie interessiert uns hier - sind drei Faktoren von Belang: die Hervorhebung einzelner Redeteile (Satzbetonung oder Satzakkzentuierung), die Gliederung in Sprechereinheiten und der Tonhöhenverlauf (Sprachmelodie oder Intonation). Letztere richtet sich nach den Sprechereinheiten und der Akzentuierung; zugleich ist sie aber maßgeblich für die Gliederung. Wir kommen auf dieses komplizierte Verhältnis zurück. Beginnen wir mit der Satzakkzentuierung.

Der Akzent hebt einzelne Silben, und zwar die „betonte Silbe“ von Wörtern, hervor. Dafür gibt es drei Möglichkeiten:

- "(1) durch den *melodischen Akzent*: die Silbe liegt höher oder tiefer als die sie umgebenden Silben, oder sie tritt durch eine besondere Melodiebewegung hervor;
- (2) durch den *temporalen Akzent*: die Silbe wird gedehnt oder gerafft, gelegentlich kann sie auch durch eine vorausgehende Pause kenntlich gemacht sein;
- (3) durch den *dynamischen Akzent*: die Silbe wird im Unterschied zu den benachbarten Silben mit größerer oder geringerer Lautstärke gesprochen." (S. 997/998).

Sie können einzeln oder gemeinsam auftreten; was im Deutschen üblich ist, wird nicht weiter erläutert. STOCK unterscheidet „für praktische Zwecke“ drei Stufen der Silbenbetonung: hauptbetont, betont, unbetont. Was auffällt, ist der Umstand, daß ganz konsequent die *Abweichung* von der Umgebung, gleich in welche Richtung, als Merkmal der Hervorhebung gesehen wird. Wenn man dies so nähme, wie es steht, würde es zu merkwürdigen Konsequenzen führen. Was ist z.B. bei „Fritz schlief“ hervorgehoben, was nicht, wenn die zweite Silbe starken Tonabfall hat, kurz und mit geringer Lautstärke gesprochen wird? Zweifellos die erste, aber laut Definition könnte es auch die zweite, das Verb, sein. Und wenn man sagt: „Fritz schlief tief, dann weicht die mittlere klar von *beiden* andern ab, müßte demnach hervorgehoben sein. Das ist eigen.

Für die Akzentverteilung im Satz ist nach STOCK die Unterscheidung in *psychologisches Subjekt* und *psychologisches Prädikat* grundlegend: „Der Sprechende verfolgt stets die Absicht, die Aufmerksamkeit des Hörers auf etwas Neues zu lenken oder auf etwas, was zwar nicht neu ist, aber im Moment als wichtig und daher als mitteilenswert empfunden wird ... Das, was eigentlich mitgeteilt werden soll, nimmt eine *zentrale* Stelle im Satz ein ... Man nennt diese Stelle *kommunikatives Zentrum* oder auch *psychologisches Prädikat* ... Daneben enthält der Satz im allgemeinen einen Ausgangspunkt für die Aussage, auch *psychologisches Subjekt* genannt ... Auf der Verbindung von Ausgangspunkt und kommunikativem Zentrum, von Be-

kanntem und Neuem, beruht die sprachliche Verständigung." (S. 997). Der Hauptton eines Satzes liegt nun stets innerhalb des kommunikativen Zentrums. In ruhiger Rede ist das letzte Satzglied kommunikatives Zentrum; im „Zustand der Erregung" (S. 999) kann es auch verschoben werden, und zwar in der Regel an den Satzanfang. Was Ausgangspunkt und was kommunikatives Zentrum ist, muß nach Kontext und Sprechsituation entschieden werden. Innerhalb des kommunikativen Zentrums liegt nun allerdings, so Stock, die Hauptbetonung nicht immer auf dem sinnstärksten Wort. Für ihre genaue Position gibt er, getrennt nach „neutraler Rede", „kontrastiver Betonungsweise" und „emphatischer Betonungsweise" eine Anzahl von Regeln an (S. 999—1001), die „aus der Beobachtung des Sprechgebrauchs und aus der sprecherzieherischen Arbeit abgeleitet wurden" (S. 999); es handelt sich dabei mehr um praktische Orientierungshilfen, die ich hier nicht einzeln erörtern möchte.

Der zweite wichtige Faktor ist die Gliederung in Sprechereinheiten. Nach Stock sind sie „meist durch Pausen voneinander getrennt" (S. 1002), oft zusätzlich durch „Tempoverzögerung" und manchmal durch diese allein. Offenbar sieht er diese suprasegmentalen Kriterien nicht als konstitutiv an, denn es heißt kurz darauf: „Die Regeln für die Untergliederung (in tatsächlich verwirklichte Sprechereinheiten, W. K.) lassen sich im wesentlichen aus der syntaktischen Struktur und aus den Akzentverhältnissen im Satz ableiten (M. BIERWISCH)." (S. 1002). Er gibt dann beispielhaft einige solcher Regeln an, die allerdings mit denen BIERWISCHS (1966) wenig zu tun haben. Auch der Melodieverlauf, so sagt STOCK, diene als Grenzsignal, ohne dies zu erläutern; was er dann zeigt, ist, wie der intonatorische Verlauf die gegebenen Sprechereinheiten weiter charakterisiert. Dazu kommen wir nun.

Dabei können wir uns kurz fassen, denn Stock folgt in nahezu allen Punkten von ESSEN (1956); er vermeidet allerdings die Begriffe „rhythmischer Körper" und „rhetorisches Syntagma"; umgekehrt führt er zusätzlich für von Essens drei Typen der Intonation die Ausdrücke „Vollschluß" (= terminal), „Halbschluß" (= progre-dient) und „Frageton" (= interrogativ) ein. Sachlich ergibt sich aber nichts Neues.

Das heißt zugleich, daß die bei von ESSEN vorgebrachte Kritik auch hier gilt; was darüber hinaus Satzakkentuierung und Gliederung in Sprechereinheiten betrifft, so ist zunächst einmal zu bemerken, daß es fast keine fundierten empirischen Untersuchungen gibt, auf die sich die gegebene Beschreibung stützen könnte. Sie ist eine Art Astronomie der Babylonier. Zum zweiten ist der zugrundegelegte Akzentbegriff, wie schon bemerkt, sicher nicht haltbar. Und drittens muß doch immerhin hervorgehoben werden, daß Stocks Darstellung recht klar und konzis ist, soweit dies beim Stand des Wissens möglich scheint.

Ich schließe damit die Diskussion der „klassischen" Betrachtungsweise ab; sie findet sich auch in anderen neueren Grammatiken, z.B. ERBEN (1961), SCHMIDT (1966), BRINKMANN (1971), selbst noch in FLÄMIG (1974); deshalb erübrigt es sich, auf diese Grammatiken besonders einzugehen.

4 Kohler (1977)

Einen ganz anderen Weg geht KOHLER in seiner „Einführung in die Phonetik des Deutschen“. Prosodische Erscheinungen behandelt er an drei Stellen²⁰): unter phonetischen Aspekten (S. 82—85), unter phonologischen (S. 118-128) und schließlich speziell aufs Deutsche bezogen (S. 191—207). KOHLER nimmt (im Anschluß an LEHISTE (1970)) drei phonologisch relevante suprasegmentale oder prosodische Größen an: *Quantität*, *Wort- bzw. Satzakkzent* und *Intonation*²¹). Phonetisch entsprechen ihnen *Dauer*, *Prominenz* und *Tonhöhe*. Quantität als phonologisches Merkmal (z.B. Langvokal vs. Kurzvokal) spielt in unserem Zusammenhang keine Rolle. „Wortakzent“ ist ein binäres phonologisches Merkmal von Silben, das durch Lautdauer, Tonhöhenverlauf, Lautqualität und Lautstärke realisiert wird; wie dies im einzelnen fürs Deutsche aussieht, wird nicht gesagt²²). Jedes deutsche Wort hat, so sagt KOHLER, eine akzentuierte Silbe; Komposita können auch mehrere haben (diesen Fall betrachte ich im folgenden nicht besonders). Wo die akzentuierte Silbe liegt, hängt von relativ festen Regeln ab, die er sehr schön beschreibt (S. 191 — 196). Dies ist recht klar und ein guter Ausgangspunkt. Völlig unklar ist, was KOHLER mit Satzakkzent als phonologischer Kategorie meint. Er charakterisiert ihn weder in funktionaler Hinsicht - wofür ist er distinktiv? - noch im Hinblick auf seine phonetische Realisierung, ebensowenig wird gesagt, in welchem Verhältnis er zum Wortakzent steht. Es heißt dazu bloß: „Das Satzakkzentrastrer muß ... im Deutschen so angelegt sein, daß es nur auf Wortakzente fällt, obwohl nicht sämtlichen Wortakzenten ein Satzakkzent zu entsprechen braucht“ (S. 124). Zudem können aber auch — bei Kontrastivität und Emphase — Silben ohne Wortakzent einen Satzakkzent haben.

Wie immer es um den phonologischen Status des Satzakkzents stehen mag: einen Satzakkzent haben nach KOHLER Silben, die (von Kontrastivität und Emphase abgesehen) auch einen Wortakzent tragen und die im Deutschen dazu tendieren, in ungefähr gleichen Zeitabständen aufeinander zu folgen; zwischen diesen „Hebungen“ kann eine unterschiedliche Anzahl nichtakzentuierter Silben stehen, deren Dauer dementsprechend aus rhythmischen Gründen variiert werden muß. Eine Hebung samt den folgenden unakzentuierten Silben bis zur nächsten Hebung nennt KOHLER „Takt“. Jede Äußerung läßt sich in eine Folge zeitlich ungefähr gleich langer Takte zerlegen (eventuell mit einem ersten „unvollständigen“ Takt ohne Hebung). Diese Taktgliederung bildet den Ausgangspunkt seiner Intonationsbeschreibung.²³)

Eine Intonationseinheit (eine Äußerung) besteht aus einer Folge von Takten, deren einer ein bestimmtes Intonationsmuster, einen *Ton*, trägt; diesen Takt nennt KOHLER *Nukleustakt*; die vorausgehenden Takte bilden den *Pränukleus*, die folgenden den *Nachlauf*; Nukleustakt und Nachlauf bilden gemeinsam den *Nukleus*. Pränukeus und Nachlauf können fehlen (bei eintaktigen Intonationseinheiten). Es werden sechs Töne unterschieden (S. 198):

Ton 1: bis in eine tiefe Lage fallender Tonhöhenverlauf

Ton 2: hoch ansteigender Tonhöhenverlauf

- Ton 3: nur bis zu einer mittleren Lage ansteigender Tonhöhenverlauf
- Ton 4: gleichbleibende mittlere Tonhöhe
- Ton 5: (von hoch) fallender und gleich wieder bis zur Mittellage ansteigender Tonhöhenverlauf
- Ton 6: (aus mittlerer Lage) ansteigender und anschließend bis in tiefe Lage abfallender Tonhöhenverlauf.

Es ist zu betonen, daß dies formale Muster sind, über deren Funktion damit noch nichts gesagt ist. Der Nukleustakt trägt also einen dieser Töne. Ist er einsilbig, so verteilt sich die Bewegung auf diese eine Silbe; ansonsten beginnt sie in der Hebung, die Kohler auch *Nukleussilbe* nennt, und setzt sich über die nichtakzentuierten Silben fort.

Über die Tonhöhenbewegung im Nachlauf wird nichts weiter gesagt; man beachte, daß der Nachlauf sehr wohl Hebungen enthält. Dies steht in krassem Gegensatz zur Tradition der deutschen Intonationsbeschreibung, bei der ja der letzten Hebung eine entscheidende Rolle zukommt. Der Tonhöhenverlauf im Prä nukleus unterliegt gewissen Beschränkungen, die von der Art des Tons im Nukleustakt abhängen. Bei Ton 1 gibt es zwei Prä nukleustypen:

- (a) hoch eben, tief eben, von tief zu hoch ansteigend (diese Formen werden offenbar als gleichwertig betrachtet),
- (b) „springend“, d.h. pro Takt ansteigend oder pro Takt abfallend.

Von Ton 2 gibt es gleichfalls zwei Prä nukleusmuster:

- (a) hoch eben oder von hoch zu tief abfallend
- (b) tief eben.

Bei den restlichen Tönen gibt es keine klaren Muster, wohl aber Beschränkungen: die letzte Silbe vor Ton 4 ist tief, bei Ton 5 und 6 hat, so KOHLER, der gesamte Prä nukleus eine mittlere Lage.

Den einzelnen Mustern, bestehend aus Prä nukleus und Ton, läßt sich nach KOHLER „keine konsistente, von allen Äußerungen losgelöste Bedeutung“ (S. 199) zuordnen. Gemeinsam mit syntaktischem Aufbau, Wortbedeutungen, Situationskontext u.a. tragen sie zur Gesamtbedeutung bei. Was ihr möglicher Bedeutungsbeitrag ist, wird bei KOHLER nur vorsichtig an einigen Beispielen angedeutet, etwa „Sätze in der syntaktischen Form des Imperativs erhalten durch Ton 1 Befehlscharakter, vor allem bei tiefem Prä nukleus ... Durch Ton 3 mit tiefem Prä nukleus nehmen sie einen freundlichen Ausdruck an ... Bei Ton 2 mit tiefem Prä nukleus wirkt dieser gleichzeitig insistierend ... Ton 2 und 3 mit hohem Prä nukleus sind übertrieben freundlich, ja patronisierend; sie kennzeichnen den typischen Kindergartenton. Ton 4 vermittelt schließlich Desinteresse, Teilnahmslosigkeit und Gleichgültigkeit.“ (S. 205). Eine spezifische grammatische Funktion hat die Intonation nicht.

KOHLERS Skizze der deutschen Intonation beruht nicht nur auf einer anderen Beschreibunggrundlage als die „traditionellen“ Ansätze, sie ist auch weitaus vorsichtiger und sorgfältiger in ihren Festlegungen. Sie wirft aber auch eine Fülle von Problemen auf, von denen ich nur einige hier erwähnen will.

1. Der größte Schwachpunkt ist zweifellos, daß sich auch Kohler auf keine systematischen empirischen Untersuchungen zur Intonation des Deutschen berufen kann. Er stützt sich auf Einzelbeobachtungen einerseits und auf Hallidays System zur Beschreibung der englischen Intonation andererseits.
2. Der zentrale Begriff „Satzakzent“ ist völlig unklar. Dies wurde bereits weiter oben erörtert. Solange nicht klar ist, was ein Satzakzent ist, wo im Satz Satzakzente liegen, wie sie realisiert werden, was über ihr Vorhandensein entscheidet, hängt die gesamte Beschreibung völlig in der Luft.
3. Damit zusammen hängt, daß auch der Begriff des Taktes empirisch sehr schlecht gesichert ist. Mit diesem Begriff steht und fällt aber ganze Ansatz.
4. Es ist mir gänzlich unklar, in welchem Sinn die sechs Töne in der Tat ein paradigmatisches System von Oppositionen darstellen. Erstens gibt es - wenn wir einmal den Begriff des Taktes als gegeben annehmen — viele weitere regelmäßig auftretende Intonationsverläufe auf Takten, z.B. von hoch auf halbhoch fallend, tief eben, hoch eben, fallend-steigend-fallend, u.a.; wieso also sind ausgerechnet diese sechs distinktiv? Und zweitend: in welchem Sinn sind sie distinktiv?
5. Sowohl der Prä nukleus wie der Nukleustakt und der Nachlauf haben Hebungen, und alle Takte haben irgendeinen Intonationsverlauf. Wie ist nun festgelegt, wo der „Ton“ ist, d.h. was der Nukleustakt ist? Betrachten wir etwa einen einfachen Dreitakter wie „Peter kochte gestern“ und nehmen an, jedem Wortakzent entspricht ein Satzakzent. Dann könnte z.B. „gestern“ Nukleustakt mit Ton 1 sein und der Prä nukleus von Typ (b) — springend, wobei beide Takte fallen — sein, d.h. wir haben

Peter kochte gestern
 — — — — —

Aber bei dieser Gesamtkontur könnte man genausogut annehmen, daß „kochte“ Nukleustakt mit Ton 1 ist, oder „Peter“, wobei der Rest Nachlauf wäre (wohl-gemerkt: der Nachlauf hat Hebungen). In diesem speziellen Fall könnte man vielleicht fordern, daß der Nachlauf tief eben ist, aber erstens tut dies KOHLER nicht, zweitens würde es in anderen Fällen nicht helfen, und drittens taucht dann die Frage auf, wie denn der „Hebungscharakter“ im Nachlauf zum Ausdruck kommt. Man könnte alternativ annehmen, daß der Nukleustakt überhaupt der letzte Takt mit Hebung ist — und damit zugleich der letzte Takt überhaupt. Dies entspräche der klassischen Betrachtungsweise bei von Essen und andern. Aber dies würde sofort in Schwierigkeiten anderer Art führen, denn dann ließe sich der Gedanke des „akzentzählenden Rhythmus“ kaum aufrecht erhalten; man erhielte beliebig lange Takte, da ja der „Ton“ schon mit der ersten Silbe einer Äußerung beginnen kann.

6. Auch der Prä nukleus besteht aus Takten, enthält also Hebungen und in der Regel nichtakzentuierte Silben. Wie wird dann der Unterschied „Hebung - Senkung“ ausgedrückt, wenn z.B. der Prä nukleus gleichmäßig tief ist, oder gleichmäßig hoch ist? Wie ist es zu erklären, daß die Kontur innerhalb eines Prä nukleustaktes ansteigen kann (Senkung höher als Hebung), aber auch fallen kann

(Hebung höher als Senkung)? Die Unterscheidung in akzentuierte und nichtakzentuierte Silben innerhalb eines Taktes müßte demnach ganz unabhängig von der Tonhöhenbewegung sein, d.h. reiner Lautstärkenakzent (oder Quantitätsakzent). Das glaubt aber kein einziger Forscher.

7. Es wird keinerlei Bezug zur syntaktischen Struktur, etwa zur Gliederung in Konstituenten, zur Inversion usw. hergestellt.

Auch hier könnte man die kritischen Einwände fortsetzen, aber die angeführten scheinen mir hinlänglich klar zu machen, daß sich Kohlers Ansatz in dieser Form sicher nicht halten läßt.

5 Isačenko-Schädlich (1966)²⁴⁾

Unter allen Versuchen, die deutsche Satzintonation zu beschreiben, ist der von ISAČENKO und SCHÄDLICH (künftig IS) der radikalste Neuanatz. IS unterscheiden expressive (oder emotionelle) und syntaktische Funktionen der Intonationen; mit ersteren befassen sie sich nicht²⁵⁾. Zur Beschreibung der syntaktischen Funktion, die die Intonation im Deutschen hat, gehen sie von der Existenz zweier (und nur zweier) Tonstufen (*Hochton* und *Tief-ton*) aus, die mindestens einen Halbton auseinanderliegen (das sind bei einer durchschnittlichen Männerstimme etwa 10 Hz); ihre absolute Höhe schwankt natürlich, je nach Sprecher, emotionalen Zuständen des Sprechers, u.a. Relevant sind nicht die Tonstufen an sich, sondern die Wechsel (*Tonbrüche*), von denen jede normale Äußerung mindestens einen hat; jede Äußerung läßt sich daher als Folge von hochtonigen und tief-tonigen Segmenten darstellen. Tonbrüche können nur an bestimmten Stellen vorkommen. IS nehmen an, daß jedes Wort²⁶⁾ ein *markierbares* (hervorhebbares) silbisches Segment hat, das sie *Iktus* nennen. In einsilbigen Wörtern sowie in sog. „endbetonten“ Wörtern ist der erste Teil der Silbe (bzw. der letzten Silbe) iktisch, der Rest nicht-iktisch; in allen andern Wörtern ist die „betonte“ Silbe der Iktus. Ein Tonbruch kann nur unmittelbar vor (präiktisch) oder unmittelbar nach eine Iktus (postiktisch) liegen. Hat ein Wort einen Tonbruch - gleich ob fallend oder steigend — so wirkt der betreffende Iktus und damit das Wort selbst hervorgehoben²⁷⁾

Für die syntaktische Funktion der Tonbrüche hat man vier Fälle zu unterscheiden: präiktisch steigend, postiktisch steigend, präiktisch fallend, postiktisch fallend. „Als invariante syntaktische Bedeutung der steigenden Tonbrüche wird die *Ankündigung eines fallenden Tonbruches* angesehen“ (S. 60). Der fallende Tonbruch gilt als merkmalloses Glied der Opposition „steigend-fallend“. „Seine syntaktische Bedeutung ist rein negativ.“ (S. 61). Diese Bestimmungen verleihen dem *letzten* Tonbruch in der Äußerung einen besonderen Rang. Steigt er, so ist ein weiterer angekündigt - der aber nicht folgt; hier kommt es nach IS nun darauf an, ob er präiktisch oder postiktisch ist; im ersten Fall wird die Äußerung als „Satzfragment“ interpretiert, im zweiten als Frage. Fällt der letzte Tonbruch, so wird nichts Besonderes ausgedrückt — außer der Hervorhebung natürlich, wie bei jedem Tonbruch; ist er postiktisch, so wird er als „kontrastierend“ empfunden. Nichtletzte fallende

Tonbrüche bilden die Grenzen von „Phrasen“; diese Grenzen können auch durch Pausen gekennzeichnet werden; zur Pausenstruktur und zur Phrasierung führen IS jedoch nichts Näheres aus.

Das System von IS ist bestechend einfach; darüber hinaus kann es sich in der Tat auf empirische Untersuchungen stützen. IS haben dazu einfache deutsche Sätze sprechen lassen, sie auf zwei Frequenzen — etwa im Abstand eines Halbtons — monotonisiert und dann Segmente aus den beiden Bändern mit Hochton und Tiefen zusammen geschnitten. So entstanden Äußerungen, die nur aus tiefen und hohen Segmenten bestehen, etwas unnatürlich klingen, aber nach IS noch alle syntaktisch relevante Information enthalten; diese Äußerungen wurden dann in einer Reihe von Experimenten durch — phonetisch nicht vorgebildete — Informanten beurteilt (Einzelheiten finden sich in ISACENKO-SCHÄDLICH (1970a) Zu diesem empirischen Vorgehen ist zweierlei zu sagen:

1. Aus der Tatsache, daß ein Leser des Arabischen oder Hebräischen einen geschriebenen Satz auch versteht, obwohl die Vokale nicht vorhanden sind, wird man nicht den Schluß ziehen, daß für diese Sprachen nur die Konsonanten relevant sind. Man kann aus den vorhandenen Informationen des Signals einfach das Fehlende aufgrund sprachlichen und situativen Wissens rekonstruieren. Wenn man aus dem Signal nicht nur Tonhöenschwankungen entfernt, sondern — wie bei geflüsterter Sprache — das übliche akustische Korrelat der Tonhöhe selbst, so sind die Äußerungen immer noch klar interpretierbar. Daraus kann man aber nicht folgern, die Tonhöhe sei belanglos; sie wird einfach zu rekonstruieren versucht, und dies gelingt auch weitgehend, wie man weiß.
2. Es wundert nicht, daß bei diesem Vorgehen fast nichts an syntaktisch relevanter Intonation übrig bleibt. Sätze werden als Frage, Nicht-Frage oder Satzfragment charakterisiert. Das ist alles. Und selbst dies ist falsch (bei Doppelfragen z.B. ist der letzte Tonbruch fallend, ebenso bei vielen Entscheidungsfragen einfacher Art — etwa bei „Kommt Peter?“, wenn „kommt“ stark „betont“ ist). Mit anderen Worten: es kommt halt fast nichts heraus.

Man kann eine Reihe weiterer Einwände vorbringen. So wird kaum eine Verbindung zur syntaktischen Struktur eines Satzes geschlagen. Es gibt viele durchaus nicht emotionelle, sondern syntaktische Phänomene, die mit der Intonation zusammenhängen, von denen man aber nicht weiß, wie sie in diesem Rahmen behandelt werden können (vgl. etwa oben in Abschnitt 2, was unter Kritikpunkt 3. zu von Essen gesagt wurde). In der gesprochenen Rede hat ein Satz wie „er kochte“ oft einen Intonationsverlauf wie – – –, d.h. Abfall von „er“ auf „koch“ und von „koch“ auf „te“; da ja nicht zwei fallende Tonbrüche aufeinanderfolgen können, zählt einer nicht; aber welcher? Ist der andere Abfall emotional, und wenn ja, warum? Mir scheint, damit sind für den Anfang genügend Probleme angedeutet. Der Ansatz von IS läßt sich sicher nicht halten. Aber man kann ohne Zweifel viel daraus lernen: was sie sagen, ist klar, und es beruht nicht auf bloßen Impressionen wie meist in der deutschen Intonationsforschung — was immer man gegen ihre empirischen Untersuchungen nun vorbringen mag.²⁸⁾

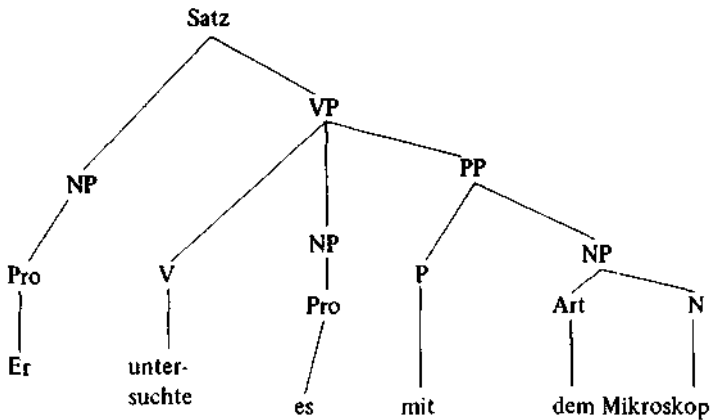
6 Bierwisch (1966)

BIERWISCHS Monographie ist die linguistisch weitaus anspruchsvollste und reflektierteste Arbeit zur deutschen Satzintonation. Sie geht von einer gut ausgearbeiteten syntaktischen Theorie, der Transformationsgrammatik jener Tage aus. Ausgangspunkt ist die syntaktische Oberflächenstruktur, auf der die verschiedenen phonologischen Regeln — unter anderem die der Intonation — arbeiten: „Die Intonation ist Teil der phonetischen Transkription ... und sie determiniert den Tohhöhenverlauf einer Äußerung bis auf emotional bedingte Varianten und die physiologisch bedingte Realisierungsform. Sie ist prädictabel aufgrund dreier Faktoren, die entweder unmittelbar zur syntaktischen Struktur gehören oder durch sie determiniert sind: Akzent, Grenzsymbole und SIM (= syntaktische Intonationsmarker, z.B. Fragemorphem, W.K.)" (S. 104).

Im folgenden will ich versuchen, unter Verzicht auf technische Formulierungen (und damit auch auf Präzision) die Grundgedanken wiederzugeben. Ausgangspunkt ist, wie gesagt, die syntaktische Oberflächenstruktur; auf ihr operieren unter anderem die Akzentregeln, die kein Teil der Intonationsregeln sind, aber eine Voraussetzung dafür. Die Akzentregeln²⁹⁾ weisen zunächst jedem lexikalischen Formativ (im wesentlichen V, N, A) das Akzentmerkmal [1 A] zu, den übrigen Formativen das Merkmal [- A], d.h. Verben, Nomina haben den Akzent 1, alle andern Elemente haben keinen Akzent. Durch (zyklisch operierende) Regeln werden dem syntaktischen Aufbau folgend nunmehr vorhandene Akzente hochgestuft, z.B. zu [2A], [3A] usw; höhere Zahl bezeichnet *schwächeren* Akzent; so könnte aus „das schöne

₁ Haus" etwa werden „das schöne Haus." Wie dies erfolgt, hängt von der jeweiligen Struktur ab; in den hier interessierenden Fällen behält die am weitesten rechts stehende Einheit ihren Wert, also [1A]. Soviel zur Grundidee. Zu beachten ist noch, daß unter besonderen Bedingungen auch andere Elemente als V, N und A einen Akzent haben können.

Gegeben sei also die Oberflächenstruktur mit Akzentmarkierung. Die ihr entsprechende Endkette wird nun in *Phrasierungseinheiten* (PE) zerlegt. Dazu werden durch eine allgemeine Konvention Grenzsymbole eingeführt, die dann durch zwei Regeln reduziert werden. Eine Konstituente, die von *i* vielen verzweigenden Knoten dominiert wird, heißt Konstituente *i*-ter Ordnung (ein Satz ist 0. Ordnung, NP und VP 1. Ordnung, usw.); eine allgemeine Konvention³⁰⁾ (K) besagt nun, daß zwischen zwei Konstituenten *i*-ter Ordnung ein Grenzsymbol „i #“ eingeführt wird. BIERWISCH gibt folgendes Beispiel (S. 110):



0 #Er 1 # untersucht 2 # es 2 # mit 3 # dem 4 # Mikroskop 0 #

Die Grenzsymbole — die äußeren sind hier hinzugefügt — rekonstruieren also die Phrasengrenzen; man beachte, daß die „Grenze“ syntaktisch umso schwächer ist, je höher der Zahlenwert des Grenzsymbols liegt. Ein Teil der Grenzen muß nun aufgehoben werden, um sinnvolle PE zu bilden. Eine erste Regel dazu betrifft Enklise und Proklise. Sie besagt, daß eine Einheit ohne Akzent sich nach rechts oder links einer Einheit mit Akzent anschließt, je nachdem, wo die schwächere Grenze ist (bei gleichstarken Grenzen geht sie nach links); diese Regel ist zyklisch; sie beginnt mit der schwächsten Grenze (jener mit dem höchsten „i#“). Für den Satz in (1) sähe dies so aus (nur „untersuchte“ und „Mikroskop“ haben Akzent):

(7)

- (a) 0 # Er 1 # untersuchte 2 # es 2 # mit 3 # dem 4 # Mikroskop 0 #
- (b) 0 # Er 1 # untersuchte 2 # es 2 # mit 3 # dem Mikroskop o L
- (c) 0 # Er 1 # untersuchte 2 # es 2 # mit dem Mikroskop 0 #
- (d) 0 # Er 1 # untersuchte es 2 # mit dem Mikroskop o #
- (e) 0 # Er untersuchte es 2 # mit dem Mikroskop 0 #

Wie (e) zeigt, hat dieser Satz also zwei PE. Weiterhin können nun Grenzsymbole aufgrund „rhythmischer Verschmelzung“ wegfallen. Diese Regel besagt, daß eine Einheit, die nicht mehr als p viele Akzente hat, mit der rechten oder linken zu einer PE verschmolzen wird — auch hier wiederum entsprechend der schwächeren Grenze (bei gleichstarken diesmal nach rechts); p ist ein Parameter, der normalerweise zwischen 0 und etwa 3 schwankt und von Faktoren wie beispielsweise der Sprechgeschwindigkeit abhängt; für $p = 0$ würde diese Regel an (7e) nichts mehr ändern, da beide PE mehr als 0 Akzente haben; für $p = 1$ würden beide PE zu einer verschmelzen, da sie nicht mehr als einen Akzent haben. Pro- und Enklise sowie rhythmische Verschmelzung faßt Bierwisch formal zu einer Regel (Pi) zusammen. Eine weitere Regel (Pii) besagt nun schließlich, daß alle etwa verbliebenen Grenzsymbole hinter dem letzten 1-Akzent beseitigt werden („Nachlauf“). (Pi) und (Pii)

legen — mit einigen Verfeinerungen, insbesondere für eingebettete Sätze — die Phrasierungseinheiten fest, auf denen nunmehr die eigentlichen Intonationsregeln operieren.

Für die Transkription nimmt BIERWISCH an, daß nicht feste Tonhöhen relevant sind, sondern *Tonhöhenveränderungen unterschiedlicher Größe* und daß es einen Unterschied macht, ob die Tonhöhenveränderungen eine Haupttonsilbe betreffen (stärkster Akzent der betreffenden PE), eine Nebensilbe oder nur unbetonte Silben; im letzten Fall spricht er von phonetischen, in den beiden ersten von phonologischen Tonschritten. Da nur Sonoranten (genauer: Vokale, Liquide, Nasale) Tonhöhenveränderung zeigen, verwendet er zur Notation ein Merkmal [n Sonorant], wobei $n = 0, 1, 2, \dots$; n bezieht sich aber nicht auf feste Tonstufen, sondern mithilfe von n wird lediglich der Unterschied zum vorausgehenden und folgenden Segment ausgedrückt. Es ist deshalb gleich, ob man (8a) oder (8b) schreibt (dabei ist X ein — möglicherweise leeres — Segment, das nicht Tonträger ist also z.B. ein stimmloser Kosonant):

(8a) X [1 Son] X [2 Son] X [4 Son] X [1 Son]

(8b) X [7 Son] X [8 Son] X [10 Son] X [7 Son]

Was zählt, sind die Unterschiede: leichter Anstieg, stärkerer Anstieg, starker Abfall. Um den Haupttonschritt einer PE zu kennzeichnen, nimmt Bierwisch zwei weitere Merkmale [\pm Anstieg], [\pm Fall] an. Sie sind auch wichtig für rein phonetische Tonhöhenveränderungen (also solche in akzentlosen Silben); so gilt etwa, daß nach Grenzsymbol oder [+ Fall] benachbarte Silben mit gleichen [Son]-Werten abwärtsgerichtet sind und nach [+ Anstieg] aufwärtsgerichtet, während sie in allen andern Fällen frei variieren; weiterhin dienen sie dazu, Tonhöhenveränderungen *innerhalb* des Akzentvokals zu kennzeichnen.

Wie schon erwähnt, hängt die Intonation nicht nur von Akzent und Phrasierung ab, sondern auch von bestimmten Morphemen, die Bierwisch „syntaktische Intonationsmarker (SIM)“ nennt; er nimmt drei SIM an; Q (löst Frageintonation aus, s.u.), E (steht nach emphatisch hervorgehobener Konstituente) und D, das [+ Fall] vor den Hauptakzent plaziert.³¹⁾

Die eigentlichen Intonationsregeln operieren nun auf einem Satz, der (a) in PE gegliedert ist, (b) eine Akzentverteilung gemäß den Akzentregeln hat, (c) bei dem jeder Vokal als [- Fall], [- Anst] markiert ist, und (d) bei dem (vereinfacht gesagt) jede Silbe als [+ Son] gekennzeichnet ist; die Intonationsregeln ändern nun diese Merkmale in geeigneter Weise. Es gibt zehn solcher Regeln, die ich im folgenden ihrem Inhalt nach umgangssprachlich paraphrasierte³²⁾ (mit Hauptakzent wird im folgenden ein Akzent der Stufen 1, 2 oder 3 bezeichnet):

- (Ai) Der letzte 1-Akzent im Satz erhält die Merkmale [+ Fall] und [3 Son], falls eine „terminale Intonation“ und als Bezugsgröße für Tonhöhenunterschiede den Wert „3“, der im Augenblick aber noch nichts besagt).
- (Aii) Die stärkstakzentuierten Silben aller anderen PE — falls Q vorhanden ist, auch die letzten PE - erhalten die Merkmale [+ Anst] und [3 Son]. (Da-

mit hat also in jeder PE der stärkste Akzent den Wert [3 Son] und das Merkmal [+ Fall] oder [+ Anst]).

- (Aiii) In jeder PE erhalten die Hauptakzente *vor* dem stärksten von rechts nach links die Werte [4 Son], [5 Son], [6 Son],... (Man erhält also eine Folge ...: x [6 Son] x [5 Son] x [4 Son] x [3 Son] x, wobei x möglicherweise leere Segmente sind, die keinen Hauptakzent aufweisen; dieser allmähliche Abfall auf die stärkstakzentuierte Silbe einer PE zu entspricht von Essens fallenden Führtönen im rhythmischen Körper).
- (Aiv) In jeder PE wird eine mit [+ Anst] markierte Silbe von [3 Son] auf [6 Son] gesetzt, falls in dieser PE Hauptakzente vorangehen (sonst könnte es eine progrediente Intonation nur *innerhalb* der stärkstbetonten Silbe geben, die aber an sich die tiefste in der Reihe der Führtöne wäre).
- (Av) Alle als [+ Son] gekennzeichneten Elemente *nach* [+ Fall] werden auf [1 Son], d.h. deutlich tiefer als die stärkstbetonte Silbe gesetzt. (Diese tiefen Silben bilden den Nachlauf).
- (Avi) In jeder PE erhalten die [+ Son]-Elemente *vor* dem ersten Hauptakzent einen Wert, der um 2 tiefer ist als der dieses ersten Hauptakzents („Vorlauf“³³).
- (Avii) Noch verbleibende [+ Son]-Elemente erhalten denselben Wert wie das vorausgehende [+ Son]-Element. (Dies betrifft vor allem Elemente zwischen den Führtönen, die also durchweg „enklitisch“ sind).
- (Aviii) Die Tonveränderung der stärkstbetonten Silbe jeder PE wird *vor* den Akzentgipfel gelegt, falls Q (bei Anstieg) oder D (bei Fall) vorhanden sind; sonst liegt sie dahinter. (Das entspricht den Beobachtungen von Isačenko und Schädlich).

Die beiden letzten Regeln betreffen lediglich Sätze mit dem Emphasemorphem E. Dabei wird vorausgesetzt, daß die „emphatische“ Konstituente stets einen 1-Akzent (und damit zunächst einmal den Wert [3 Son]) hat, sowie, daß sonst kein Hauptakzent vorhanden ist:

- (Aix) die emphatische Konsistente wird von [3 Son] auf [5 Son] gesetzt.
- (A x) Alle andern [+ Son]-Elemente erhalten [1 Son]:

Inhaltlich entsprechen diese zehn Intonationsregeln weitgehend den Befunden von Essens, sieht man einmal davon ab, daß sie wesentlich präziser sind und sich deshalb oft stärker festlegen als die manchmal unbestimmten Angaben von Essens. Die zweite Intonationsregel könnte den Eindruck vermitteln, als hätten alle Frage-sätze steigende Intonation am Ende (genauer: in der am stärksten betonten Silbe der letzten PE). Dies ist natürlich nicht so, und in einem eigenen Kapitel diskutiert Bierwisch ausführlich die Bedingungen, unter denen ein in der Tiefenstruktur vorhandenes Q eliminiert wird, bevor es für die Intonation wirksam werden kann. Darauf gehe ich aber hier nicht ein.

Der Satz besteht also aus drei PE „ein Zigeuner“, „hat einen Traktor“ und „gekauft“. Aufgrund rhythmischer Verschmelzung können PE noch einmal zusammengefügt werden, wenn sie nicht mehr als p viele Akzente enthalten; für p = 0 bleibt (d) natürlich erhalten; schon für p = 1 fallen alle PE zu einer einzigen zusammen:

(e) $\overset{2}{0 \#}$ ein $\overset{3}{Zigeuner}$ hat einen $\overset{1}{Traktor}$ gekauft $\overset{3}{0 \#}$

4. Schritt: Zuweisung der Intonation

Auf (d) bzw. (e) werden nun die Regeln (Ai) bis (Aviii) angewandt; jede Silbe ist dabei zunächst als [+ Son] gekennzeichnet; statt [+ Son], [3 Son] usw. schreibe ich im folgenden einfach +, 3 usw. unter die Silben; beginnen wir mit (d), d.h. drei PE:

$\overset{2}{0 \#}$ ein $\overset{3}{Zigeuner}$ $\overset{1}{1}$ hat einen $\overset{3}{Traktor}$ $\overset{1}{3}$ gekauft $\overset{3}{0 \#}$

+ +++ + ++ ++ ++

Ai ergibt:

+ +++ + ++ ++ + 3
[+ Fall]

Aii ergibt:

+ +3+ + ++ 3+ + 3
[+ Anst] [+ Anst] [+ Fall]

Aiii – Av laufen leer; Avi ergibt:

1 13+ 1 1 1 3+ 1 3
[+ Anst] [+ Anst] [+ Fall]

Avii schließlich führt zu:

1 133 1 1 1 3 3 1 3
[+ Anst] [+ Anst] [+ Fall]

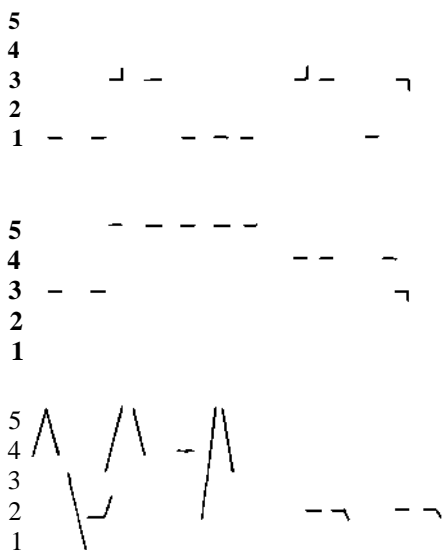
Dies ist die endgültige Struktur; Regel Aviii legt nur noch fest, wo genau in der Silbe - vor oder nach dem Akzentgipfel-Anstieg bzw. Fall liegen; hier liegen sie stets nach. Dieser Intonationsverlauf entspricht einer relativ starken Phrasierung; geht man nicht von dreien, sondern von nur einer PE aus (Fall (e)), so ergibt sich nach Ai zunächst dasselbe, nämlich:

	2		3		1
0 # ein	Zigeuner		hat einen	Traktor	gekauft 0 #
	+ + + +		+ ++ ++		+ 3
					[+ Fall]
Ai läuft leer;		Aiii führt zu			
+ + 5 +		+ ++ 4 +			+ 3
					[+ Fall]
Aiv und Av laufen leer;		Avi ergibt			
3 3 5 +		+ ++ 4 +			+ 3
					[+ Fall]
Avii schließlich		führt zu:			
3 3 5 5		5 5 5 4 4			4 3
					[+ Fall]

Nach Aviii liegt der Fall in der letzten Silbe nach dem Akzentgipfel.

Je nach Phrasierung ergeben sich also in der Tat völlig andere Intonationsverläufe. Sie sind im folgenden noch einmal in einer einfachen, aber anschaulichen Schreibweise dargestellt; dabei ist zu beachten, daß es nicht auf die absolute Höhe, sondern nur auf die Abstände ankommt. Nach demselben Schema habe ich darunter den ungefähren Verlauf einer empirisch ermittelten Tonhöhenkurve gesetzt (vgl. dazu weiter unten Abschnitt 7):

(10) ein Zigeuner hat einen Traktor gekauft



Die Übereinstimmung scheint mir in beiden Fällen nicht sehr groß, aber daß kann verschiedene Ursachen haben. Ich komme auf diesen Punkt gleich noch zurück. Bierwischs Arbeit ist bei weitem die beste Behandlung der deutschen Satzintonation überhaupt. Es gibt auch für andere Sprachen wenig Arbeiten, die vergleichbar geschlossen und durchdacht sind, und angesichts vieler anderer deutscher Arbeiten auf diesem Gebiet ist es vielleicht nicht überflüssig zu bemerken, daß man — unabhängig davon, ob man seine Ansicht teilt —jedenfalls verstehen kann, was gemeint ist. Ich habe zwei Einwände, oder besser gesagt: mögliche Einwände gegen seine Analyse.

1. Die enge Verknüpfung mit einem bestimmten Grammatikmodell, der Transformationsgrammatik um 1965, ist in vieler Hinsicht ein Vorteil, weil nur so eine derart geschlossene Behandlung möglich erscheint. Aber es ist klar, daß sie mit der Annahme dieses Modells (oder jedenfalls wesentlicher Komponenten dieses Modells) steht und fällt. Ich will dies nur an vier Punkten andeuten:

- (a) Über die „richtigen“ Oberflächenstrukturen deutscher Sätze besteht nach wie vor keinerlei Einigkeit; es ist aber klar, daß unterschiedliche Oberflächenstrukturen in der Regel zu andern PE und damit zu andern Intonationen führen.
- (b) In den neueren Versionen der Transformationsgrammatik enthält die Tiefenstruktur keine "trigger"-Symbole wie Q mehr. Sollten sie aus reinen Intonationsgründen beibehalten werden, und wenn nicht: wie wird die notwendige Information sonst aus der Tiefenstruktur abgeleitet?
- (c) In vielen späteren Arbeiten der Transformationsgrammatik seit CHOMSKY (1968) wird die Wichtigkeit von Begriffen wie „Fokus“ und „Presupposition“ für die Intonation betont. Dafür ist hier keine Stelle vorgesehen.

All dies sind Punkte — und es gibt viele ähnliche — die die interne Entwicklung der Transformationsgrammatik betreffen; sie würden Anpassungen verlangen, und es ist offen, ob solche Anpassungen ohne weiteres möglich sind. Man kann aber auch eher externe Einwände bringen:

- (d) Mit ist völlig unklar, aufgrund welcher Intuitionen die Akzentregeln gerechtfertigt werden. Die ausführlichste Arbeit dazu, soweit das Deutsche betroffen ist, stellt KIPARSKY (1966) dar (vgl. dazu auch Bierwisch (1968)); ich weiß nicht, wie KIPARSKY zu seinen Regeln kommt. Sie können ja nicht unmittelbar aus irgendwelchen Daten, z.B. Satzkorpora, abgeleitet sein; erstens würde ein echter Transformationsgrammatiker dies zweifellos ablehnen, und zweitens hat man in solchen Daten ja immer zugleich den Niederschlag der Intonation mit denselben phonetischen Parametern wie Schallstärke, Grundfrequenz, Dauer. Wir haben zwar alle eine recht klare Intuition, daß etwa das Wort „Maurer“ auf der ersten Silbe betont ist, das Wort „Betrug“ hingegen auf der zweiten; aber diese Intuition habe ich auch bei „gestern“ und „sofern“, die überhaupt keinen Akzent (im Sinne der Theorie) enthalten. Und zum ändern hilft diese Intuition nicht bei der Formulierung der zyklischen Akzentregeln selbst. Worauf also stützen sie sich?

2. Der zweite Einwand ist ganz einfach: Es ist mir nicht klar, ob die Regeln stimmen. Ist es z.B. in der Tat so, daß - wie Regel (Aiii) besagt — die Führtöne allmählich abfallen? BIERWISCH stützt sich bei der Formulierung seiner Regeln weithin auf von ESSEN, daneben auch auf eigene Einzelbeobachtungen. Über die empirischen Analysen von ESSENS braucht nichts weiter gesagt zu werden; die zusätzlichen Beobachtungen, deren Wichtigkeit gar nicht in Zweifel gezogen werden soll, sind zum einen sporadisch, und zum andern gibt es nach meinen Erfahrungen kaum ein Gebiet der Grammatik, bei dem man mit seinen Intuitionen so schnell durcheinandergerät wie die Intonation. Vielleicht habe ich auch nur besonders schlechte Ohren, und der „trainierte Ohrenphonetiker“ hat diese Probleme nicht. Aber zum einen hat LIEBERMANN (1965) schlagend nachgewiesen, daß auch trainierte Linguisten inkonsistent sind und vor allem hören, was ihnen ihre jeweilige Theorie sagt. Und zum zweiten kann ich in instrumentellen Transkriptionen einfach nicht wiederfinden, was da sein sollte — auch wenn man in Rechnung stellt, daß unsere Wahrnehmung von den physikalischen Eigenschaften des Signals oft abweicht; zumindest in groben Zügen folgt sie dem Signal.

7 **Schluß**

Ziel dieses Aufsatzes war es, die wichtigsten Ansätze zur Beschreibung der deutschen Satzintonation kurz und kritisch darzustellen. Wenn mir das bis hierhin nicht gelungen ist, wird auch eine Zusammenfassung nichts daran ändern. Stattdessen möge sich der geneigte Leser überlegen, nach welchem Verfahren er die beigefügten maschinell transkribierten Sätze am besten analysieren würde; der erste dieser beiden Sätze ist der oben besprochene „ein Zigeuner hat einen Traktor gekauft“, der zweite ist die entsprechende Frageform „hat ein Zigeuner einen Traktor gekauft?“. Angegeben ist der Verlauf der Grundfrequenz, die im großen und ganzen der Tonhöhe entspricht; die Ziffern unter den Silben geben die Rangfolge der Intensität wieder, d.h. 1 hat die höchste Intensität, 2 die zweithöchste usw.; 1 cm entspricht etwa 40 Millisekunden. Die Sätze sind „normal“, d.h. ohne besondere Emphase und ohne „Emotionalität“ gesprochen.³⁵⁾

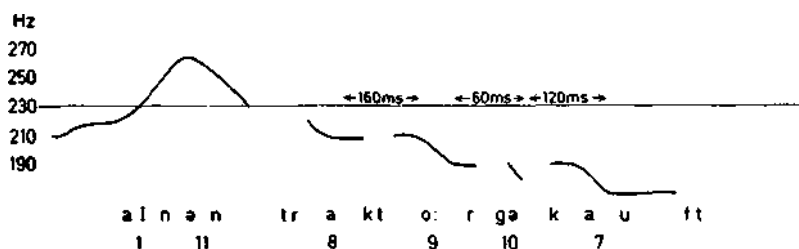
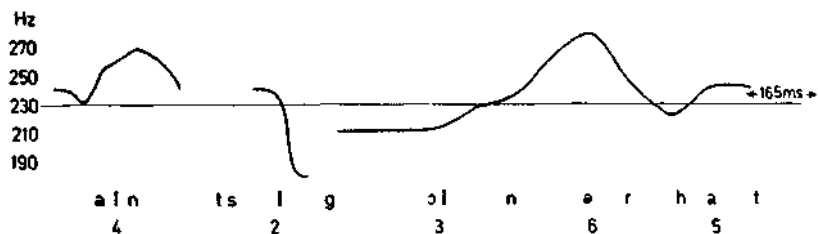


Abbildung 1

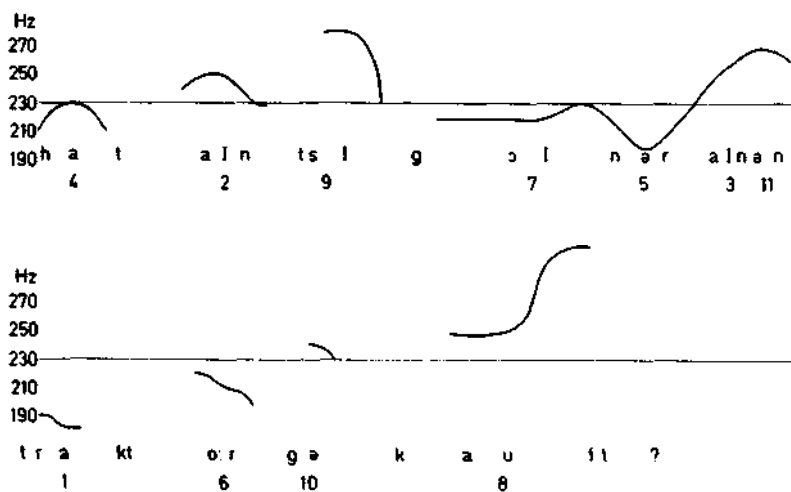


Abbildung 2

Bibliographie

Im folgenden werden neben den zitierten Arbeiten einige weitere, die ich für lesenswert halte, aufgeführt. Ausführliche Bibliographien — allerdings nicht speziell zum Deutschen finden sich in Crystal (1969), Di Cristo (1975) und Wunderli u.a. (1978).

- BALD, W. D. (1976): Contrastive studies in English and German intonation: a survey. In: Papers and Studies in Contrastive Linguistics, Vol. 4, S. 37-47 (Posnan 1976).
- BIERWISCH, M. (1966): Regeln für die Intonation deutscher Sätze. In: Studia grammatica VII, S. 99-201.
- BIERWISCH, M. (1968): Two critical problems in accent rules. In: Journ. Linguistics 4, S. 173-178.
- BOLINGER, D. L. (1972), (Hrsg.), Intonation. Harmondsworth.
- BRESNAN, J. W. (1971): Sentence stress and syntactic transformations. In: Language 47, S. 257-281.
- BRESNAN, J. W. (1972) Stress and syntax - a reply. In: Language 48, S. 326-342.
- BRINKMANN, H. (1962. ²1971): Die deutsche Sprache. Düsseldorf.
- CHOMSKY, N. (1968): Deep Structure, Surface Structure, and Semantic Interpretation. In (unter anderem): Chomsky, N., Studies on Semantics in Generative Grammar. Den Haag 1972, S. 62-119.
- CHOMSKY, N. & HALLE, M. (1968): The sound pattern of English. New York.
- CRYSTAL, D. (1969): Prosodic Systems and Intonation in English. Cambridge, Engl.
- DANES, F. (1960): Sentence intonation from a functional point of view. In: Word, 16, S. 34-54.
- DELATTRE, P. C. (1963): Comparing the prosodic features of English, German, Spanish and French. In: Int Review of Applied Linguistics 1, S. 193-210.
- DELATTRE, P. C., POENACK, E. & OLSEN, C. (1965): Some characteristics of German intonation for the expression of continuation and finality. In: Phonetica 13 (1965), S. 134-161.
- DI CRISTO, A. (1975): Soixante et dix ans de recherches en prosodie. Aix-en-Provence 1975.
- ERBEN, J. (1961): Abriß der deutschen Grammatik. Berlin (mit zahlreichen späteren Auflagen).
- VON ESSEN, O. (1956): Grundzüge der hochdeutschen Satzintonation. Ratingen (2. Auflage 1964).
- FLÄMIG, W. u. a. (1972): Skizze der deutschen Grammatik. Berlin.
- FRY, D. B. (1968): Prosodic Phenomena. In: B. Malmberg (Hrsg.), Manual of Phonetics. Amsterdam. S. 365-410.
- HALLIDAY, M. A. K. (1967): Intonation and grammar in British English. Den Haag 1967
- HEIKE, G. (1969): Suprasegmentale Analyse. Marburg.
- HERMANN, L. (1942): Probleme der Frage. In: Nachrichten der Akademie der Wissenschaften, Göttingen. Phil.-hist. Klasse, S. 121-408.
- HOFFE, W. L. (1960): Über Beziehungen von Sprachmelodie und Lautstärke. In: Phonetica 5, S. 129-158.
- ISAČENKO, A. V. & SCHÄDLICH, H. J. (1966): Untersuchungen über die deutsche Satzintonation. In: Studia grammatica 7, S. 7-67.
- ISAČENKO, A. V. & SCHÄDLICH, H. J. (1970a): A model of standard German intonation. Den Haag (erstmalig deutsch 1964).

- ISAČENKO' A. V. & SCHÄDLICH, H. J. (1970b): Die deutsche Satzintonation als Zweistufensystem. In: Kleine Enzyklopädie „Die deutsche Sprache“. Leipzig. S. 1008-1014.
- VAN KATWIJK, A. (1974): Accentuation in Dutch. Amsterdam.
- KIPARSKY, P. (1966): Über den deutschen Akzent. In: Studia grammatica VII, S. 69-98.
- KOHLER, K. J. (1977): Einführung in die Phonetik des Deutschen. Berlin.
- KUHLMANN, W. (1931): Die Tonhöhenbewegung des Aussagesatzes. Diss. Freiburg.
- KUHLMANN, W. (1952): Vergleich deutscher und englischer Tonhöhenbewegung. In: Z. Phon. 6, S. 195-207.
- LASS, N. J. (1976), (Hrsg.): Contemporary issues in experimental phonetics. New York.
- LEHISTE, I. (1970): Suprasegmentale Cambridge, Mass.
- LEON, P. R. & MARTIN, P. (1969): Prolegomènes a l'étude des structures intonatives. Montreal (Einführung in Forschungsstand und Technik).
- LIEBERMAN, P. (1965): On the acoustic basis of the perception of intonation by linguists. In: Word 21, S. 40-54.
- LIEBERMAN, P. (1967): Intonation, Perception and Language. Cambridge, Mass.
- LIEBERMAN, P. (1974): A study of prosodic features. In: T. Sebeok (Hrsg.), Current Trends in Linguistics 12, Den Haag. S. 2419-2449.
- PILCH, H. (1977): Intonation in discourse analysis. In: Phonetica 34, S. 81-92.
- SCHMIDT, W. (1967): Grundfragen der deutschen Grammatik. Berlin.
- STOCK, E. (1970a): Der gesprochene Satz. In: Kleine Enzyklopädie „Die deutsche Sprache“. Leipzig. S. 994-1008.
- STOCK, E. (1970b): Grundfragen der Melodisierung im deutschen Aussagesatz. Unveröffentl. Habilitationsschrift. Halle.
- STOCK, E. & ZACHARIAS, C. (1971): Deutsche Satzintonation. Leipzig. (2. Aufl. 1973).
- WUNDERLI, P. - BENTHIN, K. - KARASCH, A.: Französische Intonationsforschung. Tübingen.

Fußnoten

- ¹⁾ Die zweite Auflage, nach der ich zitiere, unterscheidet sich nur unwesentlich von der ersten. Eine knappe Darstellung gibt von Essen in seiner „Allgemeinen und angewandten Phonetik“, Berlin 1966 (dies ist die von mir benutzte Auflage), S. 197-208.
- ²⁾ An anderer Stelle (S. 18/19) verweist von Essen auf zwei Analysen mit dem Siemens Oszilomink und dem "pitch meter" von Frøkjær Jensen; die betreffenden Analysen sind auch im Anhang wiedergegeben. Es wird aber nicht gesagt, ob auch die erwähnten Transkriptionen, aus denen die Intonationsregeln abgeleitet sind, auf eine derartige Analyse zurückgehen. Ich bezweifle dies jedoch.
- ³⁾ Ganz am Rande sei vermerkt, daß dies selbstverständlich überhaupt nicht so ist, wie in vielen Untersuchungen gezeigt werden konnte.
- ⁴⁾ Auch dies ist falsch, wenn damit gemeint sein sollte, daß sie nicht starke Tonhöhenbewegung haben können.
- ⁵⁾ Es ist unklar, ob der Ausdruck „Akzentsilbe“ auch bei einsilbigen Wörtern gebraucht wird. Vor allem wird aber auch nicht gesagt, ob *jedes* Wort (bzw. jedes mehrsilbige Wort) eine Akzentsilbe hat, oder z.B. nur derjenigen, die einen „Begriff bezeichnen, der hervorgehoben werden soll.“
- ⁶⁾ Ich betone dies deshalb, weil z.B. in der von der Transformationsgrammatik herkommenen Intonationsforschung sehr stark die Rolle der Syntax für die Akzentabstufung berücksichtigt wird.

- 7) Es wird hier also davon ausgegangen, daß einer Silbe jeweils genau ein Ton zugeordnet werden kann. Nun zeigt jede etwas genauere Transkription, daß eine Silbe eine beträchtliche interne Tonhöhenbewegung haben kann; selbst wenn man nicht der Ansicht ist, daß die *genaue* Lage der Auf- oder Abwärtsbewegung oft ausschlaggebend ist (vgl. unten Abschnitt 6), bleibt die Frage, welcher Teil der Silbe dann für die Bemessung der Tonhöhe zählt.
- 8) Dies ist nicht ganz genau zu nehmen, denn sonst könnte es maximal vier hervorgehobene Silben geben, weil nur vier Tonhöhen unterschieden werden. Gemeint ist es besteht eine allgemeine Tendenz des langsamen Tonhöhenabfalls.
- 9) Davon kann, um bestimmter rhetorischer Wirkungen willen, abgewichen werden. Darauf gehe ich hier nicht ein, und zwar deshalb nicht, weil um bestimmter rhetorischer Wirkungen willen fast alles möglich ist
- 10) Auch dies wird, wie schon gesagt, noch durch spezifische rhetorische Abweichungen etwas aufgeweicht; das sei aber hier unberücksichtigt.
- 11) Es spielt für die Betrachtung hier keine Rolle, ob man „schrüb“ als hervorgehoben oder als nicht hervorgehoben betrachtet; hier wird letzteres angenommen.
- 12) Auch wenn man annimmt, daß es einen solchen Unterschied gibt, betrifft er Tempus oder Aspekt und ist für das vorliegende Problem belanglos. - Ein paralleles Beispiel findet sich übriggend im BIERWISCH (1968); dort geht es allerdings um ‚stress‘.
- 13) Nach van KATWIJK (1974) und anderen sind solche Tonhöhenbewegungen unter bestimmten, näher zu spezifizierenden Bedingungen maßgeblich für die Wahrnehmung eines „Akzents“ - jedenfalls im Holländischen. Fürs Deutsche kenne ich keine entsprechenden Untersuchungen; vgl. aber etwa KOHLER (1977), S. 125.
- 14) Es gibt in einer Reihe von Fällen auch die Möglichkeit, den interrogativen Charakter einer Äußerung schon früher anzuzeigen; wir wollen hier aber einmal die von den Essen angegebene „interrogative Intonation“ als Normalfall betrachten.
- 15) Ich beziehe mich im folgenden stets auf die letzte Auflage der Dudengrammatik; der einschlägige Teil ist aber von dem der früheren Auflagen nicht sehr verschieden.
- 16) Übrigens spricht WINKLER in der Ausgabe von 1973 durchweg von „Ausspruch“, wo es 1966 noch „Satz“ hieß; geblieben ist jedoch das „Satz“ im Titel.
- 17) In der Ausgabe von 1966 heißt es übrigens „zweiten“.
- 18) Dies gilt für den „ruhig gesprochenen Aussagesatz“ (S. 661); bei Emphase kann „der Drang zur Ausdrucksentladung den Sinnkern oft an den Anfang rücken (S. 647). - Es ist etwas verwirrend, daß Winkler unmittelbar zuvor „ruhige Rede“ und „Aussagesatz“ gleichzusetzen scheint: „Bei ruhiger Rede, das heißt hier im Aussagesatz,...“ (S. 646). Mir ist nicht klar, wie das auszulegen ist
- 19) Stock hat gemeinsam mit C. ZACHARIAS 1971 eine selbständige Monographie vorgelegt, die aber als praktisches Lehrbuch für Deutschlernende angelegt ist Nicht zugänglich war mir leider STOCK (19970b).
- 20) Mehr kursorisch werden sie auch an anderen Stellen gestreift
- 21) Er nennt ferner Ton, der uns hier nicht zu interessieren braucht da er im Deutschen irrelevant ist Allerdings wird „Ton“ später auch im Sinne von „Intonationsmuster“ verwendet Darauf kommen wir noch zurück.
- 22) Es scheint mir kein gutes Zeichen für den Stand der deutschen Phonetik, daß völlige Unklarheit darüber herrscht was das signalphonetische Gegenstück des „Akzents“ ist. KOHLER ist, bei aller Klarheit sonst, in diesem Punkt nicht bestimmter als WINKLER. Vgl. etwa HEIKE (1969), S. 56-71, der allerdings mit künstlichem Material arbeitet.
- 23) KOHLERS Beschreibung ist eine Adaption des Systems von HALLIDAY (1967) ans Deutsche. Sie weicht allerdings in vieler Hinsicht von HALLIDAY ab.

- ²⁴⁾ ISAČENKO und SCHÄDLICH haben ihren Ansatz an verschiedenen Stellen dargestellt: 1966, (1970a), (1970b); wesentliche Unterschiede - außer in der Ausführlichkeit - gibt es nicht; auch die beiden letztgenannten Arbeiten sind bereits früher geschrieben.
- ²⁵⁾ Dies hat keine prinzipiellen, sondern methodische Gründe, vgl. IS S. 40-43 und S. 59, (8).
- ²⁶⁾ Ausgenommen sind einige wenige klitische Elemente.
- ²⁷⁾ IS glauben gezeigt zu haben, daß Lautstärkeveränderungen für die Hervorhebung sekundär sind. Sie können zwar auch so wirken, aber falls Lautstärke und Tonbruch nicht parallel gehen, ist letzterer entscheidend.
- ²⁸⁾ Ich betone dies, weil IS in der Literatur nur selten diskutiert wurden; stattdessen findet sich Kritik an marginalen Punkten. So bemerken WUNDERLICH-BENTHIN-KARASCH (1978) als „Vorbehalt“, das IS zwar von zwei Tonstufen reden, aber genauso gut könne man von relevanten Bewegungen sprechen. Das sagen IS selber, aber das hat ja nichts mit der Richtigkeit ihrer Theorie zu tun (S. 126, S. 138; vgl. jedoch auf die etwas eingehendere Diskussion (S. 182-184)). LIEBERMAN (1974) kritisierte die nicht gemachte Behauptung, bereits Unterschiede von 5 Hz seien relevant IS sprechen davon, daß ein Halbton genügt, und auch dies nur, wenn man expressive (emotionelle) Faktoren außer Acht läßt. Daß die Isolierung expressiv und syntaktisch bedingter Schwankungen ein Problem darstellt, versteht sich.
- ²⁹⁾ Die folgenden Bemerkungen über Akzentregeln beschränken sich auf das zum Verständnis Nötigste. Vgl. im einzelnen KIPARSKY (1966) und - obwohl später erschienen - CHOMSKY-HALLE (1968), Kap. 2.1 und 3 (dort werden allerdings nur sehr einfache Phrasen behandelt); ferner BIERWISCH (1968).
- ³⁰⁾ Ich gebe hier und im folgenden jeweils nur den Grundgehalt der Regel wieder, ohne spezifische Zusatzbedingungen und Ausnahmen.
- ³¹⁾ BIERWISCH trägt damit den Beobachtungen von ISAČENKO und SCHÄDLICH über präiktischen und postiktischen letzten Fall Rechnung, interpretiert sie allerdings vorsichtig um. Die genaue Rolle von D bleibt noch zu klären.
- ³²⁾ BIERWISCH faßt sie später zu fünf komplexen Regeln zusammen (S. 153). Es muß noch einmal betont werden, daß meine Paraphrasen natürlich vergrößern.
- ³³⁾ Diese Regel gilt nicht, falls der SIM D vorhanden ist und die PE überhaupt nur einen Führton enthält; dann muß der Verlauf höher liegen als dieser Führton.
- ³⁴⁾ Man kann, als Variante, Aux direkt von S dominieren lassen; das macht im Ergebnis in der Tat oft einen Unterschied, je nachdem, wie man den Parameter p wählt, d. h. wieviel Phrasierungseinheiten man zuläßt.
- ³⁵⁾ Die Transkription wurde am Max-Planck-Institut für Psycholinguistik, Nijmegen, mit einem NSC Ubiquitous Spectrum Analyzer und einem PDP 11/55-Rechner angefertigt. Das Programm stammt von Peter WITTENBURG; gesprochen wurden die Sätze von Charlotte LAUER; ich bedanke mich bei ihnen, ferner bei Joan Leslie für bibliographische Hilfe und bei Marlene ARNS, die das Manuskript mittels einer Schreibmaschine transkribiert hat.