Unterkapitel 3.6 behandelt die **Semantik**, dasjenige Teilgebiet der Sprachwissenschaft, das sich mit der Bedeutung von Sprache beschäftigt. Zunächst werden die Grundlagen der satzsemantischen Analyse anhand der weit verbreiteten Montague-Semantik vorgestellt, wobei die Bedeutung der kleinsten Bestandteile, also der Wörter, als gegeben vorausgesetzt wird. Ausgehend von der Satzsemantik wird die Diskursrepräsentationstheorie DRT, die die Bedeutung ganzer Diskurse erfassen kann, vorgestellt. Ein wichtiges Problem der Computerlinguistik stellt im Rahmen der Semantik die Verarbeitung von Mehrdeutigkeiten dar. Dieses Problem und Ansätze zur Lösung desselben mittels unterspezifizierten Repräsentationen werden im Anschluss diskutiert. Schließlich wird die Grundannahme der Satzsemantik, die Bedeutung von Wörtern als kleinste Einheit nicht weiter zu untersuchen, problematisiert und damit der Bereich der lexikalischen Semantik näher beleuchtet.

Das nächste Unterkapitel 3.7 dieses Methoden-Kapitels behandelt einen sehr heterogenen Teil der (Computer-)Linguistik, der unter dem Begriff **Pragmatik** zusammengefasst ist. In den Abschnitten dieses Unterkapitels werden verschiedene Aspekte angesprochen, die mit kontextuellen Eigenschaften der Sprachanalyse und -verarbeitung zu tun haben. Zum Beispiel wird diskutiert, wie Bezüge innerhalb eines Diskurses hergestellt werden können, welche impliziten Aussagen hinter einer Äußerung stehen und in welcher Hinsicht das Modell eines Benutzers für ein sprachverarbeitendes System relevant ist.

Unterkapitel 3.8 ist der **Textgenerierung** gewidmet, also der Erzeugung von Texten aus semantischen Repräsentationen. Die Generierung kohärenter Texte umfasst mehr als die Generierung aneinandergereihter einzelner Sätze, denn Texte sind satzübergreifend organisiert. Daher erfordert die Textgenerierung Methoden für die Planung globaler Organisationsstrukturen für Texte und entsprechende Mittel für deren sprachliche Umsetzung.

Das letzte Unterkapitel 3.9 im Methodenteil diese Buches widmet sich einer Übersicht über die verschiedenen Programmierparadigmen und **Programmiersprachen**, die in der Computerlinguistik vornehmlich Verwendung finden.

3.1 Phonetik und Phonologie

Dafydd Gibbon

Die **Computerphonologie** und die **Computerphonetik** befassen sich mit der Modellierung und Operationalisierung linguistischer Theorien über die lautsprachlichen Formen und Strukturen der ca. 7000 Sprachen der Welt. Die lautsprachlichen Eigenschaften der Sprachen sind sehr vielfältig, ebenso die theoretischen und methodologischen Ansätze, die in der Linguistik und der Phonetik entwickelt worden sind, um diese Vielfalt zu beschreiben. Einige dieser Theorien haben eine formale Grundlage, sie werden aber in der Linguistik oft recht informell gehandhabt. In der **phonologischen** Literatur hat man es also oft mit recht informellen textuellen Beschreibungen und Visualisierungen von Formen

und Strukturen zu tun, deren formale Beschaffenheit nicht explizit gemacht wird. Eine Interpretation der Literatur zu erreichen, die computerlinguistischen Standards genügt und eine explizite Modellierung und Operationalisierung erlaubt, ist daher oft nicht einfach.

Dieser Beitrag verwendet zwar weitgehend standardsprachliche deutsche und englische Beispiele, beschränkt sich vom Anspruch her jedoch nicht darauf, sondern geht auch auf allgemeinere Aspekte von lautsprachlichen Systemen ein, die nicht nur andere Sprachen, sondern auch dialektale, soziale und stilistische **Aussprachevarianten** einzelner Sprachen betreffen. Die spontansprachlichen Eigenschaften der Varianten des Deutschen sind z.B. teilweise ‚exotisch' im Vergleich zur Standardsprache und gehen weit über die mit der Standardorthographie darstellbaren Zusammenhänge hinaus.

Auf die Fachliteratur wird nicht in den Hauptabschnitten des Beitrags, sondern in einem gesonderten Schlussabschnitt verwiesen.

Es hat sich in den letzten ca. dreißig Jahren herausgestellt, dass die theoretische, methodologische und empirische Vielfalt in der Phonologie und der Phonetik sich mit relativ einfachen formalen Mitteln modellieren und operationalisieren lässt: in erster Linie mit **regulären Modellen** (**endlichen Automaten** und **endlichen Transduktoren**) und mit Attribut-Wert-Strukturen. Die Teildisziplin der Computerphonologie existiert im Prinzip seit den 1970er Jahren, als erste **Intonationsmodelle** auf der Grundlage von endlichen Automaten entwickelt wurden und als auch entdeckt wurde, dass klassische phonologische Regeln sich mit endlichen Transduktoren modellieren lassen. In den 1980er Jahren kamen reguläre Silbenmodelle und die Zweiebenenmorphologie hinzu, in den 1990er Jahren dann reguläre Modelle der Optimalitätstheorie.

Seit den 1980er Jahren gehört auch die Einführung von statistisch gewichteten endlichen Automaten als **Hidden-Markov-Modelle** (**HMM**) in die Sprachtechnologie (s. Unterkapitel 3.2) zum weiteren Umfeld der formalen Modellierung und der computerbasierten Operationalisierung von sprachlautlichen Systemen.

In den 1980er Jahren wurden über den Bereich der regulären Modelle hinaus (und z.T. damit verbunden) Anwendungen der **Attribut-Wert-Logik** in die Computerlinguistik eingeführt, vorwiegend in der Syntax, aber auch für die Modellierung von phonetischen Merkmalen. Auch asymmetrische Markiertheitsrelationen zwischen den Werten phonologischer Attribute (Merkmale) wie bei *stimmhaft* und *stimmlos* konnten mit **defaultlogischen** und **unifikationstheoretischen** Mitteln modelliert werden.

Ziel dieses Beitrags ist es, diese wesentlichen phonetischen und phonologischen Fakten, Generalisierungen und Modellierungsstrategien so einzuführen und zu erläutern, dass die weitergehende Literatur, die am Ende des Beitrags angegeben wird, nutzbar gemacht werden kann. Zuerst soll die empirische sprachlautliche Domäne mit ihren Teildomänen Phonetik, Phonologie und Prosodie besprochen werden, um dann in den folgenden Abschnitten auf die Anwendung empirischer und formaler Methoden in diesen Teildomänen einzugehen. Besondere Aufmerksamkeit wird zum Schluß dem Bereich der lautsprachlichen Eigenschaften der Prosodie gewidmet.

3.1.1 Grundlagen der Computerphonologie

Die Lautlehre wird konventionell in drei Bereiche eingeteilt: Phonetik, Phonologie und Prosodie, die untereinander Abhängigkeiten aufweisen. Als erste Annäherung kann festgehalten werden, dass die **Phonetik** sich mit allen Details der Physiologie der Lautproduktion, der Akustik der Lautübertragung und der Physiologie der Lautrezeption meist mit experimentellen und quantitativen Methoden befasst, während die **Phonologie** sich auf die wortunterscheidenden Lauteigenschaften, die Lautstrukturen und die Relationen zwischen den Lauten und größeren Einheiten wie Morphem, Wort und Satz meist mit symbolverarbeitenden Methoden spezialisiert. Die **Prosodie** wird aus historischen Gründen oft als selbständiger Teil der Lautlehre behandelt. Es ist aber möglich, wie später in diesem Beitrag kurz gezeigt wird, eine systematische und integrative Modellierung der drei Bereiche vorzunehmen, wie sie für die Integration umfassender computerlinguistischer oder sprachtechnologischer Modelle erforderlich ist. Eine Einführung in alle Aspekte der Phonetik, Phonologie und Prosodie ist an dieser Stelle nicht möglich. Dafür wird auf den Literaturabschnitt 3.1.4 verwiesen.

Phonetik

Die Phonetik behandelt die Modellierung aller Eigenschaften von Sprachlauten und wird unterteilt nach Teildomäne und Untersuchungsmethoden. Die wichtigste Teildomänenunterscheidung basiert auf den drei Hauptphasen der Lautverarbeitung entsprechend einem einfachen Kommunikationsmodell (Abbildung 3.1): **artikulatorische Phonetik** (**Produktionsphonetik**), **akustische Phonetik** (**Übertragungsphonetik**) sowie **auditive Phonetik** (**Rezeptionsphonetik**).

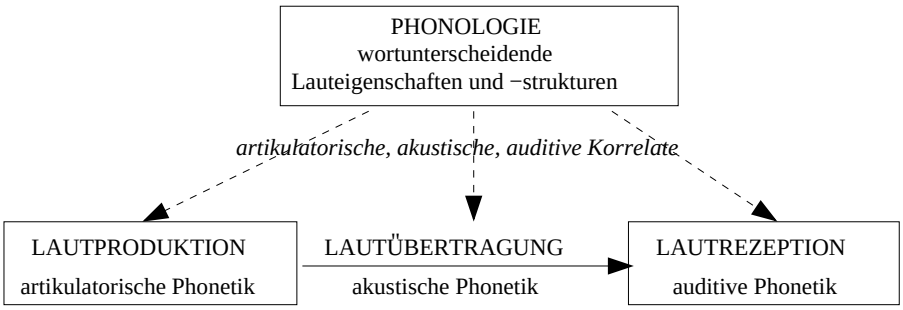

Abbildung 3.1: Korrelatrelationen zwischen der Phonologie und den phonetischen Teildomänen.

Die drei Teildomänen stellen eine hilfreiche, aber sehr vereinfachende Abstraktion dar. Die Teildomänen und die Schnittstellen zwischen ihnen sind wesentlich vielfältiger:

1. Nervensignale zwischen Gehirn und Muskeln,
2. Muskelkonfigurationen in Kehlkopf und Mundraum,
3. Gewebeoberflächenformen in Rachen und Mund,
4. Resonanzraumkonfigurationen in Rachen und Mund,
5. Akustisches Signal,
6. Transformationen im Übertragungsmedium,
7. Transformationen in den Hörorganen (Ohrkanal, Trommelfell, Gehörknöchelchen, Schneckenorgan, Gehörnerven).

Das vereinfachte Modell ist aber für Überblickszwecke nützlich und üblich.

Die drei Teildomänen der Phonetik werden an dieser Stelle überblicksweise erläutert. Für weitere Informationen steht eine reichhaltige phonetische Einführungsliteratur zur Verfügung (siehe Literaturabschnitt 3.1.4).

Laute, die in **phonologischen Kontexten** zitiert werden, sind durch Schrägstriche gekennzeichnet, z.B. /te:/ „Tee“. Laute, die in **phonetischen Kontexten** zitiert werden, sind durch eckige Klammern gekennzeichnet, z.B. [t^{h}e:], um die phonetische Realisierung des /t/ mit behauchtem [t^{h}] darzustellen.

Artikulatorische Phonetik

Als erste Annäherung kann die artikulatorische Phonetik in zwei Bereiche eingeteilt werden: **Schallquellen** und **Schallfilter**. Die Schallquellen bewirken die Erzeugung eines **Klangs** (eines harmonischen Lauts mit wohldefinierten Obertönen) oder eines **Geräuschs** (eines Lauts ohne regelmäßige Obertonstruktur) oder einer Kombination von beiden. Die Klänge sind **Vokale** und andere stimmhafte Laute, und die **Klangeigenschaft** wird im Kehlkopf durch die rapide Schließung und Öffnung der **Glottis** (Spalte zwischen den Stimmlippen) erzeugt. Die Laute mit **Geräuschanteil** sind die Obstruenten (Plosive bzw. Verschlusslaute und Frikative bzw. Reibelaute), die durch Verschluss und Öffnung bei Zunge-Gaumen-Kontakt, Luftreibung bei Lippenverengung usw. erzeugt werden. Das Filter besteht im Wesentlichen aus zwei Resonanzräumen: dem Mund-Rachenraum und dem Resonanzraum der Nase. Die Anordnung der Artikulationsorgane und Resonanzräume wird in Abbildung 3.2 schematisch wiedergegeben.

Das hier beschriebene Modell wird **Quelle-Filter-Modell** genannt: Eine Schallquelle erzeugt einen komplexen Klang oder ein komplexes Geräusch und die **Intensität** der einzelnen Frequentanteile des komplexen Schalls (aber nicht diese Frequenzen selbst) wird dann durch ein Filter verändert (im Prinzip wie der Equaliser einer Audioanlage). Das Quelle-Filter-Modell genügt zwar nicht einer sehr präzisen akustischen Modellierung, ist aber dennoch hilfreich für ein intuitives Verständnis der Grundprinzipien der Sprachschallproduktion.

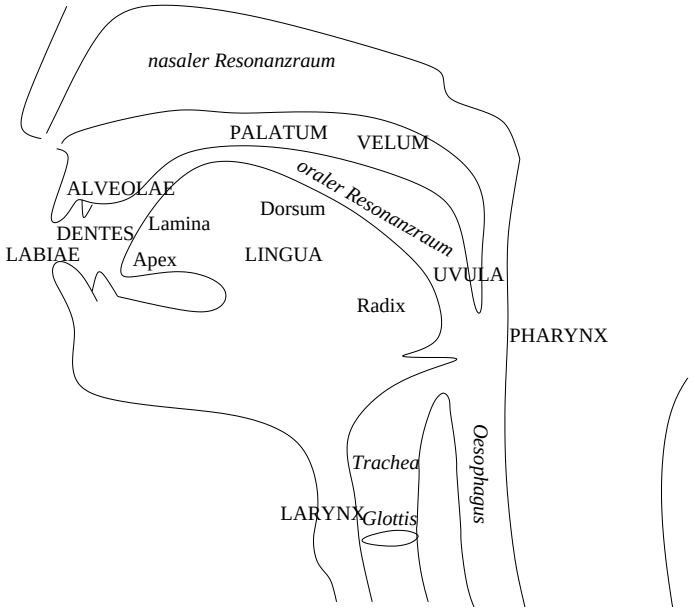

Abbildung 3.2: Schematische Darstellung des Sprechapparates. Artikulationsorgane: Labiae (Lippen), Dentes (Zähne), Lingua (Zunge), Apex (Zungenspitze), Lamina (Zungenblatt), Dorsum (Zungenrücken), Radix (Zungenwurzel), Alveolae (Zahndamm), Palatum (harter Gaumen), Velum (Gaumensegel, weicher Gaumen), Pharynx (Rachenwand), Larynx (Kehlkopf), Oesophagus (Speiseröhre), Trachea (Luftröhre). Resonanzräume: Nasenraum, Mundraum, Glottis (Stimmlippenspalte).

Der Hauptvorgang im Kehlkopf, der Schallquelle für stimmhafte Laute, ist die **Phonation**. In der normalen Phonation werden im Kehlkopf beide Stimmlippen aneinander angenähert und in der Glottis fließt der Luftstrom aus der Lunge schneller. Die Annäherung bewirkt durch den **Bernoulli-Effekt** eine Minderung des Luftdrucks zwischen den Stimmlippen, die diese aneinanderzieht (der Bernoulli-Effekt ist auch für den Auftrieb der Tragflächen von Flugzeugen, den Vorwärtstrieb von Segeln, oder auch die unerwünschte Annäherung von klammen Duschvorhängen an den Körper verantwortlich). Durch die Schließung der Glottis wird der Luftstrom blockiert, der Luftdruck steigt wieder und die Stimmlippen werden wieder auseinandergedrückt. Dieser Zyklus wiederholt sich und bestimmt die Grundfrequenz des Sprachsignals (bei Männern zwischen ca. 70 und 200 Hz, bei Frauen zwischen ca. 140 und 400 Hz, bei Kindern bis ca. 600 Hz). Während der normalen Phonation entsteht, bedingt durch die regelmä-

Quelle	Filter	Beispiel
Stimmbänder	Zunge hoch, weit vorne	**Vokal** [i]
Stimmbänder	Zunge hoch, weit hinten	**Vokal** [u]
Unterlippe, obere Zähne	vorn im Ansatzrohr	**Konsonant** [f]
Unterlippe, obere Zähne; Stimmbänder	vorn im Ansatzrohr	**Konsonant** [v]
Hinterzunge, Zäpfchen	hinten im Ansatzrohr	**Konsonant** [r]

Tabelle 3.1: Beispiele für Quelle-Filter-Konfiguationen bei der Lautproduktion.

ßige Obertonreihe, ein Signal mit etwa sägezahnförmigem Hüllkurvenverlauf. Die Frequenzen der Obertöne bzw. der Harmonischen sind als ganzzahlige Vielfache der **Grundfrequenz** definiert.

Stimmlose Laute entstehen, wenn die Stimmlippen weit auseinandergehalten werden. Die Phonationsart Flüstern entsteht, wenn die Stimmlippen aneinandergelegt werden und nur ein kleine Öffnung am Ende bleibt. Weitere Phonationsarten sind die Knarrstimme, die Hauchstimme und die Falsettstimme.

Die zweite wichtige Schallquelle ist die Friktion (Reibung) des Luftstroms in einer engen Ritze: Wenn ein bewegliches Artikulationsorgan (z.B. die Zungenabschnitte, die Unterlippe) an ein statisches Artikulationsorgan (Oberlippe, Oberzähne, Zahndamm, Gaumen) so angelegt wird, dass nur eine sehr schmale Ritze bleibt, entsteht ein Reibungsgeräusch, die Frikative (Reibelaute) wie [f, v, s, z, ʃ, ʒ, x, h] und Affrikate wie [pf, ts] charakterisiert. Die Frikative [s, z, ʃ, ʒ] entstehen an den scharfen Zahnkanten, enthalten daher besonders hochfrequente Anteile und werden Sibilanten (Zischlaute) genannt.

Der durch eine Quelle erzeugte Schall wird in den durch die Sprechorgane geformten Resonanzräumen in Mund und Rachen gefiltert.

Die wichtigsten **Konsonanttypen** nach dem Quelle-Filtermodell werden in Tabelle 3.1 gezeigt. Bei **Konsonanten** unterscheidet man beispielsweise folgende Hauptmerkmale:

Quelle: Stimmbänder (bei stimmhaften Lauten, Klängen); Zunge oder Lippen an statisches Artikulationsorgan (bei **Geräuschen**, Verschlusslauten, Reibelauten),

Art und Weise: Verschlusslaute, z.B. [p, b, t, d, k, g], Reibelaute, z.B. Frikative [f, v, ʃ, ʒ, ɕ, x, h];

Stimmhaftigkeit: Stimmlose Laute, Stimmbänder weit auseinander, nicht schwingend (stimmlose Verschluss- und Reibelaute); Stimmhafte Laute, Stimmbänder aneinandergelegt, schwingend (**Vokale**, **Nasalkonsonanten** [m, n], Liquiden [r, l], Gleitlaute [w, j]).

Bei **Vokalen** gestaltet sich das Quelle-Filter-Modell etwas einfacher: Die Quelle ist bei normalen Vokalen die Glottis mit der normalen Phonation, das Filter ist der in Form und Größe durch Zunge und Lippen variierbare Mundraum und,

bei nasalen Vokalen (und Konsonanten), der ein- und ausschaltbare und (außer durch Schnupfen) nicht variierbare Nasenraum.

Eine umfassende Notation für die artikulationsphonetischen Eigenschaften der Laute der Sprachen der Welt wird in der Symboltabelle der **Internationalen Phonetischen Assoziation** (**IPA**) bereitgestellt (vgl. Abbildung 3.15 am Ende dieses Beitrags). Der IPA-Tabelle liegt ein implizites und mit attribut- und defaultlogischen Mitteln formalisierbares universelles Modell von Lautobjekten zugrunde, deren Defaulteigenschaften in der **Konsonantentabelle** und dem **Vokaldiagramm** dargestellt werden. Die Defaulteigenschaften können mit diakritischen Zeichen durch speziellere Eigenschaften der Laute einzelner Sprachen überschrieben werden. Dieses Modell kann aus theoretischen wie empirischen Gründen kritisiert werden, aber es gilt mangels bewährter Alternativen als allgemein akzeptierter *de facto*-Standard und wird von den Experten der Internationalen Phonetischen Assoziation gepflegt. Das IPA wird nicht nur in der Allgemeinen Linguistik zur Beschreibung der Lautsysteme der Sprachen der Welt, sondern auch im Sprachunterricht, in der Lexikographie, in der klinischen Phonetik und in der Sprachtechnologie verwendet.

Für die im IPA enthaltenen Symbole existieren unzählige, miteinander nicht kompatible und auch nur partielle Fontimplementierungen, z.B. viele TTF-Fonts, das LaTeX-Font `tipa`; auch der Unicode-Standard ist nur partiell implementiert. Für die praktische Verwendung des IPA in Veröffentlichungen in der Computerphonologie, Computerphonetik und Sprachtechnologie ist daher nicht so sehr das Objektmodell des IPA, sondern die fehlende Systematisierung der bisher implementierten Fonts problematisch. Um den praktischen Datenaustausch und die einfache Verarbeitung phonetischer Daten zu ermöglichen, wurde in den 1980er Jahren im europäischen Forschungsprojekt SAM (*Speech Assessment Methods*) von Phonetikern und Sprachingenieuren eine tastaturfreundliche ASCII-Kodierung des IPA entwickelt, **SAMPA** (‚SAM Phonetic Alphabet'). Das SAMPA-Alphabet ist immer noch sehr verbreitet, weil die Unicode-Zeichenkodierung nur maschinenlesbar und eher an Druckausgabe als an Eingabeergonomie orientiert ist, weil derzeit nur unvollständige Unicode-Font-Implementierungen existieren, und auch weil Unicode die kohärente phonetische ‚Semantik' der Zeichen nicht berücksichtigt, sondern die Verwendung von Zeichen aus verschiedenen Bereichen erfordert.[1]

Akustische Phonetik

Die Schallschwingungen, die durch die artikulatorischen Quellen und Filter erzeugt wurden, werden durch die Luft und andere Medien als komplexe Druckwellen übertragen, die mit handelsüblichen Aufnahmegeräten aufgenommen werden. Das Sprachsignal wird mit den Methoden der akustischen Phonetik auf unterschiedliche Weise analysiert und dargestellt, wovon Folgende die drei wichtigsten sind:

[1] Vgl. die Internetseiten von John Wells, University College, London, zu IPA-Fonts, SAMPA und Unicode.

1. Als Signal in der **Zeitdomäne**, das als **Oszillogramm** visualisiert wird, in dem die regelmäßigen **Klangabschnitte** und die unregelmässigen **Geräuschabschnitte** des Sprachsignals relativ leicht erkennbar sind.
2. Als transformierter Signalabschnitt in der **Frequenzdomäne**, das als **Spektrum** visualisiert wird, das die **Energien** der Frequenzanteile des komplexen Signals als Funktion der Frequenz anzeigt. Am Spektrum können **Klanganteile** (**Grundfrequenz** und ihre Obertöne) sowie **Geräuschanteile** des Sprachsignals erkannt werden.
3. Als dreidimensionale Darstellung von Sequenzen von zeitlich benachbarten Spektren in einem **Spektrogramm**. Am Spektrogramm können die gegenseitigen Beeinflussungen der Lautproduktionsvorgänge (Koartikulation) studiert werden.
4. Als **Grundfrequenzspur**, die die niedrigste Frequenz (**Grundfrequenz**, F0) eines Klangs als Funktion der Zeit darstellt, deren ganzzahlige Vielfache die Obertonreihe definieren.

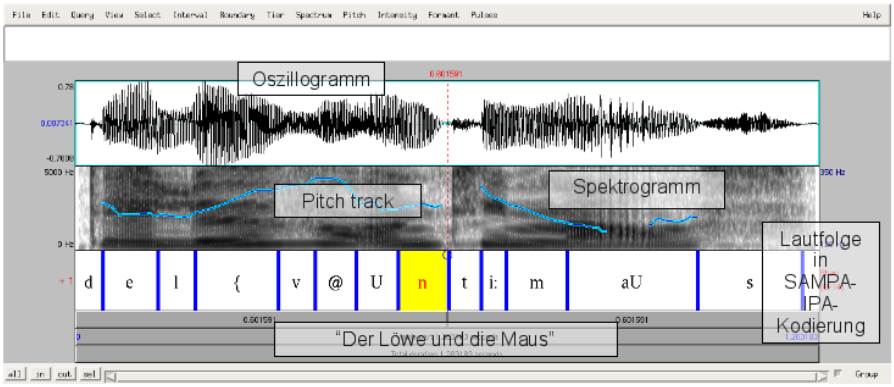

Abbildung 3.3: Visualisierungen eines Sprachsignals (weibliche Stimme) mit der Praat-Software: Oszillogramm, Spektrogramm, Grundfrequenzverlauf (pitch track), Lautfolge in SAMPA-Kodierung.

In Abbildung 3.3 können die Vokale anhand ihrer größeren Amplitude und der regelmäßigen Struktur des Oszillogramms sowie der dunkleren Bereiche des Spektrogramms (**Formantstreifen**: Verstärkungen einzelner Frequenzbereiche durch die Filterwirkung der Resonanzräume des Artikulationstrakts) erkannt werden. Das Frikativgeräusch des [s] in *Maus* ist durch die geringere Amplitude und unregelmäßige Schwingungen im Oszillogramm sowie durch die etwas dunklere Färbung in den hochfrequenten Bereichen des Spektrogramms, ebenfalls mit Formantstreifen, zu erkennen. Im Beispiel ist auch ein Abschnitt mit Knarrstimme (die spitzen Ausschläge beim SAMPA-kodierten Diphthong [aʊ]) zu sehen.

Für weitere Beschreibungen der zahlreichen Transformationen, Darstellungsweisen und Analyseverfahren für das Sprachsignal wird auf die Spezialliteratur im Literaturabschnitt 3.1.4 verwiesen; siehe auch Unterkapitel 3.2).

Auditive Phonetik

Die auditive Phonetik befasst sich mit den Vorgängen im Ohr und ist im Gegensatz zur artikulatorischen Phonetik nicht der direkten Beobachtung zugänglich. Im Gegensatz zur akustischen Phonetik ist die auditive Phonetik nicht für relativ einfache Messungen zugänglich, sondern benötigt eine Zusammenarbeit mit Fachmedizinern.

Die auditive Phonetik ist von herausragender Bedeutung in der klinischen Phonetik für die Diagnose und Therapie sowie in Zusammenarbeit mit der Sprachtechnologie und der Hörakustik für die Entwicklung von prothetischen Vorrichtungen wie Hörgeräte und Cochlearimplantate.

Aufgrund der Unzugänglichkeit dieser Teildomäne der Phonetik (außer in enger Zusammenarbeit mit klinischen Phonetikern, Medizinern und Hörakustikern) wird diesem Bereich der Phonetik keine weitere Aufmerksamkeit an dieser Stelle geschenkt. Es wird stattdessen auf die weiterführende Spezialliteratur verwiesen (vgl. den Literaturabschnitt 3.1.4).

Phonologie

Die Phonologie behandelt die Funktion, Struktur und die für die Wortunterscheidung wesentlichen Eigenschaften der Grundobjekte der Lautsprache sowie der Strukturen, zu denen diese Objekte kompositionell zusammengesetzt werden. Darüber hinaus behandelt die Phonologie die Abbildung dieser Objekte, ihrer Eigenschaften und ihrer Strukturen auf artikulatorische, akustische und auditive Korrelate in den drei phonetischen Teildomänen (artikulatorische Gesten, akustische Ereignisse und auditive Vorgänge). Die Grundeinheiten der Lautsprache bilden eine Hierarchie, die manchmal als **prosodische Hierarchie** bezeichnet wird: das **Phonem**, aus denen **Silbenkonstituenten** Anlaut, Reim, Kern und Auslaut) zusammengesetzt werden, die **Silbe**, das phonologische **Wort**, der prosodische **Takt**, und **Intonationseinheiten** verschiedener Größe.

Phoneme

Das Phonem ist die kleinste wortunterscheidende Lauteinheit und gilt traditionell als das wichtigste Grundobjekt der Phonologie. Die **Phoneminventare** der Sprachen der Welt unterscheiden sich sehr in der Größe (ca. 20 bis 50) und in den Elementen des Inventars. Generalisierungen (implikative Universalien) sind möglich: z.B. wenn Frikative in einer Sprache vorhanden sind, dann sind auch Verschlusslaute in der Sprache vorhanden. Zur Unterscheidung von der normalerweise detaillierteren Repräsentation von Lauteinheiten in der Phonetik werden Phoneme nicht in eckige Klammern, z.B. [p^h], sondern zwischen Schrägstriche gesetzt, z.B. /p/.

In der Einführungsliteratur wird ein Phonem oft als ‚die kleinste bedeutungsunterscheidende Einheit' definiert. Als formale Definition taugt diese Formulierung nicht viel, weil ein Definiens voraussetzt, dass die in ihm enthaltenen Terme entweder evident oder bereits definiert sind. Dies gilt für ‚Bedeutung' jedoch nicht, erklärt auch nicht z.B., wie Unsinnswörter oder noch nicht gelernte, nicht verstandene aber bereits existente und durch ihre Form identifizierbare Wörter unterschieden werden. Da die Einheit ‚Wort' ein formales Objekt bezeichnet, das zumindest in sehr vielen Sprachen eine intuitiv relativ gut angebbare Einheit ist, die eine Bedeutung haben kann oder auch nicht, eignet sich folgende Definition besser:

> Ein Phonem ist die kleinste sequentielle wortunterscheidende Lauteinheit.

Aber auch diese Definition ist noch relativ dürftig, vor allem, weil das Phonem eine Generalisierung bzw. Abstraktion von den phonetischen Äußerungsdetails und nicht direkt im Sprachsignal beobachtbar ist. Die kleinsten wortunterscheidenden Segmente im Sprachsignal sind **Phone**; wenn Phone in unterschiedlichen **Silben**- oder **Wortkontexten** vorkommen (komplementär verteilt sind) und im Vergleich zu anderen Phonen phonetisch ähnlich sind, gelten sie als **Allophone** desselben Phonems.

Geeigneter als diese eindimensionale Definition ist eine komplexere Charakterisierung des Phonems als semiotische Einheit. Wie andere sprachliche Einheiten kann ein Phonem als **Zeichen** verstanden werden, das anhand der semiotischen Dimensionen Struktur (intern und extern) und Interpretation (semantisch und phonetisch) definiert wird:

Struktur: Sprachliche Zeichen haben zwei Strukturdimensionen, die einerseits ihre interne Zusammensetzung und andererseits den externen Kontext, in dem sie vorkommen, kompositionell bestimmen:

- **Interne Struktur:** Phoneme sind die kleinsten sequentiellen wortunterscheidenden Lauteinheiten und haben als solche keine interne sequentiell-temporale Struktur. Sie werden aber auch als Mengen von distinktiven Eigenschaften aufgefasst. Eine traditionelle Definition lautet demnach: Phoneme sind Bündel von distinktiven **Merkmalen** und haben eine simultan-temporale Struktur. Die Merkmale übernehmen dann die Funktion der kleinsten wortunterscheidenden Lauteinheiten.
- **Externe Struktur:** Phoneme sind die kleinsten Bestandteile von **Silben** (die ihrerseits als Bestandteile von größeren Einheiten in der prosodischen Hierarchie definiert werden).

Interpretation: Zeichen haben in einer modellorientierten Sicht zwei Interpretationen im Hinblick auf eine wahrnehmbaren Realität, die als Interpretationspaar den Kern des **semiotischen** Charakters des Zeichens bilden: die mediale Interpretation („Bezeichnendes") und die semantisch-pragmatische

Interpretation („Bezeichnetes“). Für Phoneme heißt dies, dass es sich um das Paar aus der phonetischen und der semantischen Interpretation der Äußerung handelt:

Phonetische Interpretation: Phoneme werden je nach Position in der externen Struktur als unterschiedliche Allophone (Phon, die einem Phonem zugeordnet werden) interpretiert, die die kleinsten temporalen Segmente von sprachlichen Äußerungen sind. Zum Beispiel wird das Phonem /p/ in „Panne“ behaucht (aspiriert) ausgesprochen und phonetisch als [p^{h}] dargestellt, in „Spanne“ wird das Phonem aber unbehaucht ausgesprochen und folglich phonetisch mit dem Default-Symbol [p] geschrieben. Diese extensionale oder denotationelle Definition durch Interpretation lautet also: Ein Phonem wird durch eine Menge von Allophonen interpretiert, beispielsweise /p/ $=_{def}$ {[p], [p^{h}]} (tatsächlich hat /p/ noch weitere **Allophone**).

Semantische Interpretation: Phoneme haben die Funktion, Wörter zu unterscheiden. Kombinationen von wenigen Phonemen können viele Tausende einfache Wörter kodieren (die ihrerseits durch morphologische Kombinationen weitere Wörter bilden können). Diese Funktion von Phonemen kann also als **Kodierung** definiert werden.

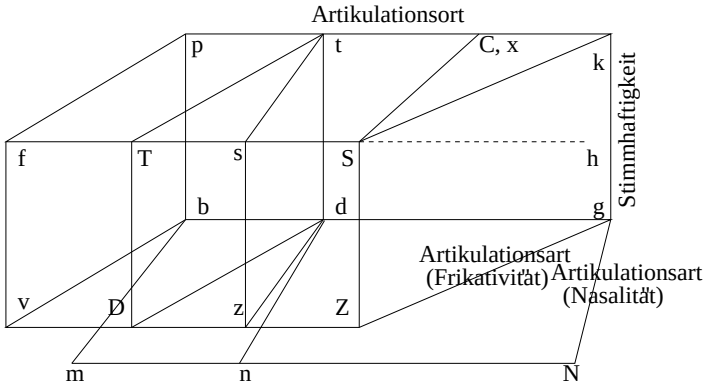

Abbildung 3.4: Paradigmatische Relationen zwischen Konsonanten.

In struktureller Hinsicht gehen Phoneme , wie andere sprachliche Einheiten, zwei Arten von Relationen miteinander ein:

1. klassifikatorische Relationen aufgrund ihrer Eigenschaften, die Ähnlichkeiten und Unterschiede zwischen den Phonemen charakterisieren und traditionell **paradigmatische Relationen** genannt werden,

2. kompositorische Relationen, die das Vorkommen von Phonemen in unterschiedlichen Positionen in Silben bestimmen und traditionell **syntagmatische Relationen** genannt werden.

Die paradigmatischen Relationen zwischen den Konsonanten des Deutschen (ohne Approximanten) werden in Abbildung 3.4 visualisiert. Die syntagmatischen Relationen werden in einem folgenden Abschnitt im Kontext der Silbenstruktur besprochen.

Das Phonem als Grundobjekt ist nicht ganz unkontrovers. In der **Generativen Phonologie** werden abstraktere Generalisierungen, **Morphophoneme**, als einzige abstrakte Grundobjekte angenommen, die direkt phonetisch interpretiert werden. Nach dieser Auffassung ist das Phonem ein Artefakt, das die lautsprachliche Struktur nicht adäquat modellieren lässt. In **prosodischen Phonologien** werden einige Merkmale mit phonematischer Funktion, deren zeitliche Ausdehnung über das einzelne Phonem hinausgeht, als gleichberechtigte Grundobjekte angesehen und **Prosodien** genannt. Die Phoneme oder **phonematischen Einheiten** sind damit unterspezifiziert und müssen durch prosodische Eigenschaften ergänzt werden. Die Stimmlosigkeit von Obstruenten (Plosive, Frikative, Affrikate) im deutschen Silbenauslaut (Auslautverhärtung, z.B. im Auslaut /kst/ von „Axt“ /akst/) wäre demnach eine Prosodie, da sie nicht ein einzelnes Phonem, sondern alle Obstruenten des Auslauts betrifft. In der **Autosegmentalen Phonologie** werden Prosodien, wie z.B. Töne, mit selbständiger klarer Struktur oder Funktionalität Autosegmente genannt und graphisch durch parallele temporal gerichtete Graphen repräsentiert, die durch Bezugskanten (*association lines*) miteinander verbunden werden.

Silbe

Die Silbe ist die kleinste Lautfolge, die als eigenständiges Wort funktionieren kann und besteht in der Regel aus einem Vokal und einem oder mehreren vorangestellten oder nachgestellten Konsonanten. Die Kombinatorik bzw. die Distribution von Phonemen wird im Kontext der Silbe beschrieben. Allerdings wird die Silbe nicht immer als universelles phonologisches Objekt anerkannt, vor allem im Kontext der indo-europäischen Sprachen, die komplexere Lautfolgen aufweisen. Dennoch dient die Silbe als nützliche Vereinfachung beim Verständnis der Lautstrukturen der Sprachen. Es soll hier lediglich angemerkt werden, dass dieser phonologische Silbenbegriff sich vom orthographischen Silbenbegriff unterscheidet.

Aus phonetischer Sicht wird die Silbe etwas anders definiert: Eine Silbe hat eine glockenförmige Sonoritätskontur (etwa: Intensitätskontur), die am Silbenanfang eine geringere Sonorität (Konsonanten) hat, mit dem Vokal die höchste Sonorität erreicht und zum Silbenende eine geringere Sonorität (Konsonanten) hat.

Die Sprachen der Welt unterscheiden sich nicht nur sehr stark in den Phoneminventaren sondern auch in ihrer Kombinatorik im **Silbenkontext**. Die germanischen Sprachen Deutsch, English, Niederländisch, Dänisch, Norwegisch, Schwedisch, Isländisch haben komplexe Silben mit bis zu 8 Phonemen (der Affrikat /pf/ zählt als ein Phonem): „strümpfst“ /ʃtrympfst/ im Kunstwort „bestrümpfst“ (vielleicht: „jemandem die Strümpfe anziehen“) mit der Struktur KKKVKKKK. An den einzelnen K-Stellen können nicht alle Konsonanten vorkommen; an der er-

sten Stelle beispielsweise nur /ʃ/, wie in diesem Beispiel, oder /s/, wie in „Skat“ /skat/. Andere Sprachen haben nur KV- oder KKV-Strukturen, wobei die zweite K-Stelle nur mit den Liquiden /l, r/ besetzt werden kann. Die Möglichkeiten sind sehr vielfältig.
Um **Silbenstrukturen** darzustellen, gibt es viele Notationsarten:

1. Constraints über Lautklassensequenzen:
   V, KV, KVK, VK, ..., KKKVKKK, ...

2. Constraints über Lauteigenschaftssequenzen:
   z.B. Deutsch: wenn $\#X$[Verschlusslaut][Liquid] eine Phonemsequenz beschreibt und X = [Konsonant], dann X = [stimmloser Zischlaut], d.h. in einer Dreikonsonantensequenz am Wortanfang (bezeichnet mit ‘#’) muss der erste Laut ein /s/ oder ein [ʃ] sein.

3. Sonoritätsunterschiede (etwa Unterschiede in der Intensität einzelner Laute):
   $\text{son}_{initial} > \text{son}_{gipfel} > \text{son}_{final}$

4. Eingebettete Strukturen, visualisiert als Baumgraphen oder Klammerungen, z.B. ($_{Anlaut}$ ʃtr ($_{Reim}$ ($_{Kern}$ ɪ) ($_{Auslaut}$ k))) „Strick“.

Alle diese Notationen haben gewisse Vorteile bei der Generalisierung über wesentliche Eigenschaften von Silbenstrukturen, alle haben aber auch Nachteile, weil sie unterschiedliche Abstraktionen darstellen und daher jeweils auf unterschiedliche Weise unvollständige, fragmentarische Modelle sind:

1. Die KVK-Notation generalisiert nicht über die Detailbeschränkungen in der Konsonantenkombinatorik.

2. Die Eigenschaftsnotationen, zu denen traditionelle phonologische Regelnotationen gehören, erfassen jeweils nur Fragmente der Gesamtstruktur.

3. Die Sonoritätsnotation, wie die erste Notation, erfordert weitere Informationen über spezifische Sprachlauttypen und ihre Positionen in der Silbe.

4. Die hierarchischen Notationen bedürfen einer Ergänzung mit zusätzlichen Einschränkungen der linearen Kombinationsmöglichkeiten (besonders im Auslaut), die quer zur Verzweigungsstruktur des Baumgraphen verlaufen.

Diese Notationen, die oft recht informell gehandhabt werden, können dennoch *lokal* explizit und präzise sein. Aber sie lassen trotzdem vieles offen, insbesondere die Fragen der Konsistenz, der Präzision, der Vollständigkeit und der Korrektheit des *globalen* lautsprachlichen Gesamtsystems. In Abschnitt 3.1.3 wird mit regulären Modellen (endlichen Automaten ; vgl. Unterkapitel 2.2) eine computerlinguistisch adäquate Antwort auf dieses Problem in der Form einer vollständigen Modellierung von Silben als Elemente regulärer Mengen gegeben.

Merkmalstheorie

Die Eigenschaften von Sprachlauten werden in der Regel nicht unabhängig voneinander (z.B. ‚stimmhaft', ‚stimmlos', usw.) aufgezählt, sondern zu kontrastierenden Elementen der Wertemengen von Attributen gruppiert. Damit partitioniert jedes Attribut die Menge der Phoneme in Teilmengen, von der jede Teilmenge mit einem der Werte in der Wertemenge des Attributs assoziiert wird. In den traditionellen phonologischen Theorien sind die Attribute binär, d.h. sie haben eine Wertemenge mit zwei Werten, heißen „Merkmale", und werden mit einer Notation dargestellt, in der der Wert dem Merkmal vorangestellt wird: [+ stimmhaft] bedeutet z.B: ‚stimmhaft', [- stimmhaft] bedeutet ‚stimmlos'. In einer in der Computerlinguistik gewohnten Schreibweise können die Merkmale als [Stimmhaft: +] und [Stimmhaft: -], oder expliziter als [Stimmhaftigkeit: stimmhaft] und [Stimmhaftigkeit: stimmlos] ausgedrückt werden. Das Phonem /p/ kann beispielsweise mit einer der gängigen auf artikulatorischen Korrelaten basierenden Merkmalstheorien folgendermaßen definiert werden:

$$/\mathrm{p}/ = \begin{bmatrix} + & \text{konsonantisch} \\ - & \text{vokalisch} \\ - & \text{kontinuierlich} \\ - & \text{stimmhaft} \end{bmatrix}$$

Die phonologische Regel , die die Auslautverhärtung von Obstruenten im Deutschen ausdrückt und dabei über die **natürliche Klasse** von Phonemen {/b/, /d/, /g/, /v/, /z/} generalisiert, lautet in Merkmalsnotation, einmal in der konventionellen unterspezifizierten Form, einmal in der voll spezifizierten Form:

1. $[+ \text{ stimmhaft}] \rightarrow [- \text{ stimmhaft}] \,/\, \begin{bmatrix} \overline{\qquad\qquad} \\ + \text{ konsonantisch} \\ - \text{ vokalisch} \end{bmatrix} \#$

2. $\begin{bmatrix} + & \text{konsonantisch} \\ - & \text{vokalisch} \\ + & \text{stimmhaft} \end{bmatrix} \# \rightarrow \begin{bmatrix} + & \text{konsonantisch} \\ - & \text{vokalisch} \\ - & \text{stimmhaft} \end{bmatrix} \#$

Phonologie und Orthographie

Die Relation zwischen Phonologie und Orthographie ist ein relativ eigenständiger Gegenstandsbereich. Die Schriftsysteme der Sprachen sind in gewisser Hinsicht komplexer und variabler als ihre Phonemsysteme. In logographischen Systemen (z.B. in der chinesischen Orthographie) kodieren die Schriftzeichen Morpheme und nicht Phoneme. Im lateinischen alphabetischen System der europäischen Sprachen werden ebenfalls Logogramme verwendet, aber nur für Zahlen und mathematische Operatoren (z.B. Ziffern 0, ..., 9, Operatoren '+', '-' usw.), die damit sprachunabhängig, aber mit völlig unterschiedlichen Aussprachen verwendet werden können. Einem ähnlichen Prinzip folgen Emoticons in schriftlichen

Kurznachrichten, wobei diese oft auch ikonischen Charakter (Ähnlichkeit zwischen Form und Bedeutung) haben.

Eine alphabetische Orthographie kann streng phonematisch mit einer eineindeutigen Beziehung zwischen Graphemen und Phonemen sein. Dies gilt vor allem für Sprachen, deren Orthographieentwicklung noch relativ neu ist oder mit phonematischer Orientierung reformiert wurde. Bei alphabetischen Orthographien, die bereits seit vielen Jahrhunderten mit relativ wenigen Veränderungen bestehen (vgl. Englisch, Französisch), hat sich die Aussprache weit mehr verändert als die Orthographie, so dass die Phonologie-Orthographie-Relation sehr komplex geworden ist: französisch „eaux“ /oː/ (Plural von „eau“ „Wasser“) und die berühmte englische ‚-ough‘-Reihe: „tough“ /tʌf/, „through“ /θruː/, „cough“ /kɔf/, „though“ /ðəʊ/, „thorough“ /θʌrə/, „bough“ /baʊ/, mit jeweils unterschiedlichen Aussprachen der Reimsequenz „ough“.

Für ältere Sprachstufen vor dem 20. Jahrhundert existieren, wenn überhaupt, nur schriftliche Zeugnisse. Daher kommt der Phonologie-Orthographie-Relation für die Rekonstruktion früherer Sprachstufen, die auch eine herausfordernde computerlinguistische Aufgabe ist, große Bedeutung zu.

Für die Sprachtechnologie hat die Phonologie-Orthographie-Relation ebenfalls eine zentrale Bedeutung in der Form von Graphem-Phonem-Übersetzungsregeln in Sprachsynthesesystemen und automatischen Spracherkennungssystemen (siehe Unterkapitel 3.2).

Die Vielfalt phonologischer Theorien

Das Panorama phonologischer Theorien ist immens, wobei der Eindruck manchmal entstehen kann, dass neue Theorienamen eher aus strategischen als aus theoretischen oder empirischen Gründen eingeführt werden. Es gibt einige nützliche Artikelsammlungen zur Geschichte der Phonologie, die einen guten Überblick ermöglichen (siehe den Literaturabschnitt 3.1.4).

In diesem Abschnitt sollen extrem kurze Charakterisierungen der wichtigsten Richtungen in der Phonologie als kleine Wegweiser für die Literaturrecherche gegeben werden, natürlich bei Gefahr grob unzulässiger Verallgemeinerungen und ohne mit der Spezialliteratur konkurrieren zu wollen.

Strukturalismus: Die strukturalistische Denkweise in der Phonologie wurde von de Saussure in den ersten beiden Jahrzehnten des 20. Jahrhunderts eingeführt. Sie postuliert, dass sprachliche Formen und Strukturen synchron (zu einer bestimmten Zeit) ein zusammenhängendes System von paradigmatischen und syntagmatischen Relationen bilden. Diese Idee stand im Kontrast zu früheren Ansätzen in der komparativen Philologie (diachrone Rekonstruktion früherer Sprachstufen), in der pädagogischen Grammatik, in der Logik, der Rhetorik und in der hermeneutischen Behandlung der Sprache in Philosophie und Theologie. Nach dieser allgemeinen Charakterisierung könnten prinzipiell alle modernen phonologischen Theorien als im weiteren Sinne strukturalistisch angesehen werden. Spätere Entwicklungen änderten diese Grundsätze zwar nicht, führten aber weitere Gesichtspunkte

ein. Repräsentanten unterschiedlicher Ausprägungen des Strukturalismus im engeren Sinne sind z.B. in Europa neben Ferdinand de Saussure auch Louis Hjelmslev (Glossematik), in den USA Leonard Bloomfield, Zellig Harris, Charles Hockett, Kenneth Pike (amerikanischer Strukturalismus, Distributionalismus). Vorgänger der Strukturalisten waren unter den komparativen Philologen die Junggrammatiker, die die Regelmäßigkeit aller diachronen Lautveränderungen betonten.

Funktionalismus: Der Funktionalismus in der Phonologie ist eine besondere Ausprägung des Strukturalismus, die die sprachlichen und situativen Kontexte fokussiert, in denen sprachliche Strukturen zu lokalisieren sind, beispielsweise im Prager Funktionalismus, der die Funktionen der Sprache in der Kommunikation und in der Kognition (Gestaltpsychologie) hervorhob und mit der Unterscheidung Sprachgebilde – Sprechakt der späteren Chomsky'schen Kompetenz – Performanz- bzw. I-Language – E-Language-Unterscheidung zuvorkam. Die Londoner Ausprägung bei Firth führte zu einer genaueren Differenzierung zwischen phonematischen und prosodischen Funktionen der Eigenschaften von Sprachlauten, die zu einem verallgemeinerten Prosodiebegriff führte. Der Funktionalismus von Halliday baut auf dem Firth'schen Funktionalismus auf, führte zur ersten theoretisch und empirisch differenzierten Intonationstheorie , ist aber im Hinblick auf den Prosodiebegriff konservativer als bei Firth. Auch der bereits erwähnte Ansatz von Pike (Tagmemik) hat starke funktionalistische Züge.

Generative (und verwandte) Phonologien: Die generative Phonologie betont im Gegensatz zum Strukturalismus eher die formalen Aspekte von phonologischen Beschreibungen als deren empirische Basis, indem von Prämissen (zugrundeliegenden lexikalischen Repräsentationen von Wörtern) und sequentiell angewendeten Ableitungsregeln (phonologischen Regeln) eine phonetische Repräsentation wie ein mathematischer Beweis hergeleitet wird. Vorläuferarbeiten von Halle, Chomsky und ihren Schülern führten zum Standardwerk *Sound Pattern of English* („SPE“) in 1968, das eine Kontroverse über die Abstraktheit phonologischer Repräsentationen auslöste. Das grundlegende Modell, das auf Phoneme und Silben verzichtete, verwendete lineare Verkettungen von Merkmalsbündeln (flachen Attribut-Wert-Strukturen) und Informationen über Wort- und Satzkonstituentengrenzen, führte aber auch eine Theorie der rekursiven Zuordnung von Betonungen und der Konstruktion komplexer Wörter ein. Die Abstraktheitskritik führte zur **Natürlichen Phonologie**, während die Kritik an der Linearität und Probleme bei der Anordnung phonologischer Regeln zur **Lexikalischen Phonologie** (Stratifikation der Rekursion in autonome Schichten), zur **Autosegmentalen Phonologie** (Abstraktion quasi-autonom strukturierter Sequenzen prosodischer Eigenschaften aus den linearen Ketten), zur **Metrischen Phonologie** (Weiterentwicklung der rekursiven Betonungstheorie) und zur **Optimalitätstheorie** (Zulassung von Constraintverletzungen) führte. In der Optimalitätstheorie, die computerlinguistisch wohl die interessanteste (wenn auch die am heftigsten

umstrittene) Theorie ist, werden geordnete Regeln als sequentiell angeordnete deklarative Constraints dargestellt, die sukzessive den Suchraum für korrekte phonetische Interpretationen eines Lexikoneintrags eingrenzen. Die Constraints lassen Merkmalveränderungen zwar nicht zu, können aber verletzt werden. Die Interpretationen mit den wenigsten Constraintverletzungen gelten als ‚optimal'. Die **Generative Phonologie** und einige ihrer hier genannten Nachfolger wurden durch Kay, Kaplan sowie Karttunen mit endlichen Automaten modelliert.

Neben diesen bekannteren Richtungen wurden einige Ansätze auf formallogischer Grundlage entwickelt, die weitere Entwicklungen zwar beeinflussten, jedoch wenig allgemeine Beachtung gefunden haben: Deklarative Phonologie (Bird, Ellison), Defaultlogische Phonologie (Gibbon), Montague-Phonologie (Bach und Wheeler), Mereologische Phonologie (Batóg).

Natürlich lässt sich die Vielfalt von Phonologien nicht vollständig in dieses einfache Schema pressen. Eigenschaften von Theorien in der einen Gruppe sind teilweise auch bei Theorien in anderen Gruppen zu finden.

Prosodie

Bezogen auf die indoeuropäischen Sprachen gehören traditionell zum Bereich der Prosodie diejenigen funktionalen lautlichen Eigenschaften, die eine längere Zeitspanne beanspruchen als ein Phonem, und die phonetisch durch die Grundfrequenz, die Intensität oder die zeitliche Organisation der Äußerung, z.B. den Rhythmus, interpretiert werden. Diese Definition ist jedoch nicht unkontrovers. Es gibt andere Eigenschaften, die länger sein können als ein Phonem, z.B. bei Assimilationen (Anpassung benachbarter Laute aneinander), in denen beispielsweise die Artikulationsart über mehr als ein Phonem beibehalten wird, etwa bei der Labialisierung von /n/ vor einem labialen Verschlußlaut: „in Bonn" /im bɔn/. In **prosodischen Phonologien** werden auch solche Eigenschaften als Prosodien in einem weiteren Sinne klassifiziert. Bereits erwähnt wurde auch die Auslautverhärtung im Deutschen als Prosodie im weiteren Sinne.

Die Sprachen der Welt bieten eine breite Palette prosodischer Eigenschaften im engeren Sinne, mit Funktionen, die teilweise ganz anders sind als in den indoeuropäischen Sprachen, z.B. **phonematische** oder **morphematische Töne** (funktionale Grundfrequenzmuster), die phonematisch als distinktive Merkmale, oder als grammatische Morpheme , oder konfigurativ als Markierungen bestimmter grammatischer Strukturen funktionieren. Die Lautdauer als phonematisches Merkmal ist auch in den indoeuropäischen Sprachen verbreitet, aber in afrikanischen Sprachen der Niger-Congo-Familie kommt die Lautdauer mitunter auch mit morphematischer Funktion, z.B. als Negativmarkierung vor.

An dieser Stelle kann nur ein kurzer Überblick über Wort-, Satz- und Diskursprosodie gegeben werden; weiterführende Lektüre wird im Literaturabschnitt 3.1.4 angegeben.

Wortprosodie

Die wortprosodischen Merkmale tragen phonematisch zur Wortunterscheidung bzw. morphematisch zur Wortbedeutung oder zur Wortstrukturmarkierung bei. Solche lautsprachlichen Eigenschaften sind vor allem von großer Bedeutung für die Sprachsynthese. Folgende Eigenschaften können unterschieden werden:

Phonematische Wortprosodie: Die phonematischen, d.h. wortunterscheidenden wortprosodischen Mittel in den Sprachen der Welt umfassen die Kategorien Ton, Tonakzent und Betonung, die möglicherweise sprachtypologisch ein Kontinuum bilden:

- **Ton:** Ein erster prosodischer Sprachtyp wird durch die phonematische Verwendung des Grundfrequenzverlaufs als Ton in Silben zur Wortunterscheidung charakterisiert. Das Mandarin-Chinesische hat z.B. 4 Töne: hoch flach, mitte-hoch steigend, mitte-tief-hoch fallend-steigend, hoch-tief fallend und zusätzlich der ‚tonlose Ton', der sich aus dem weiteren tonalen Kontext ergibt; die Interpretation der Töne im Kontext ist recht komplex. Die meisten Niger-Kongo-Sprachen in West-, Zentral-, Ost- und Südafrika haben zwei, drei oder vier sogenannte Registertöne, also Töne, die nur durch die Tonhöhe und nicht durch eine **Tonveränderung** (Kontur) charakterisiert sind; eventuell vorkommende Konturen lassen sich historisch, im Dialektvergleich und sprachintern als Kombinationen von Registertönen begründen.
- **Tonakzent:** Ein anderer Sprachtyp kennt Tonakzente , die im Gegensatz zu den Tönen in der Regel eine einzige Form haben, jedoch an unterschiedlichen Stellen im Wort vorkommen. Beispiele für solche Sprachen sind Japanisch und Schwedisch.
- **Betonung:** Ein dritter Sprachtyp, zu dem auch Deutsch, Niederländisch und Englisch gehören, verwendet Betonungen, d.h. phonetisch variable Interpretationen einer im Lexikon ausgezeichneten betonten Silbe durch Erhöhung oder Absenkung der Tonhöhe oder durch verlängerte Silbendauer. Ein deutsches Wort wie „Tenor" bedeutet „männlicher Sänger mit hoher Stimme" oder „ungefährer Inhalt", je nachdem, ob die zweite oder die erste Silbe betont wird.

Morphematische und morphosyntaktische Wortprosodie: In den meisten Niger-Kongo-Sprachen, sowie in einigen Tibeto-Burmanischen und südamerikanischen Sprachen kommen **Töne** mit grammatischer Bedeutung vor, die Flexionsmorpheme interpretieren oder die interne Grenze bei Wortkomposita markieren. Die prosodische Markierung der Wortstruktur ist im Deutschen auch zu finden, beispielsweise bei der Erstbetonung der Konstituenten von Komposita: „SCHREIBtisch", nicht „SchreibTISCH" (vgl. aber regionale Abweichungen bei Namen, beispielsweise allgemein „STEINhagen" gegenüber regional „steinHAgen").

Satz- und Diskursprosodie

In Untersuchungen zu den indoeuropäischen Sprachen ist die Satzprosodie oder **Intonation**, charakterisiert durch einen Grundfrequenzverlauf über einen Satz oder Teilsatz, wohl der klassische Bereich der Prosodie, wobei eine Unterscheidung zwischen satzorientierter Funktion und Funktion im Diskurs schwer aufrechtzuerhalten ist. Die wichtigsten satzprosodischen Funktionen, die der Intonation zugeschrieben werden, sind die **Phrasierung** (die Einteilung einer Äußerung in intonatorische Phrasierungseinheiten, **Intonationsphrasen**), die **Akzentplatzierung** (Zuordnung eines Satzakzents zu einer Satzkonstituente) und die Zuweisung eines **Terminaltons**:

Phrasierung: Sprachliche Äußerungen werden durch relativ klar erkennbare Grundfrequenzkonturen in Intonationsphrasen eingeteilt, die je nach Sprechstil in der Regel, aber nicht notwendigerweise, größeren grammatischen Einheiten wie Nominalphrasen, Satzteilen oder Sätzen zugeordnet werden. Beispiele sind in den Grundfrequenzverläufen in den Abbildungen 3.3 und 3.5 zu finden, die Gesamtkonturen mit lokalen Modulationen zeigen.

Akzentplatzierung: Innerhalb einer Intonationsphrase werden die Wortakzente (phonetische Interpretationen der abstrakten, lexikalisch festgelegten Wortbetonungen) in formellen Sprechstilen rhythmisch angeordnet, in informellen Stilen weniger rhythmisch sondern abhängig von spontanen Formulierungsprozessen und pragmatischen Constraints. Den Wortakzenten überlagert sind die Satzakzente, in der Regel nur eine pro Intonationseinheit, die Fokus-, Kontrast- und Emphasefunktionen haben können. Abbildung 3.3 zeigt eine akzentuierende Erhöhung des Tonhöhenverlaufs auf den Silben „Lö“ und „Maus“ in den Wörtern „Löwe“ und „Maus“.

Terminalton: Der steigende, fallende, komplex steigend-fallende oder fallend-steigende (seltener noch komplexere) Terminalton ist wohl das auffälligste Element der Intonation bzw. der Satzprosodie. Dem Terminalton werden in der Literatur, vor allem in der sprachdidaktischen Literatur, recht spezifische grammatische (Frage, Aufforderung, Ausruf, usw.) oder pragmatische (emotionale, wertende, usw.) Bedeutungen zugeschrieben. Solche Bedeutungen werden jedoch sehr oft assoziativ aus dem Wortlaut oder dem Situationskontext heraus interpretiert und sollten nicht allein der Intonation zugeschrieben werden.

Die **Terminalkonturen** selbst haben in der Regel lediglich die Funktion, die Abgeschlossenheit oder Nichtabgeschlossenheit einer grammatischen Einheit (z.B. zwischen Subjekt und Verb, bei einer Liste) oder eines Diskursstückes (z.B. Frage-Antwortsequenzen) anzuzeigen. Diese Funktion ist in Abbildung 3.3 am Ende der Intonationskurve zu sehen: Es handelt sich um den Titel einer Geschichte, der mit einer leichten Tonhöhensteigung endet, die die Fortsetzung durch den Hauptteil der Geschichte ankündigt.

Eine Terminalkontur sowie eine globale Tonhöhenkontur können auch soziale und emotionale Funktionen haben.

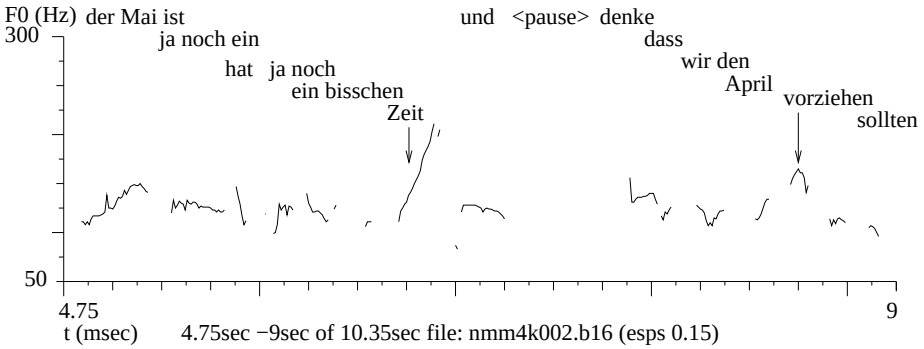

Abbildung 3.5: Grundfrequenzverlauf einer konversationellen Äußerung.

Die Abbildung des Grundfrequenzverlaufs in Abbildung 3.5 ist mit der orthographischen Transkription der Äußerung lose beschriftet, um den Frequenz-Text-Bezug anzudeuten. Die Phrasierung wird durch die Konjunktion „und“ mit nachfolgender Pause, die Akzentplatzierungen auf „Mai“ und „April“ sind durch lokale Frequenzsteigungen und durch Terminaltöne auf „Zeit“ (steigend) und „VOR“ (in „vorziehen“, fallend) durch deutliche Frequenzveränderungen markiert.

Integration von Prosodie, Phonologie und Phonetik

Wenn die Funktionen der prosodischen Objekte im Detail betrachtet werden, fällt auf, dass sie im Großen und Ganzen den Grundobjekten Phonem, Morphem, Wort und Satz zugeordnet werden können. Es liegt also nahe, auch den prosodischen Objekten eine semiotische Charakterisierung unter Bezugnahme auf die Grundobjekte zukommen zu lassen, wie bereits bei den Phonemen. Die phonetische Interpretation eines Phonems ist entweder ein Allophon oder ein Ton. Die phonetische Interpretation eines Morphems ist eine Funktion der phonetischen Interpretation der Phoneme, die ihm zugeordnet sind, und einer diesen Phonemen zugewiesenen prosodischen Einheit, z.B. ein Ton oder ein Akzent.

Die kompositionelle Hierarchie kann fortgesetzt werden: Die phonetische Interpretation eines Wortes ist eine Funktion der phonetischen Interpretation seiner Bestandteile und der prosodischen Strukturmarkierung, die phonetische Interpretation eines Satzes ist eine Funktion der phonetischen Interpretation seiner Bestandteile und der prosodischen Markierungen der Phrasierung, der Akzentsetzung und des Terminaltons. Die integrierte Hierarchie, die eine generalisierte phonetische Interpretation darstellt, wird in Abbildung 3.6 visualisiert. Diese Sichtweise ermöglicht es dem Computerlinguisten, die umfangreiche, aber recht fragmentierte Literatur zum Thema Prosodie zu systematisieren und im Rahmen bekannter kompositorischer Prinzipien zu formalisieren und implementieren.

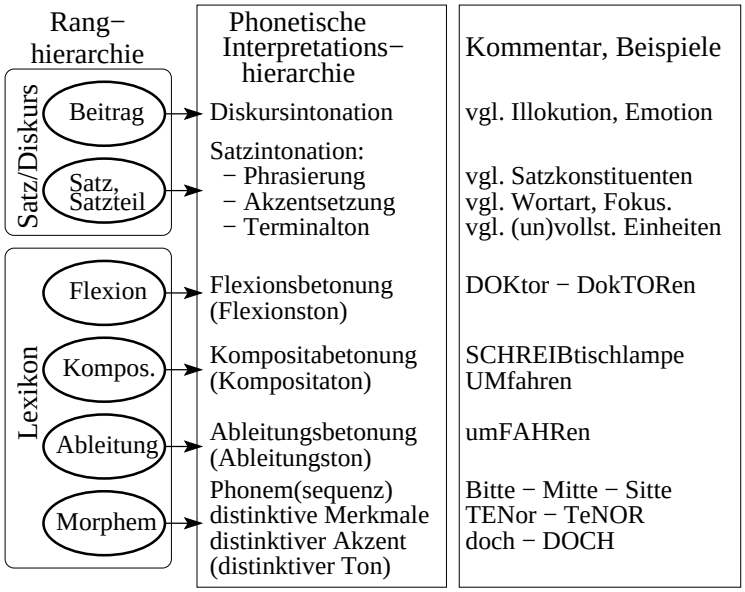

Abbildung 3.6: Generalisierte phonetische Interpretation zur Integration von phonologischen und prosodischen Einheiten.

3.1.2 Empirische Methoden

Die empirischen Grundlagen für die Phonetik und Phonologie sind im Prinzip gleich und ergeben einen dreidimensionalen empirischen Methodenraum:

Korpus: Ein Korpus ist eine Sammlung beobachteter, aufgenommener und auf Speichermedien verfügbarer sprachlicher Äußerungen, die entweder einzeln durch Selbst- oder Fremdbefragung direkt elizitiert, oder für Experimente geplant, oder als systematische oder authentische (nicht für phonetische Untersuchungszwecke erstellte) Datensammlungen aufgebaut werden. Ein Korpus enthält aber auch in der Regel eine mit Metadaten systematisch dokumentierte Menge von solchen akustischen (zunehmend auch multimedialen) Aufnahmen mit den dazugehörigen Transkriptionen, Annotationen, und eventuell auch ein Korpuslexikon.

Analyse: z.T. durch unterschiedliche Werkzeuge unterstützte Kategorisierungen von Äußerungen in einem Korpus im Hinblick auf ihren sprachlichen Status, ihre Bestandteile und die wahrnehmbaren Eigenschaften dieser Bestandteile durch den phonetisch ausgebildeten Experten. Die Kategorisierungen werden in der Regel unter Zuhilfenahme von standardisierten phonetischen Alphabeten und Merkmalssystemen und Annahmen über Silben- und Wortstruktur durchgeführt.

Werkzeuge: intellektuelle Werkzeuge (z.B. phonetische Alphabete und Merkmalsysteme, Parameterbeschreibungen, Ontologien usw.) und operationale Werkzeuge für instrumentelle Messungen, sowie deren Visualisierung und statistische Auswertung, sowie für die symbolorientierte Analyse und Modellierung von lautsprachlichen Äußerungen. Die am meisten verwendeten operationalen Werkzeuge sind Programme zur Anzeige der akustischen Eigenschaften von Sprachsignalen und zur Annotation (Zuordnung von Transkriptionen zu Sprachsignalen).

Die beiden Disziplinen Phonetik und Phonologie unterscheiden sich in ihrer Gewichtung der verschiedenen Spielarten der beiden empirischen Grundlagen. Es gibt aber nicht nur Überlappungen zwischen den Disziplinen: Die Disziplinen positionieren sich tendenziell an ganz anderen Stellen im empirischen Methodenraum. Es gibt aber keine scharfe Trennlinie zwischen phonetischen und phonologischen Methoden, wie die Bezeichnungen von Ansätzen wie „Phonology as Functional Phonetics“ oder „Laboratory Phonology“ andeuten.

Die methodologischen Überlappungen und die Schnittstellen (im Sinne von gemeinsamen Repräsentationen von Fakten und Regeln) zwischen Phonetik und Phonologie werden seit mehr als einem Jahrhundert kontrovers diskutiert. Je nach empiristischer, kognitivistischer oder anwendungsorientierter Einstellung werden die Dimensionen Korpus, linguistische Kategorisierung oder Werkzeuge in den Vordergrund gestellt. Am sinnvollsten scheint es zu sein, den gemeinsamen empirischen Methodenraum einzusetzen und einzelne Ansätze oder Studien entlang der drei Dimensionen des Methodenraums zu charakterisieren.

Die folgenden Teilabschnitte geben einen kurzen Überblick über empirische Methoden, Techniken zur Transkription und Annotation, experimentelle und korpusphonetische Methoden, sowie Anwendungsbereiche der Phonetik.

Methodenüberblick

Auf die drei Teildomänen der Phonetik lassen sich unterschiedliche Methoden anwenden. Die **Ohrenphonetik**, die auf dem geschulten Hörsinn des ausgebildeten Phonetikers aufbaut, wird von der **Instrumentalphonetik**, bei der Messinstrumente und -software verwendet werden, unterschieden. Eine ohrenphonetische Analyse zur Bestimmung des genauen Gegenstandsbereichs ist stets Voraussetzung für eine sinnvolle instrumentalphonetische Analyse. Verwirrend ist die oft anzutreffende Verwendung von „auditiv“ nicht nur für die auditive Teildomäne der Phonetik sondern auch für die ohrenphonetische Methode, die dann „auditive Methode“ heißt. Manchmal wird „impressionistische Phonetik“ statt „Ohrenphonetik“ benutzt. Die Bezeichnung „Wahrnehmungsphonetik“ wird manchmal auch in beiden Bedeutungen verwendet: für die Untersuchung der Wahrnehmung und für die Untersuchung durch Wahrnehmung, z.B. in Wahrnehmungsexperimenten. Gegenstandsbereich und Methode sollten aber auf jeden Fall konsistent auseinandergehalten werden.

Orthogonal zur Unterscheidung zwischen Ohren- und Instrumentalphonetik ist die weitere Unterscheidung zwischen **qualitativen Methoden**, bei denen

einzelne Sprachsignale beobachtet und transkribiert oder gemessen, dann analysiert und illustriert werden, von **quantitativen Methoden**, bei denen größere Datenmengen aus Experimenten und Korpora statistisch untersucht werden.

Bei den qualitativen Methoden wird weiter unterschieden zwischen teilnehmender Beobachtung (der Beobachter interagiert authentisch, d.h. nicht als Forscher zu erkennen) und nicht-teilnehmender Beobachtung (der Forscher wird klar von den untersuchten Personen unterschieden). Teilnehmen und Beobachten sind nicht unbedingt miteinander kompatibel. Bei der teilnehmenden Beobachtung muss also abgewogen werden, ob eher der teilnehmende oder der beobachtende Aspekt bevorzugt wird. Auf jeden Fall müssen ethische und juristische Gesichtspunkte bei der teilnehmenden Beobachtung berücksichtigt werden.

Qualitative Untersuchungen sind auch stets Voraussetzung für sinnvolle quantitative Untersuchungen. Insofern geht die phonetische Analysearbeit einen Weg von ohrenphonetisch-qualitativen zu ohrenphonetisch-quantitativen Untersuchungen, oder von ohrenphonetisch-qualitativen über instrumentalphonetisch-qualitativen (direkte Inspektion von Messungen) zu instrumentalphonetisch-quantitativen, statistisch auswertenden Methoden.

Bei den quantitativen Methoden unterscheidet man ferner zwischen **experimentellen Methoden**, bei denen sorgfältig strukturierte Datentypen in Rezeptions- und Produktionsexperimenten untersucht werden, und **korpusphonetischen Methoden**, bei denen große Mengen an weniger homogenen Sprachaufnahmen eines Korpus aus einem allgemeiner spezifizierten Szenario untersucht werden.

Die Verwendbarkeit der Methoden hängt von der Teildomäne ab. Mit qualitativen Methoden lassen sich die artikulatorische Domäne (durch Selbstwahrnehmung der Sprechorgane) und die akustische Domäne (durch Höreindrücke vom Schall) untersuchen, aber nicht die auditive Domäne. Vorgänge im Ohr kann man nicht direkt beobachten. **Instrumentalphonetische Methoden** lassen sich auf alle drei Teildomänen anwenden, allerdings erfordert die messphonetische Untersuchung der Produktion (teilweise) und der Rezeption (vollständig) die Zusammenarbeit mit Fachmedizinern bzw. mit Neurologen. Nur die akustische Domäne ist für medizinische Laien problemlos messtechnisch zugänglich, wenngleich erhebliche technische Kenntnisse der Akustik des Sprachsignals und der Signalverarbeitung für die erfolgreiche Arbeit notwendig sind. Diese Domäne ist für die bekanntesten Anwendungen der Phonetik in der Sprachtechnologie – automatische Sprachsynthese, Sprecher- und Spracherkennung – relevant. Technologische Anwendungen in der Produktionsdomäne (z.B. Sprechprothesen) oder in der Rezeptionsdomäne (z.B. Hörgeräte) erfordern medizintechnische Zusammenarbeit.

Ressourcen: Aufnahme, Transkription, Annotation

Die Qualität phonetischer Untersuchungen und damit auch indirekt die Qualität auch von phonologischen Untersuchungen hängt von der Qualität der empirischen Ressourcen ab, die durch den empirischen Methodenraum bereits definiert wurden: Korpus, Analyse, Werkzeuge. Der Qualitätssicherung phonetischer

Ressourcen ist viel Aufmerksamkeit gewidmet worden, vor allem im Kontext der Anwendung phonetischer Analysen in der Sprachtechnologie bei der Entwicklung von Sprachsynthese- und Spracherkennungssystemen, aber auch in hochqualitativer Dokumentation der vom Aussterben bedrohten Sprachen der Welt (siehe den Literaturabschnitt 3.1.4).

Die Direkttranskription einzelner Zufallsbeobachtungen ohne akustische Aufnahmen wird z.B. noch in der Analyse von Versprechern (die nicht leicht zu elizitieren sind) und in der Fehleranalyse im Fremdsprachenunterricht verwendet. In der deskriptiv-linguistischen Feldforschung werden auch z.T. noch Direkttranskriptionen ohne akustische Aufnahmen angefertigt; diese Methode verschwindet aber allmählich mit dem zunehmenden Bewusstsein der Bedeutung wiederverwendbarer hochqualitativer phonetischer Ressourcen.

In phonologischen Untersuchungen wurden traditionell keine oder kaum Beobachtungen im üblichen empirischen Sinne gemacht. Vor allem Muttersprachendaten wurden (und werden noch) manchmal nur introspektiv vom Phonologen erdacht. Die introspektive Methode wird vor allem von Soziolinguisten kritisiert: Es ist bekannt, dass introspektive Urteile stark durch normativ-subjektive Kategorisierungsschemata beeinflusst werden, die mit der Äußerungswirklichkeit nicht gut übereinstimmen. Diese Variante der qualitativen Methode wird immer weniger verwendet, sondern durch empirisch abgesicherte qualitative und quantitative Methoden ergänzt.

Aufnahme

Zur **Aufnahmeplanung** gehören drei Phasen, die über den eigentlichen Aufnahmevorgang hinausgehen und sorgfältig durchgeführt werden müssen, um den heutigen Ansprüchen an Wiederverwertbarkeit (*reusability*) und Nachhaltigkeit (*sustainability*) zu genügen: die Designphase (*pre-recording phase*), die Aufnahmephase (*recording phase*) und die Bearbeitungsphase (*post-recording phase*).

Designphase: In der Designphase geht es darum, den Rahmen für die Datenaufnahme zu spezifizieren: die Fragestellung, den Untersuchungstyp (z. B. Produktions-, Wahrnehmungs- oder Reaktionsexperiment, oder der Korpusdatentyp für dialogische Interaktionen), das Szenario und (bei experimentellen Fragestellungen) die Instruktionen und Vorlagen sowie Datenverwendungsvereinbarungen mit den Versuchspersonen, Aufnahmeausrüstung, sowie Aufnahmeort und -zeit. Diese Informationen gehen in die Metadaten zum Korpus ein. Probeaufnahmen werden durchgeführt, um den Aufnahmeablauf zu testen.

Aufnahmephase: In die Aufnahmephase Sprachaufnahme fällt der tatsächliche Ablauf der Datenerhebung, im Studio oder in einer natürlichen Umgebung, je nach Designspezifikation. Für die Aufnahme müssen alle Materialien bereitgestellt werden (vorbereitete Unterlagen, Geräte, Stromversorgung), Trinkwasser für die Sprecher (z.B. ein Schluck alle 5 oder 10 Minuten, um eine Austrocknung der Stimmbänder zu vermeiden, die die Aufnahmequalität beeinträchtigen würde). Während der Aufnahme muss für korrekte

Mikrofonplatzierung und Signalaussteuerung gesorgt werden. Gleichzeitig werden standardisierte Metadaten über Aufnahmematerialien und -verlauf festgehalten.

Bearbeitungsphase: Der erste Schritt in der Bearbeitungsphase ist die Archivierung der aufgenommenen Daten und der Metadaten mit systematischen und eindeutigen Dateinamen. Die folgenden Schritte der Transkription, Annotation und Korpuslexikon- oder Sprachmodellerstellung werden gesondert behandelt.

Transkription

Eine Transkription ist die Zuordnung einer symbolischen Repräsentation zu einer sprachlichen Äußerung, heutzutage normalerweise zu einer Audio- oder Videoaufnahme einer sprachlichen Äußerung. Die Möglichkeiten der symbolischen Repräsentation sind vielfältig und hängen von der Fragestellung ab. Auf jeden Fall müssen die **Transkriptionskonventionen** exakt spezifiziert werden, nicht ad hoc erfunden; hierzu ist häufig eine Testphase erforderlich, wenn noch wenige Erfahrungen mit dem Datentyp vorliegen.

In diesem Beitrag werden Transkriptionskonventionen für Videoaufnahmen von Äußerungsvorgängen und sprachlichen Interaktionen nicht behandelt. Diese sind noch relativ wenig standardisiert und werden immer weiter entwickelt (siehe aber den Literaturabschnitt 3.1.4).

Die wichtigsten Transkriptionstypen für Audiodaten verwenden die IPA-Transkriptionskonventionen (Abbildung 3.15), für die maschinelle Verarbeitung auch die SAMPA- und Unicode-Kodierungen des IPA. Die wichtigsten Transkriptionstypen werden hier beschrieben.

Orthographische Transkription: Die orthographische Transkription folgt den Standardregeln der Orthographie und bedarf in dieser Hinsicht keines weiteren Kommentars. Zitate im Textzusammenhang werden mit den üblichen Zitierzeichen gekennzeichnet.

Modifizierte orthographische Transkription: In der Konversationsanalyse oder in der Transkription der Kindersprache wird oft eine modifizierte orthographische Transkription verwendet, die Vokalisierungen, Geräusche und nicht-standardisierte Aussprachevarianten andeuten soll. Diese Transkriptionsart ist als Grundlage für funktional orientierte sprachliche Analysen entwickelt worden und für phonetische und sprachtechnologische Zwecke nicht gut geeignet. Zitate im Textzusammenhang werden mit den üblichen Zitierzeichen gekennzeichnet.

Phonotypische Transkription: Die phonotypische oder morphophonematische Transkription setzt eine morphologische Analyse der Sprache voraus und generalisiert über morphologisch bedinge Varianten, beispielsweise „Hund“ mit den Stämmen /hunt/ und /hund/ (vgl. „der Hund“ ausgesprochen

„Hunt“, „des Hundes“). Morphophoneme werden manchmal großgeschrieben und können mit Schrägstrichen zitiert werden, z.B. /hʊnD/, oder, zur Unterscheidung von phonematischen Transkriptionen, zwischen spitzen oder geschweiften Klammern.

Kanonische phonematische Transkription: Die kanonische phonematische Transkription ist die wortunterscheidende Transkription, die in einem Aussprachewörterbuch verwendet wird und die phonematischen Kriterien für die Abstraktion von phonetischen Details aufgrund der phonetischen Ähnlichkeit und der komplementären Distribution der Allophone erfüllt. Neben der orthographischen Transkription ist die kanonische phonematische Transkription die nützlichste Transkriptionsart in der Computerlinguistik und Sprachtechnologie. In der Sprachtechnologie wird eine kanonische honematische Transkription in der Regel automatisch anhand eines Aussprachelexikons und Graphem-Phonem-Übersetzungsregeln erzeugt, eine Prozedur, die Graphem-Phonem-Übersetzung oder Phonetisierung genannt wird. Kanonische phonematische Zitate im Textzusammenhang werden mit Schrägstrichen gekennzeichnet.

Weite phonetische Transkription: Die weite phonetische Transkription ist eine phonematische Transkription, die nicht unbedingt dem Kriterium der kanonischen lexikalischen Repräsentation entspricht. Diese Art der Transkription wurde primär für den Fremdsprachenunterricht entwickelt, um bestimmte Arten der Assimilation zu verdeutlichen, wie etwa „in Bonn“ /in bɔn/ in der Aussprache „im Bonn“ /im bɔn/. Diese Transkriptionsart ergänzt die kanonische phonematische Transkription, kann sie aber nicht ersetzen. Zitate im Textzusammenhang werden mit Schrägstrichen gekennzeichnet.

Enge phonetische Transkription: Die enge phonetische Transkription zeigt mehr phonetische Details der Aussprache an, als für die einfache Unterscheidung von Wörtern notwendig wäre, beispielsweise die Behauchung von stimmlosen Plosiven am Wortanfang („Tanne“ [tʰanə]). Die enge phonetische Transkription ist in der allgemeinen Linguistik, der Soziolinguistik und der Dialektologie sowie in phonetischen Detailuntersuchungen unerlässlich. In der Sprachtechnologie wird die enge phonetische Transkription in der Regel nicht verwendet. Das Sprachsignal wird direkt auf eine kanonische phonematische Transkription abgebildet. Die enge phonetische Transkription erfordert eine intensive Phonetikausbildung, ist sehr zeitaufwändig und bei einer Abbildung von Phonemen in Kontext (z.B. als Diphone oder Triphone) außer bei sehr auffälligen Allophonen nicht notwendig.

Erweiterte Transkriptionen: Die bisher aufgeführten Transkriptionsarten können um vielerlei weitere Informationen erweitert werden, wie z. B. prosodische Informationen, Informationen über Pausen, Häsitationssignale, Abbrüche und Neustarts, nonverbale Vokalisierungen (Lachen, Seufzen, Weinen, usw.), Lautungen bei Sprachbehinderungen, sowie bei Husten,

Niesen und anderen nichtsprachlichen Geräuschen. Für prosodische Transkriptionssysteme gibt es mehrere Vorschläge, von denen die wichtigsten hier nur genannt werden sollen: die IPA-Symbole (Abbildung 3.15), ToBI (Tones and Break Indices), INTSINT (International Transcription System for Intonation), SAMPROSA (aus demselben europäischen Projekt wie SAMPA , eine Zusammenstellung gebräuchlicher prosodischer Transkriptionssymbole). Im Gegensatz zu den orthographischen oder phonematischen und phonetischen Transkriptionstypen und auch zu den prosodischen Transkriptionskonventionen sind die Symbole für andere Vokalisierungen und Lautungen nicht standardisiert, daher werden hier keine weiteren Hinweise dazu gegeben. Im Internet können allerdings zahlreiche Hinweise auf die Konventionen für solche Lautungen gefunden werden.

Annotation

Zur Analyse der Schallwellen haben Phonetiker und Sprachingenieure vielfältige Software für alle gängigen Betriebssysteme zur Verfügung gestellt, die im Internet erhältlich sind und in einigen Linux-Distributionen zur automatischen Installation bereitgestellt werden. Die bekanntesten freien Software-Werkzeuge für die akustische Sprachsignalanalyse sind *Praat*, *WaveSurfer* sowie *Transcriber*. Auch andere, eher für das Editieren von Musik und Audio-Reportagen vorgesehene freie Software wie *Audacity* eignet sich für das Schneiden, Filtern usw. von Sprachaufnahmen.

Der wichtigste Schritt in der computerphonologischen , computerphonetischen und sprachtechnologischen Korpusanalyse ist die Korpusannotation, auch „labelling“, „time alignment“ oder „Etikettierung“ genannt. Die Annotation in diesem Sinne ist als eineindeutige Relation zwischen Symbolen einer Transkription und Zeitstempeln von Segmenten in einem Sprachsignal definiert.

Ein Beispiel für eine Annotation auf Phonemebene wird in Abbildung 3.3 wiedergegeben: Einzelne Phonemsymbole in der Transkription von „der Löwe und die Maus“ werden mit Zeitstempeln versehen, die eine genaue Zuordnung zum Sprachsignal ermöglichen. Die Transkriptionseinheiten, die annotiert werden, hängen von der Fragestellung ab und können Phoneme, Silben, Wörter usw. sein. Üblich sind auch orthographische Annotationen. Die Hilfe-Dokumentation der Praat-Software bietet eine ausgezeichnete Einführung in die Annotation an. Vielfältige Informationen, auch zur prosodischen Annotation, sind im Internet erhältlich, beispielsweise zum ToBI-System oder zum IntSint-System.

Korpuslexikon, Syntheseeinheiten, Sprachmodelle

Die Transkriptionen und Annotationen werden auf verschiedene Weisen weiterverarbeitet, in der Praxis oft mit Skriptsprachen wie z.B. Perl, Python oder einer UNIX-Shell-Sprache.

Ein weiterer wesentlicher Schritt vor allem bei der sprachtechnologischen Korpusbearbeitung ist die Erstellung eines Korpuslexikons, mit Häufigkeitsstati-

stiken über die für Untersuchungen relevanten Einheiten und Kombinationen von Einheiten. Aus solchen Korpuslexika werden statistische oder symbolische **Sprachmodelle** für die Suchraumeingrenzung für Wörter in der Spracherkennung erstellt.

3.1.3 Formale Methoden

Die eher signalverarbeitenden formalen Methoden in der Phonetik und die eher symbolverarbeitenden formalen Methoden in der Computerphonologie sind verschieden und werden daher getrennt behandelt.

Formale Methoden in der Phonetik

Die in der Phonetik zur Anwendung kommenden formalen Methoden betreffen einerseits quantitative Parameter des Sprachsignals, andererseits statistische Verfahren zur Analyse des Sprachsignals. Eine ausführliche Einführung in diesen Bereich kann an dieser Stelle nicht geleistet werden (s. aber den Literaturabschnitt 3.1.4).

Hier sollen lediglich die wichtigsten Begriffe in der akustischen Phonetik erläutert werden, weil diese Teildomäne wohl für die meisten Bereiche der heutigen Phonetik und Sprachtechnologie die wichtigste ist.

Die Zeitdomäne

Das Sprachsignal soll zunächst in der Zeitdomäne beschrieben werden, in der die Amplitude als Funktion der Zeit dargestellt wird. Hierzu dient Abbildung 3.7, die einen Auszug aus dem bereits in Abbildung 3.3 dargestellten Signal visualisiert. Abbildung 3.7 zeigt den Übergang vom Vokal [ʊ] (dem zweiten Teil des Diphthongs [aʊ]) auf den Frikativ [s] im Wort „Maus“ [maʊs]. Sehr deutlich zu sehen ist der Unterschied zwischen der regelmäßigen Schwingung des Klangs [ʊ] und dem unregelmässigen Verlauf des Geräuschs [s].

Für die Beschreibung des Sprachsignals in der Zeitdomäne sind die Parameter **Amplitude**, **Intensität**, **Energie** (oder **RMS-Amplitude**), **Periode**, **Frequenz** und **Phase** die wichtigsten Grundbegriffe, sowie für digitalisierte Sprachsignale das **Digitalisieren**, die **Abtastfrequenz** und das **Aliasing**.

Größe: Die Größe, die in der Zeitdomäne als Funktion der Zeit gemessen wird, ist der variable Druck der Schallwellen im Medium Luft (oder einem anderen Medium), der die Bewegungen des Trommelfells oder der Mikrofonmembran verursacht. Der Druck wird im Innenohr in Nervensignale umgesetzt, im Mikrofon in elektrische Potentiale.

Amplitude: Die Amplitude des Drucks ist die Abweichung der Druckstärke vom Ruhewert (Nullwert) und hat bei einer Schallwelle positive und negative Werte um den Nullwert. Die durchschnittliche Amplitude hat somit bei einem Signal, das um den Nullwert symmetrisch ist und vollständige Perioden enthält, den Wert 0. Bei einem nicht symmetrischen Signal heißt die

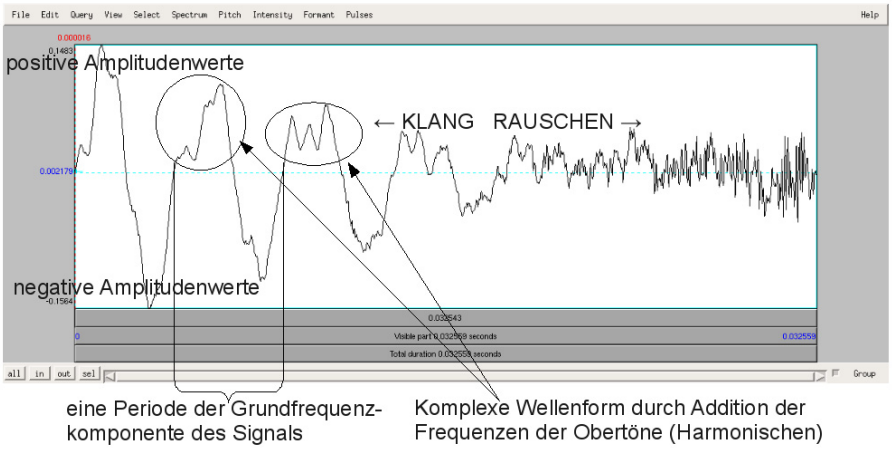

Abbildung 3.7: Übergang von [ʊ] zu [s] in [maus].

Abweichung **additive Konstante** oder ***y*-Abschnitt** (engl. *offset* oder auch *DC offset*). Das Signal in der Zeitdomäne wird mit der Amplitude als Funktion der Zeit dargestellt:

$$A(signal_x) = f(t_x)$$

Oszillogramm: Das Oszillogramm ist eine Visualisierung des Amplitudenverlaufs (der Wellenform) in der Zeit. In Abbildung 3.3 wird im oberen Bereich ein Oszillogramm gezeigt. Abbildungen 3.7 und 3.8 (rechts, oben) zeigen ebenfalls Oszillogramme.

Intensität: Die Intensität des Signals ist die Amplitude im Quadrat:

$$I = A^2$$

Intervall: Ein zeitliches Intervall ist eine Zeitspanne (Zeitdifferenz, Zeitunterschied), dargestellt als $t_i - t_{i-1}$, δ_t, d_t, usw.

RMS-Amplitude (RMS-Intensität, Energie): Die RMS-Amplitude entspricht der durchschnittlichen Intensität in einem Intervall $t_1, ...t_n$:

$$E = \sqrt{\frac{\sum_{i=1}^{n} A(x_i)^2}{n}}$$

Periode, Frequenz, Phase: Die **Periode** eines Signals ist das Intervall einer vollständigen Welle. Die **Frequenz** in Hertz ist die Anzahl der Perioden in einer Sekunde (δ_t ist die **Periodendauer**):

$$f = \frac{1}{\delta_t}$$

Ein Zyklus eines periodischen Signals fängt bei Phase 0 an und durchläuft 360°, bevor der Zyklus neu anfängt. Die Phasen verschiedener Obertöne des Signals müssen nicht unbedingt miteinander übereinstimmen. Wenn die Phasen zweier sonst gleicher Signale sich um 180° unterscheiden, heben sie sich auf. Nach diesem Prinzip funktionieren geräuschneutralisierende Kopfhörer. Die Phase eines Sprachsignals ist im Allgemeinen nicht von großer Bedeutung in der phonetischen Analyse.

Digitalisieren, Abtastfrequenz, Nykvist-Theorem: Das Digitalisieren ist die Messung der Amplitude des Sprachsignals in (gewöhnlich) regelmäßigen Abständen. Die **Abtastfrequenz** (engl. *sampling rate*) ist die Anzahl der Messungen des Signals pro Sekunde in Hertz. Das **Nykvist-Theorem** besagt:

> Wenn f die höchste zu messende Frequenz ist, dann muss die Abtastfrequenz mindestens $2f$ sein.

Andernfalls wird die Frequenz nicht korrekt gemessen, weil bei kleineren Abtastfrequenzen **Phantomfrequenzen** erscheinen, die die Messung verfälschen. Die Frequenz $2f$ heißt auch die **Nykvist-Frequenz**.

Die Abtastfrequenz für Audio-CDs beträgt beispielsweise 44100 Hz, aus 2 Gründen:

1. Wenn für die höchste von Menschen wahrnehmbare Frequenz gilt: $f =$ 22 kHz, dann: $2f = 44$ kHz.
2. Die Summe der Quadrate der ersten vier Primzahlen ergibt eine geringfügig höhere Zahl als 44 kHz und wurde gewählt, um eine möglichst vielseitige digitale Frequenzteilung ohne aufwändiges Rechnen zu ermöglichen:

$$2^2 + 3^2 + 5^2 + 7^2 = 44100$$

Mit heutigen digitalen Signalverarbeitungstechniken (**DSP-Techniken**) wäre die Berechnungsökonomie, die die Zahl 44100 ermöglicht, eigentlich nicht mehr notwendig. Die Standardabtastfrequenzen für digitale Audiobänder sind 48 kHz und 32 kHz (letztere häufig mit einer praktisch nicht wahrnehmbaren aber dennoch verlustbehafteten Signalkompression verbunden und daher nicht unbedingt für phonetische Analysen geeignet). Übliche Abtastraten für die phonetische Signalanalyse sind 16 kHz (wird seltener verwendet) und 22,05 kHz (die Hälfte von 44,1 kHz). Als Standardformat dafür hat sich das WAV-Format der Fa. Microsoft durchgesetzt. Das MP3-Format des Frauenhofer Instituts ist für viele Arten der phonetischen Analyse ungeeignet, weil das Frequenzspektrum entsprechend einem optimierenden Hörmodell verzerrt wird und also verlustbehaftet ist; Grundfrequenz und Formantfrequenzen werden aber erhalten. In spezifischen sprachtechnologischen Anwendungen können das MP3-Format und andere komprimierte Formate jedoch vorkommen.

Zeitfenster: Ein Zeitfenster ist eine Funktion, die das Signal in einem bestimmten Intervall transformiert. Die Identitätstransformation ist einfach eine Kopie des Signals in diesem Intervall. Die Funktion einer Fenstertransformation ist häufig, z.B. mit einer Kosinus- oder Gaussfunktion, eine allmähliche Absenkung der Amplitude am Anfang und am Ende des Intervalls zu bewirken, um das Vortäuschen hoher Frequenzen durch ein plötzliches Abschneiden des Signals zu vermeiden. Ein Zeitfenster ist also nicht einfach ein Intervall.

Die Frequenzdomäne

Die Darstellung des Sprachsignals in der Zeitdomäne ist die grundlegende Darstellung. Die zweite wichtige Darstellung ist die **Frequenzdomäne**, die aus der Zeitdomäne unter Verwendung einer Transformation (z.B. Fourier-Transformation) berechnet wird:

Spektralanalyse: Sprachsignale sind komplexe Signale, die durch die Spektralanalyse in ihre Teilkomponenten zerlegt werden. Wenn die Frequenzen im komplexen Signal aus einer Grundfrequenz und deren ganzzahligen Vielfachen bestehen, dann handelt es sich um einen **Klang** (vgl. Vokale). Wenn die Frequenzen im komplexen Signal in keinem einfachen Verhältnis zueinander stehen, handelt es sich um ein **Geräusch** (vgl. Reibelaute). Das Signal kann durch eine Spektralanalyse mittels der **Fourier-Transformation** in ihre einzelnen Frequenzanteile zerlegt werden. Die **Energie** der Komponenten mit den so ermittelten Frequenzen bildet das **Spektrum** und wird als Funktion der Frequenz dargestellt:

$$\mathit{Intensität}(x) = f(\mathit{Frequenz}(x))$$

Ein **Spektrogramm** ist eine Aneinanderreihung von Spektren, die eine dreidimensionale Darstellung des Verlaufs der Signalkomponenten in der Zeit ermöglicht.

Fourier-Transformation: Die am häufigsten verwendete Methode der Spektralanalyse ist die Fourier-Transformation, die von der Annahme ausgeht, dass jedes komplexe Signal als die punktweise Addition von reinen Sinusschwingungen unterschiedlicher Frequenz, Phase und Amplitude zusammengesetzt ist (**Fouriersynthese**) bzw. in solche Sinusschwingungen zerlegt werden kann (**Fourieranalyse**). Die Fourier-Transformation kann intuitiv als ein Korrelationsverfahren verstanden werden: Sinusförmige Vergleichssignale mit systematisch variierender Frequenz, Phase und Amplitude werden mit dem zu analysierenden Signal korreliert. Der Korrelationswert zeigt dann den Grad der Übereinstimming der Frequenz-, Phasen- und Amplitudeneigenschaften des Vergleichssignals mit Komponenten des zu analysierenden komplexen Signals. Die Frequenzen der Teilsignale und deren Intensität werden als Spektrum dargestellt. In der Phonetik und der Sprach-

technologie wird die Phaseninformation meist nicht benötigt. Zur Berechnung der Fourier-Transformation bei digitalen Signalen wird die **Diskrete Fourier-Transformation** (**DFT**) über die Abtastwerte verwendet, meist in einer effizienten Variante, der **Fast Fourier Transformation** (**FFT**), bei der die Punktzahl eine Zweierpotenz sein muss.

Grundfrequenz: Die Grundfrequenz ist die tiefste Frequenz in einem Klang. Die anderen Frequenzen, die in einem Klang ganzzahlige Vielfache der Grundfrequenz sind, sind die Obertöne. Die Grundfrequenz entspricht in etwa dem Höreindruck der Tonhöhe, die die Sprachmelodie (Ton, Akzent, Intonation) bestimmt, und der Phonationsrate der Glottis in der Sprachproduktion. Zur Bestimmung der Grundfrequenz können viele Methoden angewendet werden. In der Zeitdomäne können z.B. die Perioden zwischen Nulldurchgängen, zwischen Signalgipfeln, oder auch zwischen Korrelationsmaxima bei Vergleich eines Teils des Signals mit überlappenden nachfolgenden Teilen des Signals verglichen werden (**Autokorrelation**). In der Frequenzdomäne können z.B. die Abstände zwischen den Obertönen (die Abstände gleichen der Grundfrequenz) mit verschiedenen Methoden berechnet werden.

Formant: Ein Formant ist aus der Perspektive der akustischen Analyse ein Frequenzbereich, in dem Obertöne stärker erscheinen als in anderen Frequenzbereichen. Formanten dürfen nicht mit Obertönen verwechselt werden. Vor allem die Vokale werden durch ihre Formantstruktur charakterisiert. In Abbildung 3.8 werden die ersten drei Formanten des [iː] in „liegen“ [liːgən] als Spektrum (links) und als Spektrogramm (rechts) visualisiert, die in dieser Aufnahme einer weiblichen Stimme bei 330 Hz, 2700 Hz und 3700 Hz liegen. Für die Bestimmung des Vokals sind die ersten beiden Formanten am wichtigsten: Bei [iː] liegen sie weit auseinander. Bei [uː] liegen sie eng beieinander. In Abbildung 3.8 (rechts) wird außerdem die Grundfrequenz gezeigt. Die Formantfrequenzen sind prinzipiell unabhängig von der Grundfrequenz; daher können verschiedene Vokale auf derselben Tonhöhe gesprochen werden, und ein Vokal auf verschiedenen Tonhöhen.

Computerphonetische Methoden

Neben der Signalanalyse können weitere Verarbeitungen von annotierten Daten mit nicht-signalverarbeitenden Methoden vorgenommen werden. Beispiele sind die Erstellung von Korpuslexika und Diphon- und Triphonlisten, Berechnung der relativen Häufigkeit annotierter Einheiten, das Trainieren von Hidden-Markov-Modellen (HMM) in der Spracherkennung, die Analyse von Dauerrelationen zwischen annotierten Silben, und die Grundfrequenzmuster auf betonten Silben. Solche Analysen werden in vielen Arten von Anwendungen verwendet. Im Folgenden werden zwei Ansätze aus diesem Bereich angeführt, die computerlinguistisch besonders interessant sind.

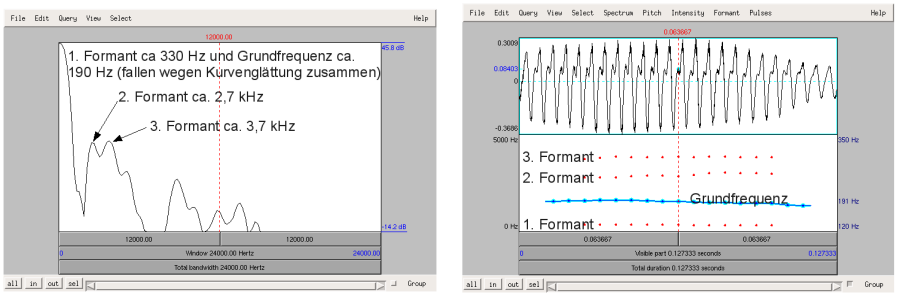

Abbildung 3.8: Spektralanalyse einer weiblichen Stimme. Links: Spektrum von [iː] mit Angabe der ersten drei Formanten. Rechts: Oszillogramm und Visualisierung der Formanten von [iː] sowie Grundfrequenzspur.

Lineare Zeitrelationen: Zur Phonetik des Rhythmus sind sehr viele Arbeiten vorhanden, bei keiner ist es aber jemals gelungen, den **Sprachrhythmus** vollständig als physikalisch-phonetisches Phänomen zu charakterisieren, ohne auf abstraktere linguistische Sprachstrukturen Bezug zu nehmen. Allgemein wird angenommen, dass das sprachliche Rhythmusempfinden eine komplexe kognitive Konstruktionsleistung ist und keine rein physikalische Regegelmäßigkeit. Dennoch werden physikalische Maße für Dauerrelationen benötigt.

Einer der bekannteren neueren Maße ist der **Pairwise Variability Index**, **PVI** (s. Literaturabschnitt 3.1.4), der über die Dauerrelationen benachbarter phonetischer Einheiten berechnet wird (Taktsequenzen, Silbenfolgen, vokalische oder konsonantische Segmentfolgen wurden in der Literatur untersucht). Der PVI ist das Mittel der normierten Differenzen zwischen den Längenunterschieden von relevanten Einheiten. Er basiert auf einer bekannten Formel, die normalerweise die Homogenität einer Wertemenge bestimmen soll. Der PVI wird mit folgender Formel berechnet (zwischen den PVI-Varianten wird hier nicht unterschieden):

$$PVI = 100 \times \sum_{1}^{n-1} \frac{|d_i - d_{i+1}|}{(d_i + d_{i+1})/2} / (n-1)$$

(d_i bezeichnet die Dauer einer annotierten Signaleinheit)

Der PVI kann Werte von 0 (gleiche Längen) asymptotisch bis näherungsweise 200 (sehr unterschiedliche Längen) annehmen. Durch Anwendung dieser Formel auf Dauerwerte von vokalischen und intervokalischen Intervallen in Annotationen von Äußerungen in unterschiedlichen Sprachen wurden interessante Verteilungen der Dauerrelationen in diesen Sprachen festgestellt.

Ob die Formel tatsächlich Rhythmus beschreibt, ist bezweifelt worden: Ähnliche Verteilungen lassen sich auch ohne das Sprachsignal allein anhand der Anzahl von Phonemtypen in entsprechenden Sequenzen ermitteln. Zudem setzt die Formel Binarität im Rhythmus voraus, was nicht unbedingt gegeben ist, und lässt kontrastive Vokaldauer außer Acht. Schließlich ist die Formel als Modell zwar vollständig, aber nicht korrekt: Es ist leicht überprüfbar, dass derselbe Indexwert zwar durch alternierende binäre Rhythmen, aber auch aufgrund der Verwendung des absoluten Werts der Dauerdifferenz durch eindeutig arhythmische Sequenzen (z.B. geometrisch ansteigende oder fallende Sequenzen oder Mischungen dieser drei Möglichkeiten) erreicht werden kann. Der PVI wurde als **Rhythmusmodell** eingeführt, ist aber aus den angeführten Gründen als solches ungeeignet. Dennoch kann der PVI aufschlussreiche empirische Informationen über die temporale Struktur von Äußerungen liefern.

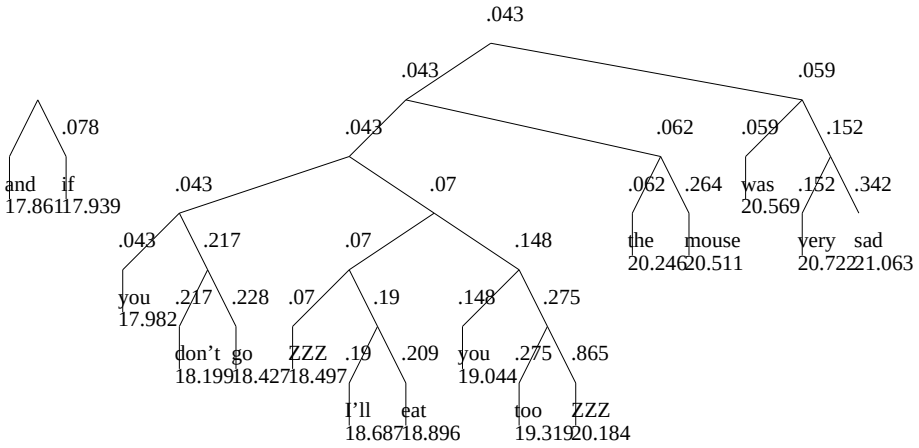

Abbildung 3.9: Durch numerisches 'Parsen' berechnete temporale Bäume über Wörter.

Zeitbäume: Dauerunterschiede zwischen benachbarten Einheiten können auch dazu verwendet werden, komplexere hierarchische Zeitrelationen als Zeitbäume zu ermitteln. Um eine Baumstruktur aufzubauen, werden im Gegensatz zur Berechnung des PVI nicht die absoluten Werte, sondern die rohen Werte der Dauerdifferenzen verwendet. Die Unterschiede zwischen positiven und negativen Differenzen werden gezielt eingesetzt, um wie bei einem Parser einen Baum aufzubauen. Eine Anwendung dieses Verfahrens, die interessante Korrelate mit syntaktischen Strukturen zeigt, wird in Abbildung 3.9 wiedergegeben.

Formale Methoden in der Computerphonologie

Merkmale, Attribute, Generalisierung, Defaults

Auf die Möglichkeit, distinktive Merkmale als Attribut-Wert-Paare darzustellen, wurde bereits eingegangen. In der Phonologie kann auch zwischen markierten und unmarkierten Werten eines Attributs unterschieden werden. In einem solchen Fall stellt der unmarkierte Wert z.B. den häufigsten Wert in Korpora, in Lexika oder unter den Flexionsformen eines Worts dar. Z.B. können durch die Neutralisierung der Wortunterscheidung an bestimmten Stellen in Silben oder Wörtern Elemente eines Spezifikationspaars mehrdeutig erscheinen.

Im Deutschen bewirkt beispielsweise die Auslautverhärtung von Obstruenten (Plosiven und Frikativen) eine Neutralisierung in den homophonen Formen „Rad“ /raːt/, „Rat“ /raːt/. Dass die Stämme sich morphophonematisch unterscheiden, zeigen die flektierten Formen „Rades“ /raːdəs/ und „Rates“ /raːtəs/. Aufgrund solcher Neutralisierungen gilt [- stimmhaft] als unmarkiert im Auslaut, [+ stimmhaft] als markiert. Es kann aber vorkommen, dass in anderen Kontexten die andere Spezifikation als unmarkiert gilt: In einigen deutschen Dialekten wird zwischenvokalisch /d/ sowohl für /d/ als auch für /t/ verwendet, z.B. „Kleid“ /klaɪt/ aber „bekleidet“ /bəglaɪdət/. In der **Generativen Phonologie** werden daher als Spezifikationen auch [u stimmhaft] und [m stimmhaft] verwendet, um diese Generalisierung zu erfassen, aus denen dann kontextspezifisch die entsprechenden Spezifikationen [+ stimmhaft] und [- stimmhaft] je nach Position in Silbe oder Wort abgeleitet werden.

Nicht alle Merkmale sind bei allen Phonemen gleichermaßen relevant. Nasale Konsonanten sind in den meisten Sprachen normalerweise stimmhaft; die Spezifikation des Merkmals [+ nasal, + stimmhaft] kann also zu [+ nasal] vereinfacht werden, wenn eine **Redundanzregel** (logisch gesehen eine Konditionalaussage: wenn nasal, dann stimmhaft) eingeführt wird:

$$[+\text{ nasal}] \rightarrow [+\text{ stimmhaft}]$$

Das Merkmal [± stimmhaft] kann unspezifiziert bleiben; das Merkmalsbündel bleibt also unterspezifiziert.

Wenn binäre (oder auch mehrwertige) Merkmale als Attribut-Wert-Paare modelliert werden, können sie mit den Operationen der Attribut-Wert-Logik (Unifikation, Generalisierung usw.) bearbeitet und computerlinguistisch implementiert werden. Solche Darstellungen wurden bereits eingeführt.

Die Modellierung von Merkmalen als Attribut-Wert-Paare eröffnet aber weitere Formalisierungsmöglichkeiten. Die Markiert-Unmarkiert-Gewichtung ist auch mit defaultlogischenDefaultlogik Mitteln behandelt und durch **Default-Unifikation** und **Default-Vererbung** computerphonologisch bearbeitet worden. In der defaultlogisch motivierten **Default-Vererbungssprache DATR** können beispielsweise Markiertheitsverhältnisse durch Unterspezifikation und Redundanzregeln durch Vererbung dargestellt werden:

```
KONSONANT:
        <KONSONANTISCH>   ==   +
        <VOKALISCH>       ==   -
        <KONTINUIERLICH>  ==   -
        <STIMMHAFT>       ==   -
        <>                ==   PHONEM.
PHONEM-P:
        <LABIAL>   ==   +
        <>         ==   KONSONANT.
PHONEM-B:
        <STIMMHAFT>   ==   +
        <>            ==   PHONEM-P.
```

Dieses Fragment der DATR-Implementierung eines **Vererbungsgraphen** modelliert folgende phonologische Generalisierungen:

1. Das Phonem /b/ erbt alle Merkmalswerte vom Phonem /p/, außer dem Stimmhaftigkeitswert [+ stimmhaft], der direkt zugewiesen wird und damit den Defaultwert überschreibt.

2. Das Phonem /p/ erbt alle Merkmalswerte von der natürlichen Klasse der Konsonanten, außer dem Wert [+ labial], der direkt zugewiesen wird und damit den Defaultwert überschreibt.

3. Die natürliche Klasse der Konsonanten spezifiziert alle unmarkierten Default-Werte der Konsonanten, alles Weitere wird von der hier nicht weiter spezifizierten Klasse der Phoneme geerbt. Konsonanten haben demnach typischerweise folgende Merkmale:

$$\begin{array}{l}\textsc{Konsonant}\\ \left[\begin{array}{ll} + & \text{konsonantisch}\\ - & \text{vokalisch}\\ - & \text{kontinuierlich}\\ - & \text{stimmhaft}\end{array}\right]\end{array}$$

Mit solchen Mitteln können ausdruckstarke lexikalische Relationen formalisiert und implementiert werden, die den Aufbau konsistenter Lexika unterstützen.

Reguläre Modelle

In den drei Domänen der Phonetik, der Phonologie und der Prosodie (und auch in der Morphologie) sind **reguläre Modelle** (d.h. **endliche Automaten** (Finite State Automaton, FSA), **endliche Übergangsnetzwerke**, **endliche Transduktoren** (Finite State Transducer, FST), **reguläre Grammatiken**, **reguläre Ausdrücke**) zu Standardmodellen für die Modellierung und Operationalisierung von kompositorischen Eigenschaften von Lautsequenzen geworden.

In der Computerphonologie werden reguläre Modelle zur Modellierung folgender Strukturen eingesetzt:

1. Silbenstrukturen,
2. phonotaktische Regeln (Morphemstrukturregeln, Redundanzregeln),
3. phonetische Interpretationsregeln,
4. die GEN-Komponente der Optimalitätstheorie (Generator des Suchraums für phonetische Interpretationen),
5. die EVAL-Komponente der Optimalitätstheorie (Constraintfilter zur Einschränkung des Suchraums für phonetische Interpretationen),
6. trainierbare, gewichtete stochastische Automaten in der Form von Hidden-Markov-Modellen in der Sprachtechnologie (siehe Unterkapitel 3.2).

Die wichtigsten Modellierungskonventionen, die die Verwendung von regulären Modellen in der Phonologie nahelegen, sind:

1. Die maximale Silbenlänge in allen Sprachen ist klein (zwischen 2 und 8); es wird also keinerlei Rekursion benötigt.
2. Die Phoneminventare in allen Sprachen sind endlich (und klein, mit ca. 20 bis 50 Elementen).
3. Das Kombinationspotential der Phoneme in Silben ist sehr beschränkt und kann z.B. mit endlichen Übergangsnetzwerken übersichtlich dargestellt werden.

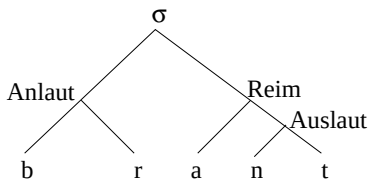

Abbildung 3.10: Baumgraph als Strukturbeschreibung einer Silbe.

4. Obwohl in der Phonologie oft Baumgraphen zur Darstellung von Silbenstrukturen verwendet werden, was einen komplexeren, kontextfreien Formalismus nahelegen könnte, haben diese Bäume eine maximale (und kleine) Tiefe (Abbildung 3.10).
5. Solche Baumgraphen können auch die linearen Einschränkungen quer zu den Baumverzweigungen nicht direkt oder anschaulich ausdrücken, wohingegen reguläre Modelle für diese Problemtik optimal geeignet sind.
6. Die Interpretation von Phonemen in linearen Kontexten kann durch endliche Transduktoren modelliert werden, entweder als Kaskaden von hintereinander geschalteten Automaten oder als parallele Automaten.

7. Auf Merkmalssystemen beruhende phonologische Theorien können von endlichen **Mehrbandautomaten** modelliert werden.

Ein Beispiel für die Verwendung von regulären Modellen ist die Formalisierung von phonotaktischen Regeln:

Regelnotation: $K \rightarrow \int \; / \; \$ \_\_ \left\{ \begin{array}{l} \left\{ \begin{array}{l} p \\ t \\ k \end{array} \right\} \left\{ \begin{array}{l} l \\ r \end{array} \right\} \\ m \\ n \end{array} \right\}$

Regulärer Ausdruck: $\int$ (((p|t|k) (r|l)) | (m|n))

Rechtslineare Grammatik:

$$\text{Silbe} \rightarrow \int \text{ KonSeq-1}$$

$$\text{KonSeq-1} \rightarrow \left\{ \begin{array}{l} \text{p KonSeq-2} \\ \text{t KonSeq-2} \\ \text{k KonSeq-2} \\ \text{m} \\ \text{n} \end{array} \right\}$$

$$\text{KonSeq-2} \rightarrow \left\{ \begin{array}{l} \text{r} \\ \text{l} \end{array} \right\}$$

Reguläre Menge: { $\int$pr, $\int$tr, $\int$kr, $\int$pl, $\int$tl, $\int$kl, $\int$m, $\int$n }

Die durch solche Morphemstrukturregeln oder Redundanzregeln angegebenen Vorkommensbeschränkungen sind wohl vollständig, indem alle Silben beschrieben werden, aber nicht korrekt, indem sie übergenerieren und Ketten beschreiben, die als Silben nicht vorkommen. Beispielsweise ist der Silbenanfang /$\int$tl/ im Deutschen nicht möglich, dies wird aber nicht direkt durch eine Redundanzregel ausgedrückt. Hierfür eignet sich eine als vollständiges Übergangsnetzwerk ausgeführte Beschreibung der ganzen Silbe eher als einzelne Regeln für Silbenteile. Mit einem solchen regulären Modell kann anhand eines relativ leicht zu implementierenden Interpreters die vollständige reguläre Menge auf einfache Weise formal und empirisch überprüft werden.

Abbildung 3.11 zeigt als Beispiel eines solchen Netzwerks ein nahezu vollständiges endliches Übergangsnetzwerk für englische Silben. Aus der Übergangskombinatorik läßt sich errechnen, dass die reguläre Menge, die durch dieses Netzwerk beschrieben wird, ca. 25.000 potentielle Silben des Englischen enthält.

Ein solches reguläres Modell der Phonotaktik lässt sich auf einfache Weise als Modell der phonetischen Interpretation verwenden, indem daraus ein endlicher Transduktor gemacht wird und die korrekten Allophone auf den entsprechenden Übergängen ihren Phonemen zugeordnet werden.

Phonetische Interpretationsregeln werden auch einzeln durch endliche Transduktoren modelliert. Folgende Regel beschreibt die Interpretation des deutschen Phonems /p/ in zwei verschiedenen Kontexten:

$$/\text{p}/ \rightarrow \left\{ \begin{array}{l} [\text{p}] \; / \; \$ \; \text{s} \_\_ \\ [\text{p}^{\text{h}}] \end{array} \right\}$$

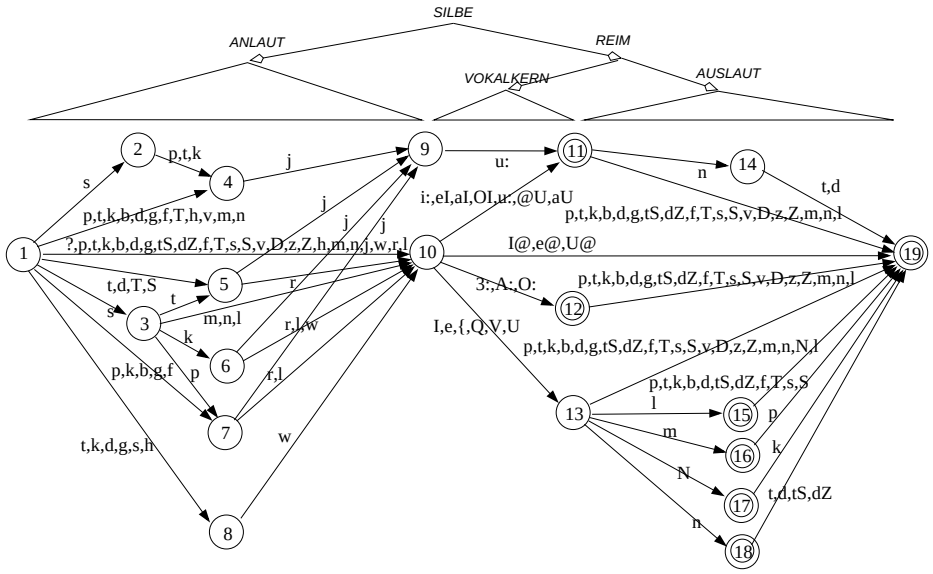

Abbildung 3.11: Endliches Übergangsnetzwerk als Grammatik für englische Silben mit Baumgraph als Generalisierung über die Graphstruktur (Beschriftung in SAMPA-Symbolen).

Die Reihenfolge der Regelalternativen ist defaultlogisch zu verstehen und bedeutet: Nach silbeninitialem /ʃ/ wird /p/ nicht behaucht, sonst wohl (oder, in der umgekehrten Reihenfolge: Typischerweise wird /p/ behaucht, nach /ʃ/ aber nicht). Als **Silbengrenze** wird hier „$“ verwendet.

Die Abbildung 3.12 zeigt einen Auszug aus einem endlichen Transduktor, der diese Regel modelliert. Am Silbenanfang läuft der endliche Transduktor an. Sollte ein /ʃ/ gefunden werden, wird es identisch übersetzt und es wird ein Plosiv gesucht. Wenn ein stimmloser Plosiv gefunden wird, wird dieser unbehaucht übersetzt und der Automat dann beendet, sonst wird der Automat direkt beendet. Sollte ein /ʃ/ nach dem Silbenanfang nicht gefunden werden, sondern ein stimmloser Plosiv, wird dieser behaucht übersetzt und der Transduktor dann beendet, sonst wird der Transduktor direkt beendet. Der vollständige Transduktor kann iterativ angewendet werden und relevante Kontexte ignorieren.

Im Ansatz von Kaplan und Kay werden solche Automaten in Kaskaden hintereinander angeordnet, um die Ableitung einer phonetischen Interpretation zu modellieren, wie dies auch für die Generative Phonologie möglich ist: Die Ausgabe eines Automaten bildet die Eingabe für den nächsten. Eine solche Kaskade kann durch eine Operation der Komposition zu einem einzigen Automaten zusammengesetzt werden. Im Ansatz von Koskenniemi, der als **Zweiebenenphonologie** (vgl. auch die Zweiebenenmorphologie) bekannt ist, werden endli-

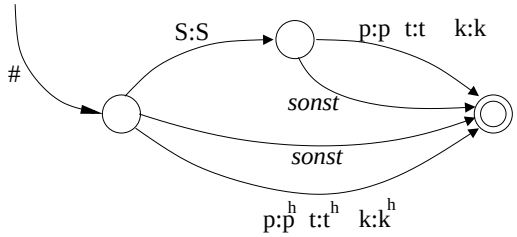

Abbildung 3.12: FST als partielles Modell der Plosivbehauchungsregel („sonst“ bedeutet die Komplementmenge der Phoneme, die auf den anderen Übergängen von einem Knoten erscheinen; der Übersicht halber werden die /p, t, k/-Übergänge zu einem Übergang zusammengefasst; „S“ bedeutet [ʃ]).

che Transduktoren verwendet, die parallel zueinander angewendet werden und aus logischer Sicht als eine Konjunktion von linearen Constraints verstanden werden. Die parallel anzuwendenden endlichen Transduktoren können ebenfalls durch eine Operation der Komposition zu einem einzelnen großen Automaten automatisch konvertiert werden.

Auch die Optimalitätstheorie lässt sich mit regulären Modellen modellieren. Die Grundidee der Optimalitätstheorie ist, dass die Abbildung von der phonologischen auf die phonetische Ebene nicht deterministisch vorgegeben ist, sondern dass ausgehend von einer phonologischen (lexikalischen) Repräsentation alle phonetischen Repräsentationsmöglichkeiten in einer Generatorkomponente GEN frei generiert werden, die dann durch eine geordnete (*ranked*) Menge von universellen Constraints CON in einer Evaluationskomponente EVAL gewichtet werden, die die Anzahl der Constraint-Verletzungen registriert, woraus schließlich die Übersetzung mit den wenigsten Constraint-Verletzungen als die optimale Übersetzung gewählt wird. Auf diese Weise wird der Suchraum für phonetische Interpretationen durch die einzelnen Constraints Schritt für Schritt verkleinert. Die Methode stammt aus der Constraintlogik und stellt im Prinzip einen Formalismus mit einer klaren Semantik dar. Karttunen hat als erster festgestellt, dass die einzelnen Constraints in der Constraint-Menge CON wie phonologische Regeln durch endliche Transduktoren modellierbar sind, ergänzt durch eine zusätzliche Default-Operation . Dieser Modellierungsansatz ist seitdem in der **Finite State Optimality Theory** vielfach angewendet worden, auch für die Modellierung der GEN-Komponente.

Ein wichtiger, aber bislang weniger bekannter Anwendungsbereich für reguläre Modelle ist die Prosodie, sowohl auf Satz- und Diskursebene als Intonationsmodelle, als auch in der phonetischen Interpretation von Tonfolgen in Tonsprachen.

Reguläre Modelle für die Intonation wurden in den 1970er Jahren von Fujisaki für Japanisch sowie von der niederländischen Arbeitsgruppe am Eindhovener Instituut voor Perceptie Onderzoek der Fa. Philips für Niederländisch und eine Reihe anderer Sprachen entwickelt. Das bekannteste reguläre **Intonationsmodell** wurde 1980 von Pierrehumbert für das Englische entwickelt (eine vereinfachte Version wird in Abbildung 3.13 gezeigt). Das Terminalvokabular des Modells besteht aus einer Relation über eine Menge von Tonbuchstaben {H, L} (für Hoch- und Tieftöne) und einer Menge diakritischer Zeichen {%, ∗}, die den Grenzton einer linguistischen Einheit („%“) oder eine Silbenbetonung („∗“) kennzeichnen.

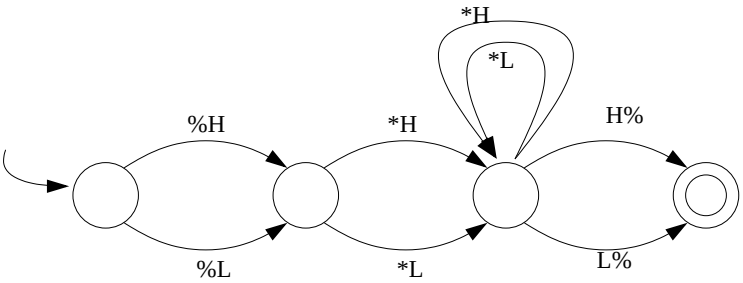

Abbildung 3.13: Vereinfachter FST für die Intonationsmodellierung.

Tonsequenzen in typologisch unterschiedlichen Tonsprachen wurden ebenfalls mit regulären Modellen beschrieben. In den meisten Niger-Kongo-Sprachen von West-, Zentral- und Südafrika werden Wörter nicht nur durch Phonemsequenzen voneinander unterschieden, sondern auch durch Töne — Silbenmelodien — mit phonematischer Funktion. Beispielsweise bedeutet in der Anyi-Sprache (Elfenbeinküste) das Wort „anouman“ /anʊmã/ mit steigender Tonkontur „Vogel“ und mit fallender Tonkontur „gestern“. Die Konturen werden als Sequenzen einzelner Töne analysiert, die eine bestimmte Tonhöhe relativ zu vorangegangenen **Tönen** einnehmen. Abbildung 3.14 zeigt einen endlichen Transduktor mit drei verschiedenen Kantenbeschriftungen, die die Operationen über solche Sequenzen für typische afrikanische Zweitonsprachen anzeigen (mit Namen der Tonregeln, die in der Literatur geläufig sind; phonetischen Interpretationen, die von den Übergängen des endlichen Transduktors modelliert werden; Dreibandoperationen, die numerische Werte und Operationen anzeigen, die hier nicht weiter kommentiert werden).

Die numerische Beschriftung erzeugt eine Annäherung an den Grundfrequenzverlauf, die weiterverarbeitet werden muss, um eine realistische Detailkontur, z.B. für die Sprachsynthese, zu erzeugen. Ein solches Modell kann auch mit einigen Modifikationen entsprechend Abbildung 3.5 für die Berechnung von Grundfrequenzverläufen in Akzent- bzw. Intonationssprachen verwendet werden. Für die Detailberechnung der Grundfrequenz werden jedoch komplexere Modelle, z.B. die Modelle von Fujisaki oder Hirst, angewendet.

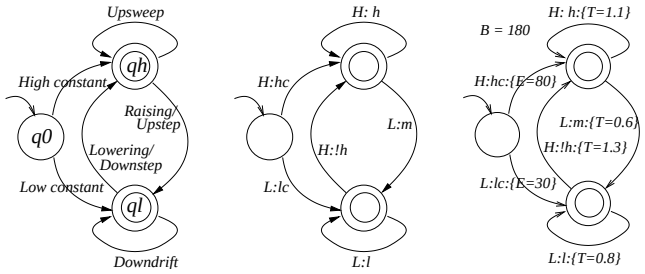

Abbildung 3.14: Endliche Transduktoren für die Tonsequenzierung in Niger-Kongo-Sprachen. (Großbuchstaben für phonologische Eingabetöne: H = Hochton, L = Tiefton; Kleinbuchstaben für phonetische Ausgabetöne: hc = konstante höhere Ansatzfrequenz, h = evt. steigende Sequenz von höheren Freqzenzen, m = angebobene tiefere Frequenz, l = evt. fallende Sequenz von tieferen Freqzenzen, !h = heruntergesetzte „downstepped
“ höhere Frequenz).

3.1.4 Zusammenfassung und weitergehende Lektüre

In diesem Beitrag werden zentrale Aspekte der Phonetik und Phonologie soweit besprochen, wie sie für gängige Forschungs- und Entwicklungsarbeiten in der Computerlinguistik und Sprachtechnologie erforderlich sind. Der Beitrag fängt mit theoretischen Konzepten aus der Phonetik und Phonologie an und stellt computerlinguistische Ansätze als Modelle für diese Konzepte vor. Die inhaltlichen linguistischen Fragestellungen werden als Aufgabenbereiche dargestellt, für die Lösungen mit empirischen und formalen computerlinguistischen, phonetischen und sprachverarbeitenden Methoden angeboten werden.

Weitere Informationen zu den computerlinguistischen und sprachtechnologischen Modellen, die in diesem Beitrag vorkommen, werden in anderen Kapiteln des Handbuchs besprochen, insbesondere im Unterkapitel 3.2.

Dieser Beitrag konkurriert nicht mit der reichhaltigen, hauptsächlich englischsprachigen Einführungs- und Handbuchliteratur zur computerlinguistischen und sprachtechnologischen Modellierung in der Phonologie und Phonetik. Eine Auswahl dieser spezialisierten technischen Literatur wird als weiterführende Lektüre zur Phonetik, Phonologie, Prosodie sowie zu einigen der spezielleren Modellierungstechniken in diesen Bereichen im Folgenden, nach Themen gegliedert, angeführt.

Phonologie: Einen Überblick über neuere Entwicklungen in der Phonologie bietet Hall (2000). Einen anspruchsvollen Einstieg in die Computerphonologie, vor allem unter Berücksichtigung regulärer Modelle, mit Anwendungen in der Sprachtechnologie, gibt Carson-Berndsen (1998).

Phonetik: Einen ersten Einstieg in die Phonetik bieten Pompino-Marschall (2003) und Ashby und Maidment (2005). Wesentlich mehr Details, recht anschaulich erklärt, sind in Reetz (2003) zu finden, während Coleman (2005) einen eher technischen Zugang bietet.

Prosodie: Die Sammelbände (Cutler und Ladd 1983) und (Gibbon und Richter 1984) beschreiben Ergebnisse der klassischen interdisziplinären Prosodieforschung in Überblicken. Neuere Forschungen aus einer phonologischen Perspektive werden in Ladd (2008) und aus interdisziplinären Perspektiven in Sudhoff et al. (2006) präsentiert.

Empirische Methoden, Ressourcen: Die aus dem europäischen EAGLES-Standardisierungsprojekt entstandenen Handbücher (Gibbon et al. 1997 und Gibbon et al. 2000) bieten einen systematischen Überblick über empirische Methoden und Evaluationsverfahren in der Sprachtechnologie, die auch für empirische Verfahren in der Computerlinguistik relevant sind. In Draxler (2008) werden sehr detaillierte Angaben zur Untersuchung von Korpora gesprochener Sprache angeboten.

Formale Methoden: Das Standardwerk (Jurafsky und Martin 2009) enthält eine Fülle von Angaben zu formalen Methoden in vielen Bereichen der Sprachtechnologien. Einen praktischen Zugang zum Programmieren (mit Python) für viele Bereiche der Computerlinguistik einschließlich Aspekte der Computerphonologie ist in Bird et al. (2009) zu finden.

Sprachtechnologien: Integrative Ansätze zu verschiedenen Teildisziplinen in den Sprach- und Texttechnologien werden in den Sammelbänden (Wahlster 2000 und Wahlster 2006) beschrieben.

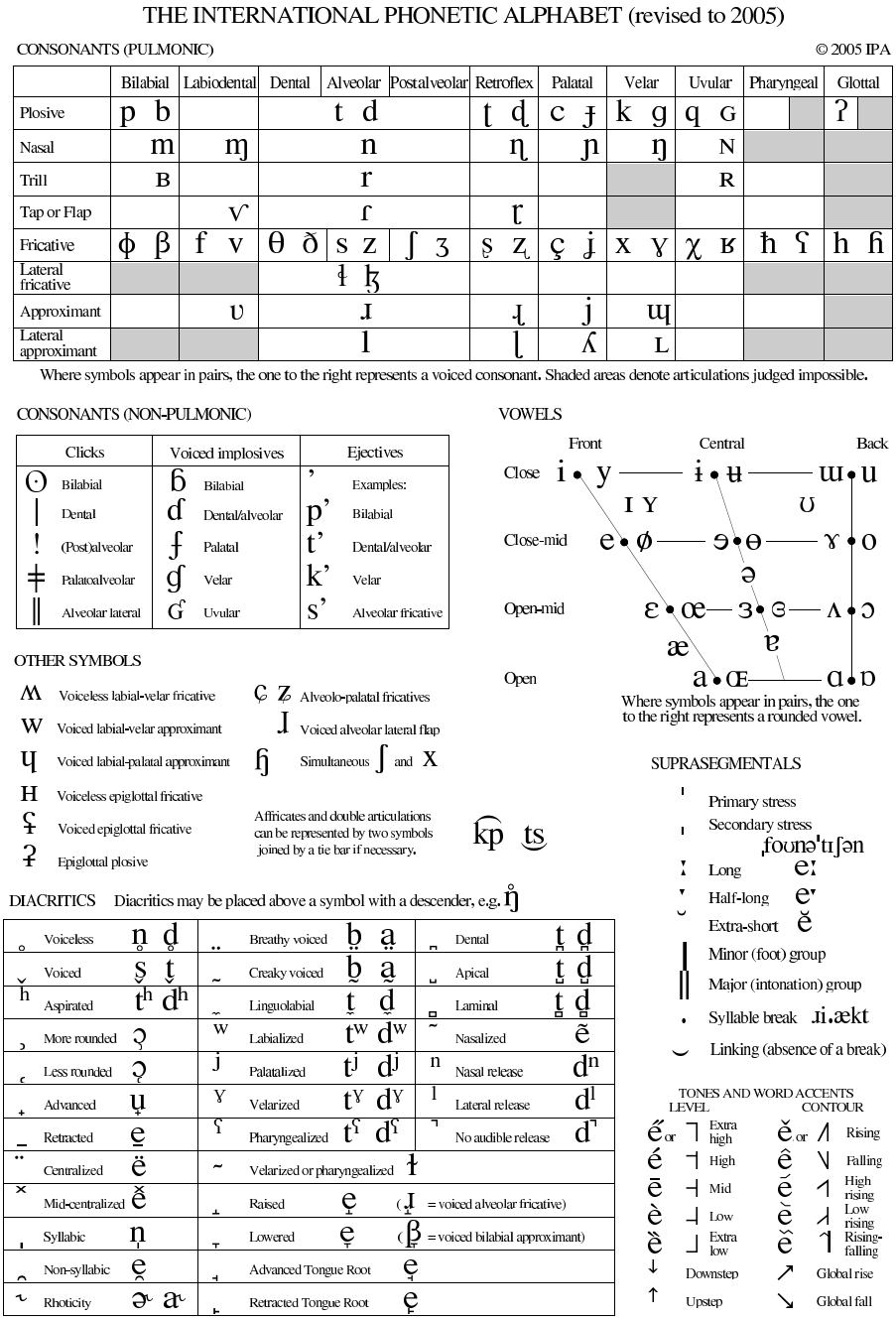

Abbildung 3.15: Symboltabelle der IPA: Konsonantensymbole, Vokalsymbole, Spezialsymbole, diakritische Zeichen zur Modifikation von Artikulationsort und -art.