Regelsystem zur Generierung von Aussprachevarianten

Bestandteil des Münchener Automatischen Segmentationssystems (MAUS)

M.-B. Wesenick

Ludwig-Maximilians-Universität München

Memo Nr. 96
November 1995

November 1995

Maria-Barbara Wesenick

Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstraße 3/II
80799 München

Tel.: (089) 2180 - 2810
e-mail: wesenick@sun1.phonetik.uni-muenchen.de

Gehört zum Antragsabschnitt: 14 VERBMOBIL/PHONDAT

Die vorliegende Arbei t wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 102 L/4 gefördert. Die Verantwortung für den Inhalt liegt beim Autor.

Inhaltsverzeichnis

1 Einführung 4

2 Korpus phonetisch-artikulatorisch motivierter Regeln 6

2.1 Die Erfassung von Variabililät der Sprache in Regeln 6

2.2 Quellen der Beobachtung von phonetischen Prozessen 7

2.3 Symbolinventar 7

3 Phonetische Prozesse 10

3.1 Assimilationen 10

3.2 Elisionen 12

3.3 Substitution von Plosiven durch den Glottalverschluß 13

3.4 Vokalisierung von /l/ 14

3.5 Sproßkonsonant 14

3.6 Lenisierung 14

3.7 /r/-Realisationen 15

4 Regelanwendung 16

5 Zusammenfassung 18

6 Literatur 18

Korpus phonetischer Regeln als Bestandteil eines automatischen Segmentationssystems

Im Rahmen von VM-TP 14.7 wird im Institut für Phonetik und Sprachliche Kommunikation München ein automatisches System zur Segmentation von gesprochener Sprache entwickelt. Ein wesentlicher Bestandteil des Segmentationssystems ist ein Korpus phonetischer Regeln, nach denen von kanonischen Formen hypothetische Aussprachevarianten abgeleitet werden können. Dieser Regelkorpus soll im folgenden beschrieben werden.

1 Einführung: Variabilität von gesprochener Sprache als Problem in der technischen Sprachverarbeitung

Ziel des Teilprojekts ist es, ein System zu entwickeln, mit dem große Mengen von Sprachmaterial automatisch segmentiert werden können. Voraussetzung ist, daß für die zu segmentierende Äußerung neben den Signaldaten eine entsprechende orthographische Form vorliegt. Das System liefert nach der Bearbeitung eine breite phonematische Transkription der Äußerung und die jeweiligen Segmentgrenzen.[1]

Anwendung finden soll das System in Bereichen, in denen umfangreiches segmentiertes Sprachmaterial benötigt wird. Dazu gehört beispielsweise das Gebiet der automatischen Spracherkennung, wenn es darum geht, Trainingsmaterial zu erzeugen. Auf der Grundlage von sehr großen Sprachdatenbanken mit segmentiertem Material können aber auch großangelegte empirische phonetische/phonologische Untersuchungen durchgeführt und beispielsweise Wortformenstatistiken erstellt werden.

Segmentationen werden bisher gewöhnlich manuell hergestellt, was aber in Anbetracht sehr großer Sprachdatenbanken wegen des äußerst großen Zeitaufwands nahezu unmöglich ist. Ein System, das ausreichend zuverlässig automatische Segmentationen herstellen kann, ist demnach wünschenswert.

Das Münchener Segmentationssystem basiert auf Hidden Markov Modellen (HMM) der Phoneme des Deutschen (und einigen Zusatzmodellen).[2] Über die Orthographie kann die kanonische Form einer Äußerung in einem Lexikon aufgefunden werden.[3] In einer früheren Version des Segmetationssystems wurden die einzelnen Lautmodelle der kanonischen Form mit dem Signal der zugehörigen Äußerung verglichen und dann zugeordnet. Auf diese Weise erhielt man eine Abbildung der kanonische Form einer Äußerung auf das Sprachsignal.

Nun ist aber Kommunikation mit gesprochener Sprache durch große phonetische Variabilität der lautsprachlichen Äußerungen gekennzeichnet. Eine kanonische Form symbolisiert nur eine von sehr vielen Möglichkeiten der Realisation ein und derselben sprachlichen Einheit.

Die gleiche Äußerung läßt sich von eineim einzelnen Sprecher nicht exakt wiederholen, d.h. die Äußerungen eines Sprechers variieren im Vergleich untereinander. Variabilität zeigt sichaber auch im Vergleich unterschiedlicher Sprecher (Intersprecher- vs. Intrasprecher-Variabilität).

---

1. Zur Terminologie: Eine Transkription mit zugehörigen Segmentgrenzen wird zusammengefaßt als *Segmentation* bezeichnet.
2. Zu den technischen Einzelheiten siehe [1]
3. Zur Terminologie: Unter *kanonischer Form* ist diejenige Ausspracheform einer Äußerung zu verstehen, die als einzige Standardform aus mehreren möglichen Zitierformen (Standardaussprache isolierter Wörter) arbiträr ausgewählt wurde. Sie wird mit Hilfe von Zeichen eines phonetischen Alphabets notiert.

Gründe für Intersprecher-Variabilität liegen z.B. in der unterschiedlichen Anatomie der Sprachtrakte, was sich auf die Eigenschaften der Sprachsignale auswirkt. Dazu kommen Faktoren wie regionale oder sozio-kulturelle Einflüsse.

Bei Intrasprecher-Variabilität könnte man z.B. unterschiedliche Kommunikationssituationen und die damit verbundenen unterschiedlichen Sprechstile oder ganz allgemein physische und psychische Verfassung als Gründe für Variabilität aufführen.

Bei der technischen Sprachverarbeitung und insbesondere auf dem Gebiet der Sprachanalyse ist die durch unterschiedlichste Faktoren hervorgerufene Variabilität der lautsprachlichen Äußerungen ein großes Problem. Man kann dem begegnen, indem man durch Wortmodelle die Variabilität statistisch miterfaßt. Sie ist dann in den Modellen erfaßt und muß nicht explizit berücksichtigt werden. Arbeitet man jedoch mit kleineren Einheiten und möchte man Informationen über Segmentrealisationen und ihre Lokalisation im Signal gewinnen, ohne daß eine exakte Transkription der Äußerung vorliegt (wie es beim Münchener Segmentationssystem der Fall ist), muß Wissen über phonetische Prozesse, die der Variabilität zugrunde liegen, verarbeitet werden.

So konnte eine entscheidende Verbesserung bei der Weiterentwicklung des Münchener Segmentationssystems dadurch erreicht werden, daß beim Vergleich von HMM mit dem Sprachsignal mögliche Aussprachevarianten der kanonischen Form alternativ angeboten werden. Jetzt wählt das System diejenige Ausspracheform einer Äußerung, die sich im Vergleich zu allen alternativen Formen optimal auf das entsprechende Signal abbilden läßt.

Die Aussprachevarianten werden durch Anwendung von Regeln eines speziell entworfenen Regelkorpus auf die entsprechende kanonische Form erzeugt, wobei auch Reduktionen an Wortgrenzen berücksichtigt werden. Die Regeln sind in einer Weise formuliert, daß soweit wie möglich alle denkbaren Varianten hergeleitet werden, wobei auch unmögliche Varianten in Kauf genommen werden. Die Regeln basieren auf beobachteten artikulatorischen Reduktionen in gesprochener Sprache, die mit Hilfe eines auf Prolog basierten Datenbankmanagementsystem systematisch vornehmlich am PhonDat II - Zugauskunftkorpus untersucht worden sind.

Ziel war es, durch die Regeln so viele Varianten zu erzeugen, daß sich unter ihnen alle Formen befinden, die unter Umständen produziert werden können (im Rahmen einer deutschen Standardlautung ohne zu große dialektale Einflüsse). Das bedeutet natürlich, daß bei der Ableitung auch solche Formen entstehen, die möglicherweise niemals in Wirklichkeit auftreten. Ein solches Vorgehen ist jedoch notwendig, da nur solche Varianten, die indirekt durch die Regeln vorgesehen sind, am Ende im Signal gefunden werden können. Bleibt eine mögliche Ausspracheform unberücksichtigt, so wird das System dem entsprechenden Signalabschnitt die Variante zuordnen, die der tatsächlichen Form, die das System ja nicht kennt, am ähnlichsten ist. Unmögliche Varianten bleiben dagegen unberücksichtigt, denn die zutreffende Form wird dem zu beurteilenden Signalabschnitt immer ähnlicher sein.

Durch Anwendung des automatischen Segmentationssystems und Beurteilung der Resultate wird das Regelsystemnach und nach verfeinert und verbessert. Nur durch große empirische Untersuchungen und Analyse umfangreichen Sprachmaterials können mögliche phonetische Prozesse in gewisser Vollständigkeit erfaßt (in das Regelsystem aufgenommen) und Aussagen über tatsächliche Aussprachevarianten des Deutschen gemacht werden.

Der Unterschied zu freilaufenden Phonemerkennungssystemen, bei denen jedes Modell jedem Signalabschnitt zugeordnet werden kann, besteht darin, daß durch die Vorgabe der kanonischen Form und ihren Varianten der Suchraum für das beste Lautmodell noch relativ stark eingeschränkt ist. Das bedeutet, daß die Rechengeschwindigkeit höher ist und zudem die resultierende Segmentation durch fundierte Hypothesen über die möglichen Aussprachevarianten verläßlicher sind.

2 Korpus phonetisch-artikulatorisch motivierter Regeln

2.1 Die Erfassung von Variabililät der Sprache in Regeln

Variabilität lautsprachlicher Äußerungen ist abhängig z.B. vom unmittelbaren Kontext der Äußerung, vom Sprechtempo, vom Sprechstil und der jeweiligen Kommunikationssituation oder von der Komplexität des Äußerungsinhalts. Sie zeigt sich in den unterschiedlichen Möglichkeiten der Realisation einer geplanten Äußerung, die auf einer kontinuierlichen Skala von sehr deutlich, langsam und genau bis zu stark reduziert und schnell liegen kann. Man spricht von dieser Skala auch als Hyper-Hypo-Kontinuum gesprochener Sprache[4]. Hinter dem Bestreben des Sprechers, eine Äußerung möglichst deutlich und mit präziser Artikulation zu produzieren (sogenannte "Hyperformen"), steht die Absicht, für seinen Kommunikationspartner möglichst verständlich zu sein, um das Kommunikationsziel erreichen zu können. Dem entgegen steht die generelle Tendenz, den artikulatorischen Aufwand so niedrig wie möglich zu halten und Bewegungsabläufe mit möglichst wenig Energie durchzuführen, was zur Produktion der "Hypoformen" führt.

Zwischen diesen beiden Zielen, der Einschränkung des artikulatorischen Aufwands und dem Erreichen des Kommunikationsziels, muß der Sprecher im Laufe einer Äußerung ständig Kompromisse finden. Das Ergebnis ist gesprochene Sprache, deren Äußerungen in einem Kontinuum von Hypo- und Hyperformen variieren.

Mit dem Regelsystem zur Generierung von Aussprachevarianten soll versucht werden, die unterschiedlichen Formen aus dem Hypo-Hyper-Kontinuum auf symbolischer Ebene unter Berücksichtigung der artikulatorischen Vorgänge zu erfassen. Dabei beziehen wir uns nicht auf mögliche Formen eines Einzelwortes als sprachliche Einheit, sondern auf eine gesamte Äußerung, die in der orthographischen Form als Satz festgehalten ist.

Die extremste Hyper-Wortform, d. h. diejenige Form, die in isolierter Aussprache bei größter artikulatorischer Genauigkeit entsteht, soll in der kanonischen Form als symbolische Repräsentation zum Ausdruck kommen (die kanonische Form ist nicht unbedingt die üblichste Aussprachevariante einer sprachlichen Einheit). Alle anderen möglichen Formen des Kontinuums bis zu den Hypo-Formen sind durch Einsparung artikulatorischen Aufwands gekennzeichnet und sollen als ebenfalls symbolische Repräsentationen durch Regeln von der kanonischen Form abgeleitet werden. Die Regeln beschreiben demnach auf abstrakte Weise segmentale Unterschiede zwischen der kanonischen Form und Aussprachevarianten, die durch die Reduktion des artikulatorischen Aufwands zustande kommen.[5]

Die unterschiedlichen Formen variieren in jeder Hinsicht entlang eines Kontinuums, d.h. insbesondere, daß die Reduktion der Artikulationsbewegungen kontinuierlich ist, was instrumentelle Untersuchungen deutlich zeigen.[6] Um Aussprachеformen schriftlich festhalten zu können, bedient man sich jedoch eines phonetischen Alphabets mit diskreten Symbolen. Auch die im automatischen Segmentationssystem verwendeten HMM für Sprachlaute sind phonetischen Symbolen zugeordnet. Diese Symbole aber repräsentieren Kategorien und können Abstufungen innerhalb von Kategorien nicht erfassen. Da Aussprachevarianten wiederum auch symbolisch beschrieben werden müssen und Unterschiede zur kanonischen Form nur in den Symbolen zu findensind bzw. in der Wahl eines anderen HMM, können diese kontinuierlichen Veränderungen im Grunde durch Regeln nicht erfaßt werden. Die Vorgehensweise, zu beobachtende Reduktionsprozesse in Form von Regeln zu formulieren, folgt ausschließlich pragmatischen Forderungen bei

4. Dieser Begriff stammt von B. Lindblom, der ihn im Rahmen der Darlegung seiner H&H(hyper-hypo-articulation)-Theorie eingeführt hat. Siehe dazu Lindblom [4].
5. Eine Ausnahme bilden allerdings z.B. die unterschiedlichen /r/-Realisationen, die z.T. beliebig variieren können.
6. siehe dazu z.B. Kühnert [5]

der Entwicklung eines automatischen Systems. Die Regeln stellen in keiner Weise den Versuch dar, ein System von wirklichen Prozessen wiederzuspiegeln, das etwa beim Sprecher/Hörer eine interne Realität hätte. Die Regeln sind ein Werkzeug, um die symbolische Repräsentation der kanonischen Form so zu verändern, daß andere Formen entstehen, die idealerweise eine symbolische Repräsentation von tatsächlich möglichen Ausspracheformen darstellen. Es ist irrelevant, daß sie mit tatsächlichen Prozessen bei der Produktion von lautsprachlichen Äußerungen wenig gemein haben, da sie praktisch dazu beitragen können, die Analyse großer Sprachdatenmengen möglich zu machen.

2.2 Quellen der Beobachtung von phonetischen Prozessen

Die Informationen über das, was die Regeln beschreiben müssen, stammen in erster Linie aus Transkriptionen, die in der phonetischen Datenbank des IPSK München enthalten sind. Die Datenbank enthält große Teile des PhonDat-II-Korpus, der 1992 in Zusammenarbeit der phonetischen Institute in Kiel, Bonn und München entstanden ist und dessen Sätze aus dem Szenario der Zugauskunft stammen.[7] Die symbolischen Daten sind in einer Datenbankkomponente der Prolog-
umbebung *eclipse* [ECRC 93] gespeichert. Durch Zugriff auf die Daten mit Prolog über Datenbank-Interface Prädikate sind systematische Analysen der Transkriptionen möglich. Man kann dabei sowohl überblicksartig, als auch sehr gezielt die Beziehung von kanonischer Form zur Realisierung untersuchen.

Auf diese Weise ist der grundlegende Teil von Regeln entstanden, mit denen die geläufigsten Reduktionserscheinungen beschrieben und auf kanonische Formen übertragen werden können. Die entstehenden Formen repräsentieren Formen, die durch reduzierte artikulatorische Abläufe bei der lautsprachlichen Kommunikation entstehen.

Weiterhin sind aber im Korpus auch solche Regeln enthalten, die aufgrund von Annahmen über mögliche phonetische Prozesse aufgestellt worden sind. Solche Regeln aufzunehmen war nötig, da das untersuchte Phondat-Material keinen großen Wortschatz umfaßt und nicht phonetisch ausgewogen ist. Diese Regeln basieren z.T. auf allgemeinen Erkenntnissen der Phonetik, die in der Literatur zu finden sind, aber auch aus eigenen Hyphothesen und Vermutungen über mögliche artikulatorische Abläufe.

Dieses Vorgehen ist in Hinblick auf das statistische Verfahren des automatischen Segmentierers zu rechtfertigen und entspricht dem Grundgedanken, so viele zur kanonischen Form alternative Varianten wie möglich anzubieten. Aus der Menge der hypothetischen Formen wird durch Viterbi-Alignment und Angaben über die Wahrscheinlichkeit einer Form, die zutreffendste Form ausgewählt. Ist aber die “richtige” Form nicht vorgesehen, kann sie auch nicht gefunden werden.

So wird der Kompromiß eingegangen, nicht ausschließlich belegbare Formen durch die Regeln zu erzeugen, sondern viele hypothetische, um sicher zu gehen, nahezu alle möglichen Formen erfaßt zu haben. Der Regelkorpus umfaßt in seiner gegenwärtigen Form 1550 Regeln.

2.3 Symbolinventar

Die verwendeten Symbole stammen aus dem phonetischen SAM-Alphabet zur Beschreibung des Deutschen.[8] Darin ist prinzipiell für jedes Phonem der deutschen Sprache ein Zeichen vorgesehen.

---

7. zum PhonDat - Projekt siehe [6]
8. Das SAM-Alphabet für das Deutsche ist abgedruckt in Pompino-Marschall [6].

Die Symbole stehen damit zunächst einmal für Phoneme und wir haben es bei der Transkription mit einer phonologischen Repräsentation des Gesprochenen zu tun. Es wird jedoch betont, daß mit der Verwendung von Transkriptionssymbolen, die nach phonologischen Gesichtspunkten ausgewählt wurden, bei der Entwicklung des Segmentationssystems keine theoretischen Implikationen der Phonologie oder Psycholinguistik eingebracht werden. Das Vorgehen beim Umgang mit Sprachsignalen ist rein praktischer Natur und die verwendeten phonemähnlichen Einheiten haben prinzipiell eine rein signalbeschreibende Funktion. Sie haben sich als äußerst nützliches Werkzeug zur Beschreibung des akustischen Signals erwiesen und sind ausschließlich als solches zu verstehen.

komplette Liste der HMM bzw. der erlaubten Lautsymbole für die Regeln:

- Vollvokale:
  `a, a:, e:, I, i:, O, o:, U, u:, E, E:, 9, 2:, Y, y:`
- reduzierte Vokale:
  `@, 6`
- Diphthonge:
  `aI, aU, OY`
- Plosive:
  `p, b, t, d, k, g, Q`
- Frikative:
  `f, v, s, z, S, Z, C, j, x, h`
- Nasale:
  `m, n, N`
- Lateral, Vibrant:
  `l, r`
- und zwei Pausenmodelle

Es ist zu beachten, daß bei Plosiven eine Markierung der Aspiration nicht vorkommt. Zum Unterschied zwischen “stimmhaften” und “stimmlosen” Plosiven ist zu bemerken, daß im Deutschen das Merkmal “fortis” vs. “lenis” distinktiv ist, wobei Lenis-Plosive phonetisch stimmhaft oder phonetisch stimmlos sein können. Bei Verwendung von Symbolen für stimmhafte Plosive ist daher nicht die phonetische Stimmhaftigkeit beschrieben, sondern eine Lenis-Artikulation.

Dem Daueraspekt kann in statistischen Modellen der verwendeten Art nicht in befriedigender Weise Rechnung getragen werden. Ein Laut kann richtig erkannt werden, egal ob er 30 ms oder 100 ms dauert. Aus diesem Grunde ist eine Information über die Dauer eines Segments in den Regeln (z.B. durch Kennzeichnung von Geminaten) nicht enthalten. Das gleiche Segment steht daher nie zweimal hintereinander. Tritt beispielsweise im Wort ausschließen eine Assimilation in der Verbindung `sS` auf, ist der Frikativ zunächst einmal von längerer Dauer, als ein “normales” Segment. Die Regel müßte logisch lauten: `sS>SS`. Aus den genannten Gründen lautet sie jedoch `sS>S`.

2.4 Regelsyntax

Neben den Lautsymbolen werden in den Regeln folgende Sonderzeichen verwendet:

`!v`	Vokale
`!K`	Konsonanten
`!N`	Nasale
`#`	Wortgrenze
`&`	arbiträre Wortgrenze

Das Zeichen "#" zur Kennzeichnung einer Wortgrenze wird benötigt, da viele Lautkombinationen nur über Wortgrenzen hinweg überhaupt möglich sind. Wird "#" nicht verwendet, findet die Regel nur innerhalb eines Wortes und nicht an Wortgrenzen Anwendung. Das Zeichen "&" für eine arbiträre Wortgrenze wird verwendet, wenn nach einem Reduktionsprozess die Wortgrenze innerhalb eines Segments liegt. Im Fall kann man `/kan man/ -> /kaman/` liegt die Wortgrenze im Segment `/m/`, das Zeichen "&" wird konventionsgemäß diesem Laut vorangestellt: `/ka&man/`.

Eine Regel r_i, i = 0...N-1 des Regelkorpus besteht aus einem rechten und einem linken Teil, die durch ">" getrennt sind. Der linke Teil enthält einen Symbolstring $\mathbf{a_i} = <a_i(0), ... a_i(K_i - 1)>$, der einem Teilstring der kanonischen Form entsprechen muß. Auf der rechten Seite steht ein Symbolstring $\mathbf{b_i} = <a_i(0), ... a_i(K_i - 1)>$, der den variierten Symbolstring a_i der kanonischen Form repräsentiert. $a_i(k)$ und $b_i(l)$, $k = 0...K_i$, $l = 0...L_i$ sind Symbole des Phonetischen SAM-Alphabets und die oben aufgeführten Sonderzeichen, ausgenommen "&" für die rechte Seite.

Wird auf der linken Seite der Regel eines der vier Sondersymbole `!v, !K, !N` oder `#` verwendet, muß es auch auf der rechten Seite an entsprechender Stelle auftauchen. Für "#" kann auf der rechten Seite im gegebenen Falle "&" stehen. Auf der linken Seite kann "&" nicht stehen.

Jeder Regel wird eine Ziffer vorangestellt, die aber weiter keine Bedeutung hat (diese Syntaxvorgabe ist aus älteren Softwareanforderungen entstanden). Sie dient dazu, gegebenenfalls die Anwendung von Regeln auf bestimmte markierte Regeln zu beschränken, bzw. bestimmte Regeln auszuschließen. Die Markierung kann mit Hilfe der Ziffer vorgenommen werden.

Beispiele:

`1nf>mf`	die Symbole `nf` können durch `mf` ersetzt werden, aber nur wenn sie im Wortinneren vorkommen
`1#pf>#f`	`pf` am Wortanfang kann durch `f` ersetzt werden
`1p#j>p#C`	an Wortgrenzen kann `pj` zu `pC` werden
`1g@n#>gN#`	`g@n` am Wortende kann durch `gN` ersetzt werden
`1t#t>&t`	Treffen zwei t an einer Wortgrenze aufeinander, so können sie durch ein einziges ersetzt werden. Die Wortgrenze ist dann arbiträr und wird konventionsgemäß vor das betreffende Segment gesetzt.
`1!vtp>!vQp`	`tp` kann nach Vokal zu `Qp` werden
`1!vr!K#>!v6!K#`	`r` kann am Wortende nach Vokal und vor Konsonant vokalisiert und durch `6` ersetzt werden

3 Phonetische Prozesse

3.1 Assimilationen

Unter Assimilation versteht man die Angleichung eines Segments an ein benachbartes Segment in einem oder mehreren Parametern. Die Assimilation kann den Ort oder die Art der Artikulation und die Stimmhaftigkeit betreffen und dabei total oder partiell auf ein folgendes Segment wirken (progressiv) oder auf ein vorhergehendes (regressiv). Assimilationen können innerhalb von Morphemen, aber auch über Morphem- und Wortgrenzen hinweg stattfinden.

3.1.1 Regressive Assimilation des Artikulationsorts

ausgewählte Regeln:	*beispielsweise in:*
`1tp>p`	[Mutprobe]
`1t#p>&p`	[mit Papier]
`1tk>k`	[mitkriegen]
`1t#k>&k`	[hat kurze]

In diesen Fällen wird der Artikulationsort des ersten Plosivs an den des folgenden Plosivs angeglichen als Folge der Abschwächung des ersten Plosivs am Ende einer kontinuierlichen Reduktion.

Im ersten Beispiel wird der apikale alveolare Plosiv an den bilabialen angeglichen. Apikale Laute sind von Assimilationen und Elisionen besonders betroffen, da die Zungenspitze als artikulierendes Organ sehr präzise ist, deswegen aber auch sehr genau kontrolliert werden muß. Der artikulatorische Aufwand ist verhältnismäßig hoch und wird häufig eingespart, wenn durch schnelleres oder weniger sorgfältiges Sprechen die Kontrolle über die Artikulationsorgane abnimmt. Auch in diesem zweiten Fall sind graduelle Abstufungen der Assimilation möglich. Die Lippen können sich beispielsweise schließen bevor der alveolare Verschluß gelöst worden ist, die Aspiration des ersten Plosivs entfiele dadurch. Auch eine Doppelartikulation der Plosive ist denkbar und eine abgestufte Einsparung der apikalen Geste.

Im zweiten Beispiel wird der alveolare Plosiv an den folgenden velaren angeglichen. Auch hier ist wieder ein apikaler Laut assimiliert worden. Und auch hier sind Übergänge in den Reduktionsstufen möglich. Der erste Plosiv könnte z.B. erst dann gelöst werden, wenn schon ein velarer Verschluß gebildet worden ist, so daß keine Aspiration entsteht, da kein Luftstrom vorhanden ist. Eine alveolar-velare Doppelartikulation ist möglich, genauso wie die graduelle Assimilation der apikalen Geste bis zu ihrem vollständigen Verschwinden.[9]

Die graduellen Veränderungen kommen wie erwähnt mit den zur Verfügung stehenden Symbolen nicht gut zum Ausdruck, sind jedoch implizit durch die Abstraktion in den Regeln enthalten.

3.1.2 Progressive Assimilation des Artikulationsortes

ausgewählte Regeln:	*beispielsweise in:*
`1b@n#>bm#`	[geben]
`1k@n>kN`	[Hakennase]
`1k@n#>kN#`	[trocken]
`1x@n#>xN#`	[Drachen]

Die progressive Ortsassimilation findet vornehmlich nach Schwa-Elision statt. So ist etwa im ersten Beispiel in der Silbe -b@n das /@/ ausgefallen. Jetzt treffen /b/ und /n/ aufeinander, wobei

9. siehe dazu die Untersuchungen von Barbara Kühnert [5].

der Nasal seinen Artikulationsort an den des Plosivs angleicht: Der alveolare Nasal wird nach einem bilabialen Plosiv bilabial realisiert.

3.1.3 Regressive Assimilation der Artiulationsart

ausgewählte Regeln:	*beispielsweise in:*
`1b@n>m`	[Abend]
`1gn>Nn`	[Magnet]
`1pf>f`	[Abfahrtszeit]

Assimilationen der Artikulationsart kommen am häufigsten als Nasalassimilationen vor. In den Wörtern der obigen Beispiele kann bei der Artikulation jeweils ein Segment an die Artikulationsart des folgenden Segments angeglichen werden. Im ersten Fall müßte zuvor das /@/ elidiert (b@n -> bn) und der Artikulationsort assimiliert werden (bn -> bm), ehe die Nasalassimilation stattfindet (bm -> m).

Im zweiten Beispiel wird der velare Plosiv unter dem Einfluß des nachfolgenden alveolaren Nasals gleichfalls zum Nasal. Führt man sich den artikulatorischen Ablauf vor Augen, so ist hier die Koordination des Velums und der Zunge je nach Artikulation unterschiedlich. Wenn der velare Plosiv bestehen bleibt, senkt sich das Velum erst, wenn der alveolare Verschluß gebildet wird. Entsteht jedoch kein Plosiv, senkt sich das Velum sehr früh, vor oder mit der Bildung des velaren Verschlusses. Die Zunge könnte dann in einer kontinuierlichen Vorwärtsbewegung den alveolaren Verschluß bilden, wobei es durchaus zur alveolar-velaren Doppelartikulation kommen kann oder zur Bildung eines palatalen Nasals.

Im dritten Beispiel wird vor einem Frikativ anstelle eines Plosivs ein Frikativ gebildet. Man muß sich auch hier wieder einen graduellen Assimilationsprozeß vorstellen: Der bilabiale Plosiv könnte labio-dental artikuliert werden, also an den Artikulationsort des Frikativs angeglichen werden. Er kann dabei relativ schwach sein und mit wenig Druck gelöst werden. Und weiterhin wird der Verschluß nicht mehr vollständig gebildet, so daß dann ein Frikativ entsteht. Er fällt mit dem folgenden Frikativ zusammen. Dieser gelängte Frikativ kann in seiner Dauer auf die Länge eines einfachen reduziert werden.

Auch hier werden die unterschiedlichen Zwischenstufen jeweils durch die in den Regeln bezeichneten statistischen Modelle abgedeckt.

3.1.4 Progressive Assimilation der Artikulationsart

ausgewählte Regeln:	*beispielsweise in:*
`1nd>n`	[Verbindung]
`1mb>m`	[Hamburg]
`1st#d>&s`	[hast du]

Bei der progressiven Assimilation der Artikulationsart beeinflußt ein Segment das nachfolgende so, daß dieses auf die gleiche Art produziert wird. Auch hier ist in den meisten Fällen ein Nasal beteiligt. In den beiden ersten Beispielen wird jeweils ein stimmhafter Plosiv in seiner Artikulationsart an einen vorausgehenden Nasal angeglichen. Im dritten Beispiel tritt die progressive Assimilation der Artikulationsart bei einem Frikativ und darauffolgendem Plosiv auf. Durch Angleichung des Plosivs entsteht ein Frikativ.

3.1.5 Assimilation der Stimmhaftigkeit

ausgewählte Regeln:	*beispielsweise in:*
`1pz>ps`	[absonderlich]
`1t#d>&d`	[hat der]
`1pz>bz`	[absonderlich]
`1t#d>&t`	[hat der]

Bei der Assimilation der Stimmhaftigkeit wird ein eigentlich stimmhaftes Segment unter dem Einfluß eines benachbarten stimmlosen Segments ebenfalls stimmlos. Umgekehrt kann auch ein stimmloses Segment stimmhaft werden.

Im ersten Beispiel trifft ein stimmloser Plosiv auf einen stimmhaften Frikativ. Der nachfolgende Frikativ wird unter seinem Einfluß ebenfalls stimmlos produziert. Dabei kann der Frikativ artikulatorisch betrachtet auch lediglich entstimmt werden und muß in seiner Intensität nicht einem "normalen" stimmlosen Frikativ gleichen.

Im zweiten Beispiel trifft ein stimmloser alveolarer Plosiv auf einen stimmhaften alveolaren Plosiv. Es entsteht ein einzelner stimmhafter Plosiv in intervokalischer Position. In abgestuften Reduktionsprozessen kann dabei zunächst die Lösung des stimmlosen Plosivs entfallen, so daß eine einzige Verschlußphase entsteht. Diese wird verkürzt und schließlich im vokalischen Kontext stimmhaft.

Es ist für beide Beispiele jedoch auch denkbar, daß das jeweils andere Segment den Einfluß auf das angrenzende ausübt. Wodurch bestimmt wird, welche Form auftritt, welche häufiger ist und welche Segmente betroffen sind, kann nur durch größere empirische Untersuchungen festgestellt werden.

Im Regelsystem ist daher generell vorgesehen, daß sich Konsonanten im Parameter der Stimmhaftigkeit an die Umgebung anpassen. Für jeden Konsonanten ist eine alternative Form mit dem jeweiligen stimmhaften bzw. stimmlosen Pendant generierbar.

3.2 Elisionen

Unter einer Elision versteht man die Auslassung von Segmenten, die bei der Prokuktion einer kanonischen Form zu erwarten sind. Es gibt typische Elisionen, die häufig auftreten und eine gewisse Regelmäßigkeit aufweisen. Dazu gehört beispielsweise die Elision von /t/ oder die Tilgung von /@/ in Endsilben. Weniger häufig zu beobachten und seltener beschrieben ist der im untersuchten Material gelegentlich beobachtete Ausfall von hinteren Frikativen.

Im folgenden werden Elisionen der genannten Segmente beschrieben.

3.2.1 Elisionen von /t/

ausgewählte Regeln:	*beispielsweise in:*
`1nt>n`	[freundschaftlich]
`1ftl>fl`	[freundschaftlich]
`1st>s`	[höchstens]

Die Auslassung des apikalen Plosivs ist verhältnismäßig oft und in allen möglichen Kontexten zu beobachten. Am häufigsten tritt die /t/-Elision auf, wenn dieser Plosiv laut kanonischer Form zwischen anderen Konsonanten steht. Das erste Beispiel zeigt einen typischen Fall für die Elision von Plosiven: Nach einem homorganen Nasal und vor einem Frikativ ist die Elision eines Plosivs regelmäßig. Außerdem ist im Wort des ersten Beispiels die Elision des Plosivs zwischen Frikativ und Lateral zu beobachten. Die dritte Regel ist häufig anwendbar: In der Verbindung [st] fällt [t] oft aus. Weitere Beispielwörter für diesen Fall sind: kannst, sonst, möglichst, wenigstens, spätestens.

3.2.2 Elisionen von /@/

ausgewählte Regeln:	*beispielsweise in:*
`1b@n#>bn#`	[haben]
`1g@l#>gl#`	[Igel]
`1m@n#>mn#`	[schwimmen]
`1m@n#>m#`	[schwimmen]

Von der Elision des Reduktionsvokals `/@/` sind in der Regel Endsilben wie -en, -el und -em betroffen. Wie schon erwähnt, kann die Schwa-Elision zu vielfältigen Assimilationen führen, wenn dadurch bestimmte Konsonantenverbindungen entstehen. Im ersten Beispiel entfällt `/@/` in der Infinitivendung *-en*. Dadurch treffen der bilabiale Plosiv `/b/` und der alveolare Nasal `/n/` aufeinander. Jetzt kann es zu einer Assimilation des Artikulationsortes kommen: `[bn]->[bm]`, darauf zu einer Assimilation der Artikulationsart: `[bm]->[mm]` und schließlich zur Reduktion: `[mm]->[m]`.

Im zweiten Beispiel fällt `/@/` vor einem `/l/` aus. Dies ist nun alleiniger Bestandteil einer Silbe und wird entsprechend silbisch realisiert.

Im letzten Beispiel tritt der Ausfall des Reduktionsvokals in zwei aufeinanderfolgenden unbetonten Silben auf. Auch hier kann es dann zu Assimilationen kommen:

`[hEr'aUsra:gNdn]->[hEr'aUsra:NNnn]->[hEr'aUsra:Nn]->[hEr'aUsra:N]`

Wie weit die Reduktionen gehen, hängt natürlich wieder stark von Sprechstil und -geschwindigkeit ab. So sind die letzten Formen vermutlich nur in stark verschliffener Sprache zu finden, in diesem Fall wahrscheinlich aber ebenso regelmäßig wie häufig anzutreffen.

3.2.3 Elisionen von /C/, /x/ und /h/

ausgewählte Regeln:	*beispielsweise in:*
`1xm>m`	[Nachmittag]
`1Cs>s`	[nächste]
`1x#h>&x`	[nach Hamburg]

Im Datenmaterial wurde die Elision der hinteren Frikative /C/,/x/ und /h/ beobachtet. Es wurden daher Regeln aufgenommen, die solche Elisionen in vielfältigen Kombinationen mit anderen Lauten sowohl innerhalb von Wörtern, als auch an Wortgrenzen vorsehen.

Auch dieser Reduktionsprozeß ist kontinuierlich, wobei die totale Elision eine Endstufe bei der Abschwächung des Frikativs bildet.

3.3 Substitution von Plosiven durch den Glottalverschluß

ausgewählte Regeln:	*beispielsweise in:*
`1tm>Qm`	[mitmachen]
`1k#m>Q#m`	[guck mal]
`1t@n#>Qn#`	[retten]

Bei den Segmentationsarbeiten ist immer wieder aufgefallen, daß von den Sprechern die stimmlosen Plosive `/p/, /t/, /k/` vermutlich glottalisiert wurden, d. h. mit geschlossener Glottis produziert bzw. durch einen Glottalverschluß ersetzt wurden. Dieses Phänomen tritt in Phonemverbindungen nach einem Vokal auf, wobei ein stimmloser Plosiv und mindestens ein weiterer Konsonant folgen. Es wurden nicht alle Kontexte und alle möglichen Ersetzungen oder Doppelartikulationen im Material gefunden. In den Regeln sind jedoch alle Möglichkeiten vorgesehen. Dieser Prozeß ist äußerst interessant, zumal er beispielsweise im Englischen (aber auch im Schwedischen und Dänischen, sowie amerikanischen Englisch) sehr häufig zu beobachten ist,

wo der Glottalverschluß von Sprechern als Verstärkung des Plosivs (glottal reinforcement) und als Ersatz von Plosiven (glottal replacement) eingesetzt werden kann.

3.4 Vokalisierung von /l/

ausgewählte Regeln:	*beispielsweise in:*
`1lg>@g`	[allgemein]
`1lz>@z`	[also]
`1lS#>@S#`	[falsch]

Anstelle der Artikulation eines alveolaren Laterals ist in zusammenhängender gesprochener Sprache häufig eine Vokalisierung zu beobachten. Der entstehende Vokal hat je nach Kontext eine andere Qualität, ist aber generell ein relativ ungespannter zentraler Laut und wird daher in den Regeln grundsätzlich durch `/@/` beschrieben.

3.5 Sproßkonsonant

ausgewählte Regeln:	*beispielsweise in:*
`1mC>mpC`	[Lämmchen]
`1mpC>mC`	[Lämpchen]
`1nf#>ntf#`	[fünf]

Bei der Artikulation von Konsonantverbindungen, die aus Nasal oder Lateral und nachfolgendem Frikativ oder Plosiv bestehen, kann ein sogenannter Sproßkonsonant entstehen. Es ist der mit dem Nasal homorgane Plosiv. Desgleichen kann dieser Plosiv ausfallen, wenn das Artikulationsziel von vorneherein die Kombination Nasal/Lateral mit darauffolgendem homorganen Plosiv und Frikativ/Plosiv ist.

Der Sproßkonsonant ist die Folge von unterschiedlicher Koordination der beteiligten Artikulationsorgane. So kann es sein, daß Wörter wie ganz und Gans phonetisch nicht unterscheidbar sind, da in ganz - `/gants/` die Affrikate `/ts/` regelmäßig auch als `[s]` realisiert wird und somit zur Ausspracheform `[gans]` wird, die mit der Ausspracheform des Wortes Gans übereinstimmt. In Gans dagegen kann zwischen `/n/` und `/s/` regelmäßig der Sproßkonsonant `[t]` entstehen, was zur Ausspracheform `[gants]` führt und mit der kanonischen Form von *ganz* zusammenfällt.[10]

3.6 Lenisierung

ausgewählte Regeln:	*beispielsweise in:*
`1t@n#>dn#`	[guten Tag]
`1Cs>js`	[nächste]
`1fp>vp`	[aufpassen]

Unter Lenisierung versteht man die Veränderung einzelner oder mehrerer Segmente, die aufgrund höherer Sprechgeschwindigkeit und weniger sorgfältiger Artikulation und abgeschwächter artikulatorischer Energie zustande kommt. Stimmlose Frikative können stimmhaft werden und in einer weiteren Reduktionsstufe zu friktionslosen Approximanten. Fortis-Plosive können unaspiriert realisiert werden und zu Lenis-Plosiven und -Frikativen werden.

Kohler[11] beschreibt die Artikulationsvorgänge dieser Prozesse und nennt als eine Ursache unter anderem, daß "Sprecher ebenfalls ihren artikulatorischen Aufwand in den einzelnen Bewe-

10. Genauere Beschreibung der artikulatorischen Vorgänge sind in [7] nachzulesen.
11. [3], pp. 29 - 31

gungen, d. h. also Muskelkraft und Bewegungsenergie, reduzieren. Die Schließbewegungen werden wiederum abgekürzt und die Verschlüsse folglich geschwächt."

Da es, wie schon erwähnt, in SAMPA keine Möglichkeit gibt, stimmlose Fortis- und Lenis-Konsonanten zu unterscheiden, wird in den Regeln für Lenis-Konsonanten das Symbol für den entsprechenden stimmhaften Laut gewählt. Das ist nicht nur deswegen gerechtfertigt, weil das "Stimmhaftwerden" ohnehin eine Stufe im Prozeß der Lenisierung darstellt, sondern auch, weil die in das Modelltraining eingegangenen Daten unter dem Label der stimmhaften Konsonanten unter anderem auch die Lenis-Entsprechung eines Lautes enthalten.

3.7 /r/-Realisationen

Das /r/-Phonem kann in sehr unterschiedlicher Weise realisiert werden, besonders in Abhängigkeit von Dialekt und Stil. Das gerollte Zungenspitzen-r ist inzwischen weitgehend als dialektal zu werten und ist beispielsweise in Bayern/Österreich, Schleswig-Holstein und im alemannischen Sprachraum zu finden.

Gebräuchlicher sind der uvulare Frikativ [ ], [X] und seine friktionslosen Varianten. In bestimmten Kontexten wird das /r/ als zentraler ungespannter Vokal realisiert. Im folgenden wird von der vokalischen und konsonantischen Form des /r/-Phonems die Rede sein, denn diese zwei Realisationsmöglichkeiten sind in der SAMPA-Notation beschreibbar: Unter dem Symbol /r/ sind alle konsonantischen Formen, also [ ], [X], [r] (auch als Frikativ oder Flap) und [ ]zusammengefaßt. Mit dem Symbol /6/ wird die vokalische Form beschrieben, wobei bei einer Verschmelzung von /6/ mit /a:/ oder /a/ nach einer Zwischenstufe [a:6], [a6] schließlich ein [a:] und in Verbindungen mit anderen Vokalen eine Art Diphthong entsteht.

Die Verteilung der /r/-Allophone der kanonischen Ausspracheform ist weitgehend komplementär. Es gibt jedoch Kontexte in denen die beiden Realisierungen in freier Variation auftreten.

Verteilung der /r/-Allophone

A konsonantische Realisierung des /r/:

silbeninitial vor Vokalen	*intervokalisch*	*nach Konsonanten*
Reise [r'aIz@]	Ural [Qu:r'a:l]	drei [draI]
Ruhe [ru:@]	fahren [f'a:r@n]	fröhlich [fr'2lIC]
hilfreich [h'IlfraIC]	Nieren [ni:r@n]	auftreiben [Q'aUftraIb@n]

B vokalische Realisierung des /r/:

im Silbenauslaut nach gespanntem Vokal, wobei Konsonanten folgen können	*in der Silbe <-er>, auf die Konsonanten folgen können*	*in Vorsilben wie <ver->, <zer->, <er->*
Tier [ti:6]	Sprecher [Spr'EC6]	verlieren [fE6li:r@n]
fuhrst [fu:6st]	abermals [Q'a:b6ma:ls]	zerlassen [tsE6las@n]
fahrt [f'a:t]	folgerst [f'Olg6st]	erleben [QE6l'e:b@n]

C sowohl vokalische als auch konsonantische Realisierung des /r/:

im Silbenauslaut nach ungespanntem Vokal, wobei Konsonanten folgen können
Irrtum [Q'Irtu:m], [Q'I6tu:m]

im Silbenauslaut nach ungespanntem Vokal, wobei Konsonanten folgen können
Karte [kart@], [ka:t@]
Hirsch [h'IrS], [h'I6S]

Die Verteilung der /r/-Allophone wurde für das zu entwerfende Regelsystem in entsprechende Regeln gefaßt, die sich auf ein abstraktes /r/-Symbol beziehen. Die obige Beschreibung der Verteilung der /r/-Realisationen enthält Angaben wie silbeninitial und im Silbenauslaut, die in den kanonischen Formen nicht enthalten sind. Daher ist in den kanonischen Formen in jedem Kontext für das /r/-Phonem das Symbol `r` notiert. Da dies der Notation für die konsonantische Realisation entspricht, wurden Regeln formuliert, die davon die vokalischen Formen ableiten. Dadurch entstehen höchst unwahrscheinliche Formen, in denen eigentlich nur ein vokalisches /r/ auftritt. Da es keine geeigneten Angaben über Silbengrenzen und Kompositagrenzen gibt, müssen alle möglichen Kontexte berücksichtigt werden.

4 Regelanwendung

Die kanonischen Formen der Wörter einer Äußerung werden für eine effektive Weiterverarbeitung bei der Prozedur des Viterbi-Alignments in Form eines einfachen, gerichteten Graphen verkettet abgespeichert, der durch Anwendung der Regeln erweitert wird.[12] Die hypothetischen Aussprachevarianten der gesamten betreffenden Äußerung sind in dem entstehenden Variantengraphen repräsentiert.

In dem Zusammenhang soll nachdrücklich nochmals darauf aufmerksam gemacht werden, daß auch wortübergreifende Reduktionen mitverarbeitet werden können.

Die einzelnen Laute werden jeweils als ein Knoten des Graphen, der außerdem einen Anfangs- und einen Endknoten besitzt, gespeichert. Die Pfeile zeigen die erlaubten Übergänge von einem Knoten zum nächsten an. Sowohl Knoten als auch Übergänge werden mit Ziffern bezeichnet, wobei jeder Knoten seine eigene Nummer erhält, ein Übergang jedoch die Nummer derjenigen Regel trägt, durch die er entstanden ist. Übergänge aus dem einfache Graphen der kanonischen Form tragen die Ziffer “0”.

Der Variantengraphen für die kanonische Form des Einzelwortes `/ h a: b @ n /` ist in Abbildung 1 dargestellt. Er besteht aus acht Knoten, von denen einer durch die Anwendung einer Regel entstanden ist. Es konnten drei verschiedene Regeln auf die kanonische Form angewendet werden, wobei vier neue Übergänge entstanden sind.

Die drei Regeln lauten:

1. `1b@n>bn`
2. `1b@n>bm`
3. `1b@n>m`

12. zum Algorithmus zur Regelanwendung auf den kanonischen Graphen siehe [1].

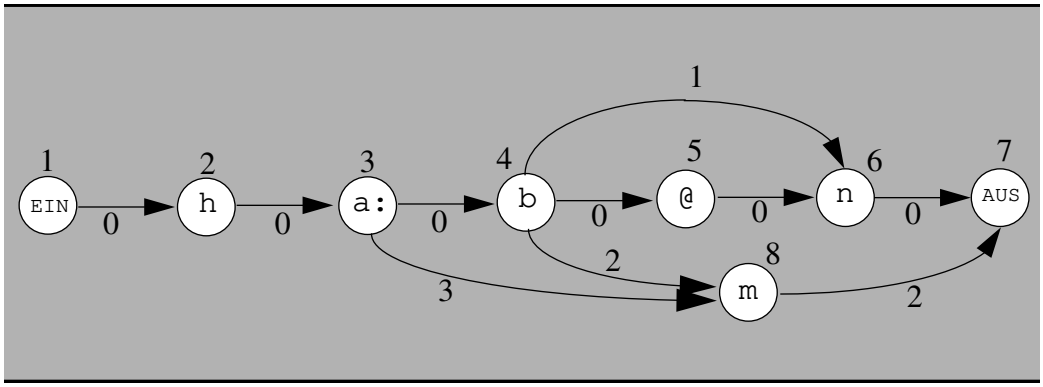

Abbildung 7: Variantengraph für das Einzelwort / h a: b @ n /

Von dem Graphen lassen sich folgende Aussprachevarianten des Wortes / h a: b @ n / ableiten:

1. / h a: b @ n /
2. / h a: b n /
3. / h a: b m /
4. / h a: m /

Ein Ausschnitt aus dem Variantengraph für eine Äußerung, die aus mehreren Wörtern besteht, ist in Abbildung 2 zu sehen: ... / n # z i: # Q aI n @ n # m / ... Hier sind Reduktionen an einem Wortübergang abgebildet. Drei Regeln konnten angewendet werden:

1. 1n@n#>n#
2. 1n#m>&m
3. 1#QaIn@n#>#n@n#

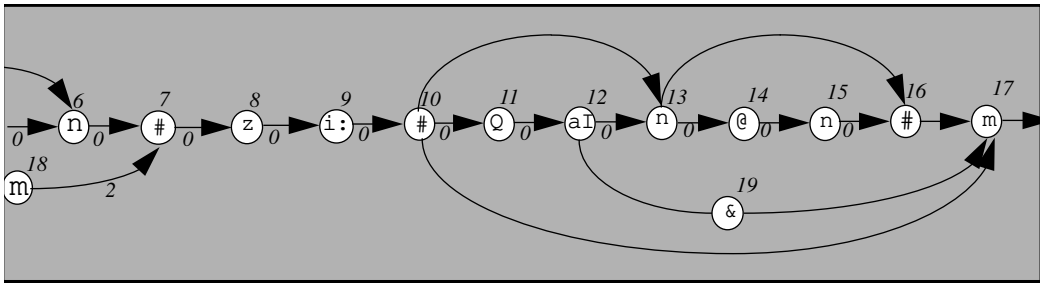

Abbildung 8: Ausschnitt aus Variantengraph: ... / n # z i: # Q aI n @ n # m / ...

Die folgenden Aussprachevarianten lassen sich ableiten:

1. .../ n # z i: # Q aI n @ n # m / ...
2. ... / n # z i: # Q aI n # m / ...
3. ... / n # z i: # Q aI n @ & m / ...
4. ... / n # z i: # Q aI n & m / ...
5. ... / n # z i: # Q aI m / ...
6. ... / n # z i: # n @ n # m / ...
7. ... / n # z i: # n # m / ...
8. ... / n # z i: # m / ...

5 Zusammenfassung

Im vorliegenden Dokument wurde ein Regelsystem zur Generierung von Aussprachevarianten der kanonischen Form einer Äußerung vorgestellt. Es ist Bestandteil des auf einem statistischen Verfahren basierenden automatischen Segmentationssystems, das im Teilprojekt 14.7 in München entwickelt wird. Die technische Beschreibung findet sich in einem gesonderten Dokument ([1]). Mit dem Regelsystem wird die Variabilität gesprochener Sprache modelliert, und es konnte eine erste Version des automatischen Segmentationssystems bereits verbessern.

Die Regeln basieren auf Beobachtungen von Reduktionserscheinungen von gesprochener Sprache, wie sie in der Münchener Phondat II - Prologdatenbank zu finden sind. Sie wurden systematisch um Regeln erweitert, so daß dieser insgesamt unterspezifizierende Regelkorpus zu einer großen Anzahl von hypothetischen Aussprachevarianten führt.

Die kanonische Form einer Äußerung und ihre Varianten werden durch einen gerichteten Graphen repräsentiert. Im Verlauf eines Viterbi-Alignments wird nach dem optimalen Pfad durch den Variantengraphen gesucht und die im Graphen repräsentierten Lautmodelle dem zu segmentierenden Signalabschnit zugeordnet

Ergebnis ist im Optimalfall die automatische Segmentation eines Sprachsignals unter Verwendung der Symbole der zutreffenden Aussprachevariante und eine richtige Zuweisung von Segmentgrenzen.

6 Literatur

[1] Kipp, Andreas (1995): *Automatisches Segmentations- und Etikettierungssystem*, Verbmobil-Memo 95.

[2] Kohler, Klaus (1977): *Einführung in die Phonetik des Deutschen*. Berlin: E. Schmidt.

[3] Kohler, Klaus (1979): "Kommunikative Aspekte satzphonetischer Prozesse im Deutschen." in: Vater, Heinz (Hrsg.): *Phonologische Probleme des Deutschen.* Tübingen: Narr.

[4] Lindblom, Björn (1990): "Explaining phonetic variation: A sketch of the H and H theory." In: W. J. Hardcastle and A. Marchal (eds.): *Speech Production and Speech Modelling*, pp. 403-439. Kluwer: Dordrecht.

[5] Kühnert, Barbara (1993): "Some Kinematic Aspects of Alveolar-Velar Assimilations." In: Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 31, 263-272.

[6] Pompino-Marschall, B. (Hrsg.)(1992): "PHONDAT. Verbundvorhaben zum Aufbau einer Sprachsignaldatenbank für gesprochenes Deutsch." In: Forschungsbericht des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 30, 99 - 128.

[7] Wesenick, M.-B. (1994): *Entwurf eines Regelsystems der Aussprache des Deutschen als Basis für empirische Untersuchungen.* Magisterarbeit, Institut für Phonetik und Sprachliche Kommunikation München.

[8] Wesenick, M.-B., F. Schiel (1994): "Applying Speech Verification to a Large Data Base of German to obtain a Statistical Survey about Rules of Pronunciation", Proceedings of ICSLP 1994, Yokohama, pp. 279 - 282.