

Regelbasierte Modellierung fremdsprachlich akzentbehafteter Aussprachevarianten

Grundlagen, Entwurf und Implementierung eines Regelsystems für
sprachtechnologische Anwendungen

Dissertation
zum Erwerb des akademischen Grades Dr. phil.

vorgelegt von

Stefan Schaden, geb. in Essen

im Fach Kommunikationswissenschaft
Fachbereich Geisteswissenschaften
Universität Duisburg-Essen

Gutachter der Arbeit:
Prof. Dr. Achim Eschbach, Prof. Dr. Ute Jekosch

Datum der Disputation: 16. Juni 2006

Vorwort

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit am Institut für Kommunikationsakustik (IKA) der Ruhr-Universität Bochum von 1999 bis 2004. Wesentliche Teile der Arbeit wurden von der Deutschen Forschungsgemeinschaft (DFG) im Rahmen des Projektes *Typisierung und regelbasierte Modellierung akzentgefärbter Aussprachevarianten für die maschinelle Sprachverarbeitung* (BL189/27-1) finanziell gefördert.

Mein Dank gilt allen ehemaligen Kollegen am IKA, insbesondere Prof. Ute Jekosch (heute TU Dresden) für die Übernahme der Betreuung der Arbeit, die Unterstützung bei der Realisierung des damit verbundenen Forschungsprojektes und fachliche Diskussionen, die dazu beitrugen, die Zielrichtung dieser Arbeit klarer hervortreten zu lassen. Danken möchte ich auch Prof. Jens Blauert für die Möglichkeit, diese Arbeit am IKA durchzuführen. Unter den Kollegen am IKA möchte ich mich besonders bei Dr.-Ing. Sebastian Möller, Dr.-Ing. Alexander Raake, Dr.-Ing. Jörg Buchholz und Dipl.-Ing. Jan Krebber für die angenehme Arbeitsatmosphäre und die Unterstützung in vielen kleinen und größeren Dingen bedanken. Herrn Dr.-Ing. Hans-Wilhelm Rühl (Wetzlar) gebührt der Dank für den ersten Anstoß zur Behandlung des Themas dieser Arbeit sowie auch für ihre Förderung in der Anfangsphase.

Prof. Achim Eschbach (Universität Duisburg-Essen) gilt mein ganz besonderer Dank nicht nur für die Übernahme des Referates, sondern vor allem auch für die Ermutigung zur Aufnahme eines Dissertationsprojektes und die freundschaftliche Unterstützung in den vergangenen Jahren.

Keinesfalls unerwähnt lassen möchte ich den Einsatz der studentischen Mitarbeiterinnen und Mitarbeiter, ohne deren Hilfe die umfangreiche Datengrundlage der Arbeit nicht zustande gekommen wäre: Andreea Niculescu für die Durchführung von Sprachaufnahmen und die Betreuung von „Versuchspersonen“ sowie Anders Krosch für die unermüdliche Anwerbung von Sprechern, die Realisierung von Sprachaufnahmen in Bochum und Udine sowie die Bearbeitung der Audio-daten. Mein besonderer Dank gilt zudem Katja Keßelmeier und Anneli von Könnemann, die mit Sachkenntnis (und Geduld!) die phonetische Transkription der Sprachdaten übernommen haben und sich dabei trotz zum Teil unlösbarer phonetischer Detailfragen nicht entmutigen ließen. Schließlich geht mein Dank auch an Nino Peterek und Alena Böhmová (Prag) für ihre freundliche Hilfe bei der Transkription tschechischer Namen.

Und obwohl sie von den fachlichen Details dieser Arbeit (weitgehend) verschont geblieben sind, haben Dr. Thorsten Kieren und Melanie Elsner dennoch einen ebenso großen Anteil an ihrer Fertigstellung. Danke auch an Euch.

Dolcédo, im Mai 2006

INHALT

1	Einleitung	1
1.1	Ausgangspunkt: Variabilität gesprochener Sprache.....	1
1.2	Ziel dieser Arbeit	3
1.3	Eingrenzung des Untersuchungsgebietes	5
1.4	Übersicht der Kapitel.....	6
I.	HINTERGRUND UND PROBLEMSKIZZE	9
2	Fremdsprachlicher Akzent: Sprachwissenschaftliche Perspektiven	10
2.1	„Akzent“: Eine erste Begriffsbestimmung.....	11
2.1.1	Akzente aus der Perspektive des Sprechers.....	11
2.1.2	Akzente aus der Perspektive des Hörers.....	12
2.2	Sprachvergleich und Fehlerprognosen.....	13
2.2.1	Kontrast, Transfer und Interferenz.....	13
2.2.2	Interferenz und Fehler im Fremdsprachenerwerb	17
2.2.3	Beispiel einer kontrastiven Analyse	18
2.2.4	Kritik der Kontrastivhypothese	21
2.3	Interlanguage und approximative Systeme	25
2.3.1	Lernersprachen als eigenständige Sprachsysteme.....	25
2.3.2	Approximationen und Übergangsstufen	26
2.4	Perzeptive Assimilation und Äquivalenzklassifizierung.....	28
2.4.1	Einfluß der Perzeption auf die Produktion.....	28
2.4.2	Äquivalenzklassifizierung und Stabilisierung von Lautkategorien.....	30
2.5	Überwindung von Akzenten und „Fossilisierung“.....	34
2.6	Weitere Einflüsse	37
2.7	Implikationen für die vorliegende Arbeit	39
2.7.1	Systematizität der L2-Produktion.....	39
2.7.2	Direkte Implikationen für die Modellbildung	41

3	Varietäten und Akzente in der Sprachtechnologie	43
3.1	Automatische Spracherkennung.....	44
3.1.1	Grundprobleme	44
3.1.2	Adaptionstechniken.....	47
3.1.3	Spracherkennung und fremdsprachliche Akzente.....	49
3.2	Generierung von Varietäten in der Sprachsynthese	53
3.3	Exkurs: Eigennamen als sprachtechnologisches Problem.....	56
3.3.1	Linguistische Besonderheiten von Eigennamen.....	56
3.3.2	Sprachtechnologische Studien zu Eigennamen.....	60
3.4	Varietäten und Akzente in der Sprachtechnologie: Zwischenbilanz.....	62
4	Vorüberlegungen zu einer regelbasierten Modellierung	65
4.1	Grundproblem: Die Bestimmung von Zielformen.....	65
4.2	Das Verfahren der lexikalischen Adaption	66
4.3	Interindividuelle Variabilität als Modellierungsproblem.....	70
4.3.1	Die Situation bei Dialekten.....	70
4.3.2	Die Situation bei fremdsprachlichen Akzenten.....	73
4.3.3	Finite-State-Modelle	73
4.4	Das Modell prototypischer Akzentstufen	78
II.	DATENERHEBUNG UND ANALYSE	83
5	Erstellung einer domänenspezifischen Sprachdatenbank	84
5.1	Zur Rolle von Sprachdatensammlungen in Linguistik und Sprachtechnologie	84
5.2	Beschreibung der Datenbank	87
5.2.1	Sprachen, Sprachenpaare und Sprachrichtungen.....	87
5.2.2	Vorbemerkung zur Selektion von Sprechern.....	90
5.2.3	Sprecher	92
5.2.4	Sprachmaterial	93
5.2.5	Versuchsdesign und -durchführung	96
5.2.6	Teilaufgaben/Szenarien.....	98
5.3	Lexikon und phonetische Transkription	100

5.3.1	Kanonische Transkription.....	100
5.3.2	Transkription der Sprachdaten.....	101
5.4	Von sprachlichen Daten zu Regeln.....	103
5.5	Sprachdatensammlung: Grenzen, Probleme, Ausblick.....	105
6	Fehlertypen und Implikationen für eine regelbasierte Modellierung	108
6.1	Grundlage und Motivation der Fehlertypologie.....	108
6.2	Substitutionen, Einfügungen, Elisionen	109
6.2.1	Vorbemerkung: Die segmentale Betrachtungsweise	109
6.2.2	Typen des Lautersatzes	110
6.2.3	Alternationen der Vokalqualität.....	112
6.2.4	Alternation allophonischer Realisierungen	114
6.2.5	Elisionen	116
6.2.6	Einfügungen.....	117
6.2.7	Kontextfaktoren bei Lautsubstitutionen.....	118
6.3	Die Rolle phonologischer Regeln aus L1 oder L2.....	119
6.3.1	Sprachspezifische phonologische Regeln	119
6.3.2	Transfer phonologischer Regeln aus L1 auf L2	123
6.3.3	Nichtanwendung phonologischer Regeln aus L2.....	126
6.4	Einflüsse der Schriftaussprache	129
6.4.1	Transfer der Graphem-Phonem-Beziehungen	129
6.4.2	Interferenz der Graphem-Inventare und Diakritika	132
6.4.3	Einflüsse der Schriftaussprache: Vorläufiges Fazit.....	136
6.5	Einflüsse der Morphologie	137
6.5.1	„Dual Route“-Strategien bei der Aussprache von Eigennamen	137
6.5.2	Fehlerhafte Identifizierung von lexikalischen Morphemen	141
6.6	Kumulative Fehler und Grenzen der Typologie.....	142
III.	PROGRAMMTECHNISCHE UMSETZUNG UND EVALUIERUNG	146
7	Regelbasierte Modellierung von Aussprachefehlern: Methoden, Regeltypen, Beispiele	147

7.1	Allgemeiner Aufbau des Regelsystems.....	148
7.1.1	Modell I: Erweiterter Graphem-nach-Phonem- Umsetzer	148
7.1.2	Modell II: Postlexikalische Modifikation.....	151
7.2	Komponenten des Regelsystems.....	153
7.2.1	Regeln und Regelinterpreter-Programm.....	153
7.2.2	Phonem-Graphem-Zuordnungen	155
7.2.3	Akzentstufenhierarchie und inkrementelle Regelanwendung	158
7.3	Regeltypen.....	161
7.3.1	Kontextfreie phonetische Substitutionsregeln.....	161
7.3.2	Kontextbedingungen (positionsbedingte Substitutionen).....	163
7.3.3	Formale Darstellung von Kontextspezifizierungen	165
7.3.4	Graphemabhängige Lautsubstitutionen.....	167
7.4	Beispielanwendungen der Regeln	169
7.4.1	Rekonstruktion „stummer Phoneme“	169
7.4.2	Rekonstruktion von Reduktionsvokalen	171
7.4.3	Graphemische Kontexte.....	172
7.4.4	Sprachspezifische Grapheme	174
7.5	Regeltypen: Vorläufiges Fazit	176
8	Evaluierung der Regeln	179
8.1	Kriterien einer Evaluierung.....	179
8.1.1	Systembezogene oder linguistische Evaluierung?	179
8.1.2	„Fehler“ bei einer Modellierung von Fehlern	181
8.2	Beschreibung des Evaluierungsverfahrens.....	183
8.2.1	Ermittlung der besten Approximation	183
8.2.2	Definition eines phonetischen Distanzmaßes: Edit- Distanz	184
8.2.3	Phonetische Gewichtung durch Merkmalssysteme.....	188
8.3	Implementierung des Verfahrens	192
8.4	Anwendung des Verfahrens.....	196
8.5	Ergebnisse der Evaluierung.....	200
8.5.1	Sprechergruppen und Sprachmaterial	200
8.5.2	Ergebnisse Sprechergruppe B	201

8.5.3	Ergebnisse bei Ortsnamen, isolierte Wörter	202
8.5.4	Ergebnisse bei neuem Vokabular.....	203
8.5.5	Ergebnisse Sprechergruppe C.....	205
8.6	Möglichkeiten einer weiterführenden Evaluierung	206
9	Zusammenfassung: Möglichkeiten und Grenzen einer Modellierung von Akzenten	208
9.1	Perspektiven einer Weiterentwicklung.....	208
9.2	Grenzen der Untersuchung	209
10	Literatur	213
11	Notationskonventionen	228

1 Einleitung

Jedenfalls aber gehört es nicht zu dem Allersebstverständlichsten, daß ein menschlicher Stimmapparat als Sender einige Zehntausende kurzer Lautbilder so wohlgeprägt erzeugen kann, daß jedes der Tausende von einem menschlichen Hörapparat als Empfänger mühelos als das und das erfaßt und von allen anderen unterschieden wird. (Bühler 1934: 277)

1.1 Ausgangspunkt: Variabilität gesprochener Sprache

Oft sind es vertraute und scheinbar selbstverständliche Eigenschaften menschlicher Sprache, die sich bei genauer Betrachtung als erstaunlich erweisen und die sich einer einfachen – alltagsweltlichen oder wissenschaftlichen – Erklärung entziehen. So gehört es zu den in unserer alltäglichen Erfahrung mit Sprache kaum beachteten, aber durchaus bemerkenswerten Fähigkeiten eines jeden Sprechers und Hörers, lautsprachliche Äußerungen auch dann noch weitgehend zuverlässig identifizieren zu können, wenn der Zeichenträger – d.h. das akustische Sprachsignal – nur in lückenhafter, verzerrter oder anderweitig defizienter, von der Norm abweichender Form vorliegt. Trotz einer erheblichen Bandbreite akustischer und artikulatorischer Variabilität weist das lautsprachliche Zeichen einen scheinbar unveränderlichen, stabilen Kern auf, der es uns ermöglicht, sprachliche Äußerungen selbst dann verlässlich zu erkennen, wenn der Zeichenträger (z.B. aufgrund bestimmter Eigenschaften der Quelle oder des Kanals) von seiner normgerechten Realisierung stark abweicht.

Variabilität als generelle Eigenschaft menschlicher Sprache wird bei der Untersuchung ihrer phonetischen Realisierungen in besonderem Maße manifest: Aus akustisch-physikalischer Perspektive existieren keine zwei sprachlichen Äußerungen, die exakt identisch sind, auch dann nicht, wenn sie von ein und demselben Sprecher unter identischen Bedingungen produziert wurden. Bird (2002: 3) gibt ein prägnantes Beispiel: Läßt man das Wort *zehn* von zehn verschiedenen Sprechern je zehn Mal sprechen, so erhält man bereits 100 voneinander verschiedene phonetische Realisierungen des Wortes und seiner einzelnen Laute. Was sich dem Phonetiker jedoch als ein extremes Variationsspektrum darbietet, ist im alltäglichen Sprachgebrauch keinesfalls ein Hindernis: So ist ein Hörer in der Regel ohne Schwierigkeiten imstande, jede dieser 100 Varianten als eine Realisierung von *zehn* zu erkennen.

Für wissenschaftliche Untersuchungen menschlicher Lautsprache bedeutet dies, daß man sich einerseits mit einer stets präsenten, dem Gegenstand inhärenten Variabilität konfrontiert sieht, andererseits diese Variabilität jedoch nicht unbegrenzt ist: Sie darf ein bestimmtes Spektrum nicht überschreiten, da andernfalls die Unterscheidbarkeit und Erkennbarkeit des lautsprachlichen Zeichens nicht mehr gewährleistet ist. In diesem Spannungsfeld besteht eine zentrale Aufgabe wissenschaftlicher Untersuchungen letztlich darin, *Invarianten innerhalb der Variation* zu ermitteln. Auch die vorliegende Untersuchung verfolgt dieses Ziel.

Auf besondere Weise evident wird der Zusammenhang von Variabilität und Konstanz im Bereich der *Sprachtechnologie*, d.h. der automatischen Verarbeitung menschlicher Sprache: Variation in der konkreten Realisierung sprachlicher Äußerungen stellt für den Menschen in der Regel kein Hindernis im praktischen Sprachgebrauch dar. Der Versuch jedoch, dieses Vermögen *technisch* zu beschreiben und zu simulieren, läßt die Komplexität dieser menschlichen Fähigkeit und die damit verbundenen Teilprobleme deutlich hervortreten. So sind beispielsweise Programme zur automatischen Spracherkennung – also der maschinellen Umsetzung gesprochener Sprache in einen orthographischen Text, der das Gesagte wiedergibt – noch immer nicht in der Lage, „natürliche Sprache“ in all ihren Erscheinungsformen fehlerfrei zu „erkennen“. Trotz großer technischer Fortschritte auf diesem Gebiet innerhalb der letzten ca. zwei Jahrzehnte sind derartige Programme in ihrer Erkennungsleistung dem Menschen noch immer deutlich unterlegen. Es ist eine weiterhin offene Frage, ob und mit welchen Mitteln dieses Defizit technisch überwunden werden kann.

Generell läßt sich aber die These vertreten, daß ein großer Teil der noch ungelösten Probleme letztlich auf die der Sprache inhärenten Variabilität zurückgeht: So nimmt die Leistungsfähigkeit sprachtechnologischer Systeme immer dann ab, wenn das Sprachverhalten des Benutzers signifikant von der erwarteten Sprachnorm abweicht. Zwar bestehen stets gewisse Toleranzgrenzen; ist jedoch die Diskrepanz zwischen den vom System antizipierten Äußerungen und den tatsächlichen Äußerungen zu groß, werden Fehler wahrscheinlich. Möglicherweise ist das Resultat sogar ein vollständiges Scheitern des sprachlichen „Dialogs mit Maschinen“.

Ein wesentlicher Schritt zur Überwindung dieses Problems besteht darin, zu ermitteln, ob die beobachteten Abweichungen eine innere Systematizität aufweisen. Ist dies der Fall, kann das Wissen über diese Systematizität in die Gestaltung des sprachverarbeitenden Systems einfließen. Ist dies nicht der Fall – handelt es sich also weitgehend um idiosynkratische, unvorhersehbare Abweichungen –, so ist eine zuverlässige Lösung deutlich erschwert.

1.2 Ziel dieser Arbeit

Sprachliche Variabilität manifestiert sich in vielfältigen Formen, von denen die meisten uns aus dem täglichen Umgang mit Sprache vertraut sind. Auf der Ebene der Phonetik bzw. der Lautung einer Sprache gehören *fremdsprachliche Akzente* zu den auffälligsten Erscheinungsformen von Variabilität.

Ein Akzent, so kann man vorläufig sagen, tritt auf, wenn ein Sprecher einer beliebigen Muttersprache (im folgenden auch als L1 bezeichnet), der über nur „unvollständige“ Kenntnisse bzw. Fertigkeiten der Phonetik einer Zielsprache (L2) verfügt, sprachliches Material der Zielsprache artikuliert. Diese nur partielle Kenntnis manifestiert sich u.a. in charakteristischen lautlichen Erscheinungen wie z.B. der Substitution von zielsprachlichen Lauten durch abweichende Laute, dem Transfer phonologischer Regelmäßigkeiten oder prosodischer Muster der Muttersprache auf die Zielsprache, oder - im Falle gelesener Sprache - einer Übertragung der muttersprachlichen Schriftaussprache (d.h. der Graphem-Phonem-Beziehungen) auf die Zielsprache.

Derartige Erscheinungen, die bisher vor allem als rein sprachwissenschaftliche oder auch fremdsprachendidaktische Probleme in Betracht kamen, sind in den vergangenen Jahren auch in den Fokus sprachtechnologischer Forschung gelangt. So wurde im vorausgehenden Abschnitt gesagt, daß Norm-Abweichungen, die ein Toleranzspektrum überschreiten, oftmals zu Leistungseinbußen sprachverarbeitender Systeme führen. Dies gilt auch und gerade für *phonetische* Abweichungen, die in der sprachlichen Eingabe solcher Systeme vorliegen. Bei nicht-muttersprachlichen Sprechern sind derartige Abweichungen oftmals besonders ausgeprägt. Eine Folge davon ist z.B., daß automatische Spracherkennungssysteme, deren Leistungsfähigkeit unter regulären Bedingungen zufriedenstellend ist, bei akzentbehafteter Aussprache oftmals nur noch eine stark reduzierte Erkennungsleistung aufweisen. Gängige Adaptions- und Optimierungsmethoden erweisen sich dabei als unzureichend zur Lösung dieses Problems.

Es wird daher nach neuen Methoden gesucht, um die bei Nicht-Muttersprachlern verschiedener Herkunft auftretenden Aussprachevarianten in einer Weise systematisch zu erfassen, zu beschreiben und zu modellieren, die eine Integration in technische Anwendungen erlaubt. Eine solche Methode wird in der vorliegenden Arbeit entwickelt und vorgestellt. Das übergeordnete Ziel besteht dabei darin, ausgewählte phonetische Eigenschaften fremdsprachlich akzentgefärbter Aussprachevarianten zu erfassen, zu typisieren und auf der Symbolebene zu modellieren. In einem nächsten Schritt wird dieses Modell in ein Computerprogramm umgesetzt, welches in der Lage ist, prototypische Aussprachevarianten, die in akzentgefärbter Sprache zu beobachten sind, mittels Regeln zu generieren.

„Modellierung auf der Symbolebene“ bedeutet hierbei, daß Deskription und Modellbildung auf der Basis symbolischer Transkriptionen lautsprachlicher Ereignisse erfolgen. Es werden im Rahmen dieser Arbeit keinerlei akustisch-phonetische Messungen oder Techniken der Signalverarbeitung herangezogen. Auch die programmtechnische Implementierung des vorgestellten Modells, welche Teil dieser Arbeit ist, erfolgt ausschließlich auf der Ebene phonetischer Transkriptionen und deren Manipulation mittels linguistischer Regeln.

Dieses Vorgehen ermöglicht es nicht nur, an etablierte linguistische Beschreibungsverfahren anzuknüpfen, sondern ist auch aus einer technischen Perspektive durchaus sinnvoll. So bilden symbolphonetische Codierungen (d.h. phonetische Transkriptionen der Aussprache) einen festen Bestandteil heutiger sprachverarbeitender Systeme: Sowohl in der automatischen Spracherkennung als auch in der Sprachsynthese bildet das sogenannte Lexikon – d.h. eine „Liste“ des vom System nutzbaren Vokabulars mitsamt einer Codierung seiner Aussprache – eine zentrale Systemkomponente. Da es bei beiden Systemtypen ähnlich aufgebaut und strukturiert ist, können Methoden der Adaption und Modellierung, die auf der Ebene des Lexikons operieren, für beide Technologien gleichermaßen genutzt werden. Bezogen auf das vorliegende Untersuchungsgebiet fremdsprachlicher Akzente bedeutet dies z.B., daß die im Rahmen dieser Arbeit vorgestellte Methode gleichermaßen eingesetzt werden kann (a) zur Adaption von automatischen Spracherkennungssystemen für nicht-muttersprachliche Nutzer, d.h. zur *Erkennung* von Sprache mit fremdsprachlichen Akzenten oder (b) zur Adaption von Sprachsynthesystemen zur *Simulation* fremdsprachlicher Akzente.

Es ist somit nicht das primäre Ziel dieser Arbeit, ein einzelnes bestehendes technisches System (z.B. einen automatischen Spracherkenner) so zu manipulieren, daß seine Leistungsfähigkeit in bezug auf eine spezifische Sprechergruppe optimiert wird. Eine solche ingenieurwissenschaftliche Umsetzung (Systementwicklung und -optimierung) kann und soll durchaus auf der Grundlage dieser Arbeit erfolgen, sie ist jedoch nicht selbst ihr Bestandteil. Vielmehr ist beabsichtigt, zunächst das hierfür notwendige Wissen in einer Weise zu systematisieren und zu generalisieren, die eine solche Systemoptimierung erst ermöglicht. Dies geschieht exemplarisch am Beispiel einer begrenzten Sprechergruppe; jedoch ist die vorgestellte Methode so konzipiert, daß sie grundsätzlich auf neue Sprechergruppen übertragbar ist. So wird die Untersuchung zwar anhand einer – notwendigerweise – begrenzten Anzahl von Sprechern, Sprachen und Sprachenkombinationen durchgeführt; die daraus hervorgehenden Methoden sind aber prinzipiell auf neue Sprachen und Sprecher übertragbar.

Auch wird bewußt offengehalten, für welchen Typ sprachverarbeitender Systeme (Spracherkennung oder -synthese) das Verfahren letztlich eingesetzt wird. Es ist beabsichtigt, die Ergebnisse so zu formulieren und zu formalisieren, daß

eine Integration in verschiedenartige Systeme leicht erreichbar ist. Einer solchen übergreifenden Anwendbarkeit des Ansatzes wurde stets Priorität gegenüber einer Optimierung eines spezifischen, existierenden Systems eingeräumt. Damit soll gewährleistet werden, daß der entwickelte Ansatz in seiner Anwendung nicht auf ein einzelnes System bzw. einen einzelnen Systemtyp beschränkt bleibt, sondern prinzipiell für verschiedene potentielle Einsatzbereiche geeignet ist, von denen einige möglicherweise erst zukünftig realisiert werden.

1.3 Eingrenzung des Untersuchungsgebietes

Angesichts der Vielfalt beobachtbaren Sprecherverhaltens ist es praktisch unerlässlich, die Untersuchung auf exemplarische Fälle einzugrenzen. Die vorliegende Arbeit beruht in weiten Teilen auf empirischen Untersuchungen mit Sprechern verschiedener Muttersprachen. Angesichts einer faktisch unendlichen Zahl von potentiell zu untersuchenden Sprechern und Sprachen sind einer solchen Untersuchung Grenzen gesetzt, welche eine praktische Beschränkung der Untersuchungsdomäne erforderlich machen. Diese Beschränkung erfolgt u.a. auf folgenden Ebenen:

Eingrenzung der zu untersuchenden Sprachen. Die in dieser Arbeit vorwiegend berücksichtigten Sprachen sind **Deutsch, Englisch und Französisch**, die in unterschiedlichen Kombinationen als Ausgangs- und Zielsprache untersucht werden. Die anhand dieser Sprachen gewonnenen Ergebnisse werden jedoch so formuliert, daß sie in weiten Teilen auf andere Sprachen übertragen werden können¹. Zum Teil fließen auch Beobachtungen und Teilergebnisse für andere Sprachen ein (hier vor allem **Italienisch**). Dies ist u.a. dadurch begründet, daß die Sprachdatensammlung, welche dieser Arbeit zugrundeliegt, weitere Sprachen beinhaltet (siehe hierzu Kap. 3.3). Eine Deskription und Auswertung sämtlicher Daten würde jedoch den Rahmen der vorliegenden Arbeit bei weitem überschreiten.

¹ Dies gilt mit einer nicht unwesentlichen Einschränkung: In ihrer derzeitigen Form ist die Methode nur anwendbar, wenn sowohl Ausgangs- als auch Zielsprache das lateinische Schriftsystem verwenden. Der Grund hierfür liegt u.a. in der in Kap. 7.3.4 dargelegten Modellierung „graphemabhängiger Lautersetzungen“, welche sich nur bei gemeinsamen Schriftsystemen sinnvoll anwenden läßt. Bei einem Aufeinandertreffen disparater Schriftsysteme sind andere bzw. weitere Aussprachefehler zu berücksichtigen, die im Zusammenhang mit den unterschiedlichen graphischen Repräsentationssystemen stehen.

Eingrenzung des zu untersuchenden Sprachmaterials. Als Sprachmaterial werden (vorwiegend) *Ortsnamen* aus den zu untersuchenden Zielsprachen verwendet. Eigennamen erscheinen mit Blick auf die durchzuführenden Untersuchungen aus folgenden Gründen geeignet: (a) sie sind für die Sprecher (Versuchspersonen) nicht oder nur sehr eingeschränkt unter inhaltlich-semantischen Gesichtspunkten analysierbar, so daß allein orthographische und lautliche Formeigenschaften im Mittelpunkt stehen; (b) sie sind dennoch Träger charakteristischer orthographischer und phonetischer Merkmale der Zielsprache; (c) im Gegensatz zum Standard-Wortschatz stellt die Aussprache fremdsprachiger Ortsnamen auch für solche Versuchspersonen ein realistisches Szenario dar, die über keinerlei Kenntnisse der Zielsprache verfügen (z.B. in muttersprachliche Sätze eingebettete fremde Ortsnamen in einer Reisesituation). Diese Sprechergruppe soll ausdrücklich in die Untersuchung einbezogen werden. Ortsnamen (bzw. Eigennamen allgemein) weisen jedoch eine Reihe von linguistischen Besonderheiten auf, welche auch im Kontext sprachtechnologischer Anwendungen von Bedeutung sind. Einige dieser Besonderheiten werden in Kap. 3.3 dieser Arbeit dargelegt.

Eingrenzung der untersuchten phonetischen Phänomene. Die vorliegende Arbeit beschränkt sich auf die Untersuchung segmental-phonetischer Variablen auf der Wortebene. Untersucht werden ausschließlich solche phonologischen Prozesse, die als Substitution, Tilgung, Metathese (Lautumstellung) oder Epenthese (Einschub von Lauten) von diskreten Lauteinheiten beschreibbar sind. Obwohl zweifellos einige charakteristische Akzenterscheinungen auf der suprasegmentalen Ebene angesiedelt sind (prosodische Muster, Tonakzente etc.), wird vorläufig nur die segmentale Ebene berücksichtigt. Weiterhin beschränkt sich die Untersuchung auf wortinterne Variation; wortübergreifende Effekte der Koartikulation oder Assimilation werden vorläufig nicht berücksichtigt. Jedoch ist das entwickelte Verfahren der Modellierung von Aussprachevarianten ohne technische Modifikationen auch auf diese Domäne übertragbar.

1.4 Übersicht der Kapitel

Trotz der sprachtechnologischen Zielrichtung soll die vorliegende Studie keinesfalls losgelöst von linguistischen Erkenntnissen über die Entstehung und die lautliche Beschaffenheit fremdsprachlicher Akzente durchgeführt werden.

Im nachfolgenden **Kapitel 2** werden daher zunächst in Form einer Übersichtsdarstellung einige wesentliche Überlegungen, Hypothesen und Modelle dargestellt, die in den vergangenen etwa fünf Jahrzehnten die Forschung zur Phonetik und Phonologie des Zweitspracherwerbs geprägt haben. Hieraus

ergeben sich eine Reihe von Implikationen hinsichtlich der angestrebten regelbasierten Modellierung von Aussprachevarianten; zum Teil können erste Grundlagen einer Modellierungstechnik bereits aus den vorgestellten theoretischen Modellen abgeleitet werden. Zugleich wird aber auch deutlich, welche Grenzen einem solchen Vorhaben beim derzeitigen Stand des Wissens realistischerweise gesetzt werden sollten.

Das sich daran anschließende **Kapitel 3** vollzieht zunächst einen Perspektivenwechsel: Hier wird das Problem phonetischer Varietäten und fremdsprachlicher Akzente aus sprachtechnologischer Sicht erörtert. Es soll ein Überblick der zentralen Probleme gegeben werden, die sich in diesem Zusammenhang stellen; zugleich werden die grundlegenden derzeit gängigen Techniken skizziert, mit denen man diese Probleme zu lösen versucht. Ziel des Kapitels ist es dabei nicht, eine detaillierte technische Darstellung von Modellierungstechniken für die automatische Spracherkennung oder Sprachsynthese zu geben. Vielmehr ist beabsichtigt, die vorliegende Arbeit in einem aktuellen Forschungskontext zu verorten und aufzuzeigen, welche Relevanz das vorgestellte Modell in diesem Kontext hat und an welche bereits erprobten und etablierten Methoden und Techniken dabei angeknüpft wird.

Kapitel 4 beinhaltet einige Vorüberlegungen zu den Möglichkeiten (und Grenzen) einer regelbasierten Modellierung akzentbehafteter Aussprache, die sich auf die vorausgehende Darstellung der sprachwissenschaftlichen und sprachtechnologischen Forschungslage gründen. Dabei geht es vor allem um die Frage, mit welchen Modellen und Methoden die große Bandbreite phonetischer Erscheinungen und Aussprachevarianten, die innerhalb der Domäne fremdsprachlicher Akzente zu erwarten ist, sinnvoll erfaßt werden kann. Hier wird schließlich das Modell prototypischer *Akzentstufen* eingeführt, mittels dessen die Vielfalt der möglichen Aussprachevarianten auf einige wenige prototypische Varianten reduziert wird. Das später in Kapitel 7 entwickelte Regelsystem basiert auf diesem Modell.

Eine wesentliche Grundlage dieser Arbeit bildet eine umfangreiche Sammlung von Sprachdaten mit Sprechern verschiedener Muttersprachen, die eigens zu diesem Zweck aufgezeichnet wurden. Konzeption und Aufbau der Datensammlung werden in **Kapitel 5** beschrieben. Das „Endprodukt“ dieser Arbeit – ein System zur regelbasierten Modellierung akzentbehafteter Aussprachevarianten – beruht zu großen Teilen auf Beobachtungen, die dieser Sprachdatensammlung entstammen. Durch diese empirische Fundierung soll zweierlei erreicht werden: Erstens soll vermieden werden, daß die erstellten Regeln sich allein auf linguistische Intuitionen oder Theorien stützen. Zweitens lassen sich mit Hilfe dieser Sprachdaten die mittels der Regeln generierten Varianten unmittelbar auf ihre Plausibilität prüfen. Teile der Datensammlung bilden somit die empirische Da-

tengrundlage, andere Teile dienen einer Prüfung bzw. Evaluierung des entwickelten Modells.

In **Kapitel 6** erfolgt eine Analyse der Daten bzw. der auftretenden Aussprachevarianten. Hauptziel dieser Auswertung ist es jedoch nicht, einzelne phonetische Variablen und ihre Realisierung durch verschiedene Sprecher einer detaillierten Untersuchung zu unterziehen. Obwohl die Datensammlung reichhaltiges Material hierfür bietet, das sich aufgrund seines Umfangs auch in möglichen Folgestudien nicht erschöpfen wird, ist das Hauptziel der Datenanalyse innerhalb dieser Arbeit ein anderes: Es soll eine *Fehlertypologie* entworfen werden, in der die auftretenden Aussprachevarianten bzw. -fehler nach linguistischen Ebenen (z.B. Graphem-Phonem-Beziehungen, Phonologie, Morphologie) klassifiziert werden. Der Entwurf einer solchen Typologie dient als notwendige Grundlage für den Entwurf von Regeln und ihrer linguistisch sinnvollen und ökonomischen Formulierung. Die Sprachaufzeichnungen liefern hierzu die Datengrundlage.

Schließlich wird in **Kapitel 7** die programmtechnische Implementierung des zuvor entwickelten Modells beschrieben. Hierzu werden für jedes zu untersuchende Sprachenpaar (d.h. jede Kombination aus Muttersprache des Sprechers und Zielsprache) phonologische Variationsregeln entworfen, mit denen die bei den Sprechern auftretenden prototypischen Aussprachevarianten auf der Symbolenebene nachgebildet werden. Die so erarbeiteten Regeln werden in einem sprachübergreifend arbeitenden Programm implementiert, das zu diesem Zweck erstellt wurde und somit ebenfalls Bestandteil dieser Arbeit ist.

Dieses Programm wird abschließend einer Evaluierung unterzogen, die in **Kapitel 8** beschrieben ist. Dabei wird ermittelt und quantifiziert, ob und in welchem Maße die Regeln imstande sind, das Sprecherverhalten (d.h. die Aussprachefehler der Sprecher) adäquat zu modellieren. Zu diesem Zweck wird unter anderem ein *phonetisches Distanzmaß* entwickelt, programmtechnisch implementiert und auf die Daten angewendet. Mit seiner Hilfe kann die Ähnlichkeit bzw. Approximation phonetischer Sequenzen auf der Symbolebene berechnet werden. Dieses Maß ermöglicht es, die Qualität der mittels der Regeln erzeugten Aussprachevarianten anhand ihrer Übereinstimmung mit sprachlichen Daten zu prüfen. Für diese abschließende Evaluierung ist wiederum die erstellte Sprachdatensammlung von großen Nutzen, da ein Teil dieser Daten herangezogen werden kann, um als Referenz zu fungieren. Somit greifen die einzelnen Arbeitsschritte dieser Studie auf vielfältige Arten ineinander.

I.

HINTERGRUND UND PROBLEMSKIZZE

2 Fremdsprachlicher Akzent: Sprachwissenschaftliche Perspektiven

Das folgende Kapitel nähert sich dem Thema fremdsprachlicher Akzente zunächst aus einer sprachwissenschaftlichen Perspektive. Ziel ist es, Überlegungen, Hypothesen und Modelle zusammenzufassen, die im Laufe der vergangenen ca. 50 Jahre die wissenschaftliche Untersuchung der Phonetik und Phonologie des Zweitspracherwerbs in besonderem Maße geprägt haben.

Obwohl fremdsprachliche Akzente bereits in älteren Arbeiten thematisiert wurden und sich dort Erörterungen finden, die auch für die aktuelle Theoriebildung durchaus noch relevant sind (z.B. bei Boas 1889; Polivanov 1931; Trubetzkoy 1939/89), konnte sich das Thema erst etwa seit den 1950er Jahren allmählich als eigenständiger Forschungsgegenstand etablieren. Dabei wurde es über lange Zeit als Teilgebiet der *kontrastiven Linguistik* betrachtet, woraus sich eine sehr spezifische Forschungsperspektive ergab, die im Laufe dieses Kapitels noch ausführlicher erörtert werden wird. Studien, die sich dabei dem spezifischen Problem der *Aussprache* von Fremdsprachen widmen, waren bis in die 1980er Jahre hinein noch relativ wenig verbreitet; seitdem jedoch wurde eine Vielzahl von Arbeiten zu diesem Themenkreis publiziert. Zu den traditionellen kontrastiv-linguistischen Analysen traten zudem neue Perspektiven und Erkenntnisinteressen, so daß mittlerweile ein recht breit gefächertes Spektrum von Untersuchungen vorliegt.

Darunter finden sich z.B. auch sprachpsychologische Studien, in denen die Produktion und Perzeption fremdsprachlicher Laute mit zumeist experimentellen Methoden untersucht wird; weiterhin Arbeiten, die primär von fremdsprachendidaktischen Zielen geleitet sind (Verbesserung der L2-Aussprache) sowie auch vereinzelte Untersuchungen, die sich pragmatischen Aspekten akzentbehafteter Aussprache widmen (z.B. Akzeptanz und Verständlichkeit von Akzenten). In jüngster Zeit sind weiterhin Arbeiten hinzugetreten, in denen fremdsprachliche Akzente unter spezifischen Aspekten behandelt werden, die sich im Zusammenhang mit sprachtechnologischen Anwendungen ergeben. Aufgrund ihrer besonderen Relevanz werden diese Studien gesondert im nachfolgenden Kapitel 3 behandelt, während im vorliegenden Kapitel zunächst nur solche Arbeiten dargestellt werden, die sich dem Thema aus einer vornehmlich sprachwissenschaftlichen Perspektive zuwenden.

Die Fülle von Einzelstudien läßt es jedoch unrealistisch erscheinen, im Rahmen dieser Arbeit ein vollständiges Bild der Forschungslandschaft zu geben. Eine gute Übersicht und Einführung in zentrale Forschungsfragen bieten hier u.a. Tarone (1987), Flege (1988a, 1992, 1995), Leather (1999), Major (2001) sowie

die Beiträge in den Sammelbänden Ioup & Weinberger (1987) und James & Leather (1997). Dennoch soll die Forschungslage im folgenden so weit umrissen werden, daß einige zentrale Hypothesen hervortreten, welche die Untersuchung von fremdsprachlichen Akzenten in den letzten ca. 50 Jahren dominiert haben.

Auch mit Blick auf die in dieser Arbeit angestrebte Modellierung von akzent-behafteter Aussprache für sprachtechnologische Anwendungen erscheint es durchaus lohnend, die Grundannahmen dieser „großen Hypothesen“ (cf. Bausch & Kasper 1979) nachzuzeichnen. Zum Teil ergeben sich hieraus unmittelbare oder mittelbare Implikationen für eine regelbasierte Modellierung; zum anderen Teil zeigt sich hierbei aber auch, welche prinzipiellen Grenzen bei der Deskription und Prognose des konkreten Sprecherverhaltens bestehen und welche Ansätze einer technischen Modellierung daher – zumindest vorläufig – nicht realistisch erscheinen.

2.1 „Akzent“: Eine erste Begriffsbestimmung

2.1.1 Akzente aus der Perspektive des Sprechers

Jedem Sprecher und Hörer einer beliebigen Muttersprache sind „Akzente“ vertraut. In unserer Rolle als *Sprecher* werden wir selbst zum Produzenten eines Akzents, sobald wir den Versuch unternehmen, eine fremde Sprache zu artikulieren. Hierbei werden in der Regel spezifische phonetische Abweichungen von der zielsprachlichen Norm produziert, in denen sich der Akzent manifestiert. Akzente treten aber generell bereits vor dem Überschreiten sprachlicher Grenzen auf: Auch regionale oder lokale Eigenarten der Aussprache, die phonetisch von der Norm abweichen, werden als „Akzent“ bezeichnet. Die verwandten Termini *Dialekt* und *Akzent* werden dabei üblicherweise so voneinander abgegrenzt, daß *Akzent* sich allein auf lautliche Eigenschaften der Sprechweise – also auf die Aussprache – bezieht, während *Dialekt* auch andere linguistische Merkmale wie Lexik, Syntax oder Morphologie umfaßt (cf. Hudson 1980: 43).

Wenn man die angesprochene ‚Abweichung von einer Norm‘ vorläufig als definierendes Merkmal eines Akzents heranzieht, so folgt daraus zwangsläufig die Frage, welches die Norm ist, die hierbei als Maßstab dient. Üblich ist es, als Norm die jeweilige „Hochlautung“ der betreffenden Sprache heranzuziehen, z.B. „Hochdeutsch“ für das Deutsche oder die *Received Pronunciation* für britisches Englisch. Jede davon abweichende Sprechweise würde demnach als Akzent gelten.

Es ist aber leicht zu sehen, daß diese Begriffsbestimmung sich auf ein unzuverlässiges Hilfskriterium stützt – denn letztlich ist auch die Norm selbst nur ein „Akzent“ unter vielen: Die Hochlautung qualifiziert sich keineswegs durch inhärente lautliche Eigenschaften als die geltende phonetische Norm, sondern durch willkürliche Setzungen und Konventionen der Sprachgemeinschaft. Damit ist aber auch sie ständig möglichen Wandlungen und Variationen unterworfen (siehe z.B. die Ausführungen von Wiese (2003) zur Wandlung der Normaussprache des /R/-Phonems im Deutschen). Wenn die Norm selbst jedoch nicht statisch, sondern veränderlich ist, so muß auch die phonetische Bestimmung von Akzenten permanent mit der jeweils aktuellen Norm abgeglichen und ggf. modifiziert werden. Darüber hinaus hat die Norm auch *innerhalb eines Sprachraums* keine unveränderliche und uneingeschränkte Gültigkeit: So wird beispielsweise ein „sächsischer“ oder „bayerischer Akzent“ immer dann als deutlich abweichend wahrgenommen, wenn die deutsche Hochlautung als Maßstab bzw. Norm herangezogen wird. In einer Situation dagegen, in der die überwiegende Mehrheit der Sprecher einen solchen Akzent aufweist, kann sich das Verhältnis von Norm und Akzent durchaus umkehren: Hier würde nun möglicherweise die Hochlautung, welche zuvor als Norm fungierte, ihrerseits als „abweichend“ wahrgenommen werden.

Ein Akzent kann demnach zwar prinzipiell als sprachliche „Normabweichung“ bestimmt werden – *welche* Norm dabei jedoch im Einzelfall herangezogen wird, ist letztlich von soziolinguistischen und situativen Bezugsgrößen abhängig, die in hohem Maße variabel sind. Infolgedessen wird in der Sprachwissenschaft oft eine Definition herangezogen, die den Begriff der Norm ganz vermeidet. Danach bezeichnet *Akzent* zunächst nichts anderes als eine „Art und Weise des Sprechens“ (cf. Laver 1994, Kapitel III). Es werden dabei keinerlei Aussagen über die Spezifik der so bezeichneten Sprechweise getroffen, beispielsweise über die Herkunft des Sprechers, über mögliche Abweichungen von einer Sprachnorm oder über sonstige Merkmale, die über das eigentliche phonetische Phänomen bereits hinausweisen. So weist dieser Definition zufolge *jeder Sprecher* einen Akzent auf.

Daß es dennoch gängige wissenschaftliche Praxis ist, die Hochlautung einer Sprache als Referenzform heranzuziehen, ist nicht zuletzt forschungspraktisch motiviert: Denn andernfalls müßte man sich vielfach darauf beschränken, die Existenz unzähliger voneinander abweichender „Akzente“ lediglich zu konstatieren, ohne sie mittels einer gemeinsamen Größe aufeinander beziehen zu können.

2.1.2 Akzente aus der Perspektive des Hörers

Als *Hörer* von Sprache verfügen wir über eine erstaunlich ausgeprägte Fähigkeit, Akzente wahrzunehmen. Sofern es sich um unsere Muttersprache handelt, in der gesprochen wird, sind wir zumeist zweifelsfrei in der Lage, das Vorliegen eines

Akzents zu erkennen. Die akustische Diskriminationsfähigkeit ist dabei so hoch, daß auch minimale Abweichungen von der Standard-Aussprache identifiziert werden können. Neuere Studien zeigen sogar, daß Hörer in der Lage sind, das Vorliegen eines Akzents in Sprachaufzeichnungen festzustellen, die *rückwärts* wiedergegeben werden (cf. Munro, Derwing & Burgess 2003).

Darüber hinaus kann oftmals auch die Herkunft des Sprechers benannt werden („englischer Akzent“, „deutscher Akzent“ usw.). Dies gilt selbst dann, wenn der Hörer über keinerlei aktive Kenntnisse der Herkunftssprache des Sprechers verfügt. Trotz der Verschiedenheit und individuellen Prägung von Sprechweisen scheint also ein *prototypischer* Akzent zu existieren, dessen Identifizierung uns zumeist leicht fällt.

Weitaus größere Schwierigkeiten entstehen hingegen, wenn man aufgefordert ist, den akustischen Gesamteindruck zu analysieren und z.B. einzelne lautliche Merkmale zu benennen, in denen sich die Abweichungen eines „englischen Akzents“ oder eines „japanischen Akzents“ manifestieren. Trotz der oft mühelosen Identifikation von Akzenten erscheint es schwierig, präzise anzugeben, *worin* die Abweichungen bestehen. Der umgangssprachliche Ausdruck „Akzent“ bezeichnet also offenbar einen ganzheitlichen akustischen Eindruck, der nicht ohne weiteres in seine Bestandteile zergliedert und analysiert werden kann.

Es ist eines der Ziele der vorliegenden Arbeit, eine solche Analyse vorzunehmen. Dies ist eine Grundvoraussetzung für eine regelbasierte Modellierung akzentbehafteter Aussprache. Damit wird jedoch nicht der Anspruch erhoben, die konstitutiven Merkmale akzentbehafter Aussprache vollständig zu beschreiben. Der akustische Gesamteindruck, von dem oben die Rede war, resultiert aus einem komplexen Gefüge phonetischer, prosodischer, lexikalischer und grammatischer Merkmale, die erst in ihrem Zusammenwirken eine Sprechweise konstituieren, die wir als typischen fremdsprachlichen Akzent bezeichnen würden. Die segmental-phonetischen Erscheinungen, die im Rahmen dieser Studie behandelt werden, bilden nur einen Teil dieses Gefüges.

2.2 Sprachvergleich und Fehlerprognosen

2.2.1 Kontrast, Transfer und Interferenz

Eine zwar inzwischen in vielen Annahmen revidierte, aber dennoch wichtige und prägende Phase der Forschung zum Zweitsprachenerwerb ist von einer theoretischen Richtung dominiert, die im Rückblick oft als „starke Kontrastivhypothese“ (*strong contrastive hypothesis*) bezeichnet wurde. Geprägt wurde diese Phase, die

man etwa auf den Zeitraum von 1950 bis 1970 datieren kann, vor allem durch die Arbeiten von Fries (1945), Weinreich (1953), Haugen (1953) und Lado (1957).

Wollte man die Grundannahmen der „Kontrastivhypothese“ in aller Kürze thesenhaft zusammenfassen, so könnte eine solche Quintessenz wie folgt lauten:

(1) *Strukturelle Kontraste zwischen den Sprachsystemen L1 und L2 bilden die Hauptschwierigkeit bei der Produktion von L2-Äußerungen durch einen Sprecher von L1.*

(2) *Eine fehlerhafte Übertragung von Elementen aus L1 auf L2 ist die zentrale Fehlerquelle in der L2-Produktion und damit die Hauptursache für akzent-behaftete Sprache.*

(3) *Es besteht eine unmittelbare Beziehung zwischen der Stärke des linguistischen Kontrastes von L1 und L2 einerseits und der Art und Stärke der auftretenden Fehler andererseits.*

(4) *Daher ermöglicht die Identifizierung der linguistischen Kontraste eine Erklärung und Prognose von Fehlern.*

Diese aufeinander aufbauenden Thesen bildeten im Kern das Forschungsprogramm einer kontrastiven Linguistik, deren Ziel nicht allein in einer vergleichenden Gegenüberstellung zweier Sprachen A und B besteht, sondern auch in einer unmittelbar daraus abgeleiteten *Deskription*, *Erklärung* und *Prognose* von sprachlichen Fehlern, die in Situationen des Kontakts zwischen A und B auftreten. Ausgangspunkt ist dabei stets die Ermittlung sprachlicher Kontraste zwischen zwei Sprachsystemen L1 und L2.

Dabei steht die folgende Überlegung im Mittelpunkt: Ein Sprecher einer Muttersprache L1, der eine Fremdsprache L2 erlernt, ist zunächst mit einem Sprachsystem L2 konfrontiert, welches auf allen linguistischen Ebenen – von der Phonetik bis zur Semantik und Pragmatik – Abweichungen von L1 aufweist. Je nach zu erlernender L2 sind diese Abweichungen unterschiedlich stark ausgeprägt; zudem können die Abweichungen auf verschiedenen linguistischen Ebenen verschieden stark sein. Ein Sprecher des Deutschen z.B., der Niederländisch als L2 erlernt, wird in den Teilbereichen Lexik, Morphologie und Syntax nur relativ moderate Abweichungen zu seiner L1 vorfinden. Erlernt derselbe Sprecher hingegen z.B. Japanisch als L2, so treten in der Zielsprache extreme Abweichungen auf. Eine der Kernannahmen der starken Kontrastivhypothese liegt nun darin, daß zwischen der Stärke des linguistischen Kontrastes der beteiligten Sprachsysteme L1 und L2 und der Art und Stärke der von Lernern produzierten Fehlern ein unmittelbarer Zusammenhang besteht.

Im Rahmen dieses Forschungsparadigmas spielten die Begriffe *Kontrast*, *Transfer* und *Interferenz* eine zentrale Rolle. Es erscheint angebracht, die ge-

nannten Termini kurz zu erörtern und voneinander abzugrenzen, zumal sich bei Sichtung der diesbezüglichen Literatur der Eindruck einstellt, daß *Interferenz* und *Transfer* (z.T. ergänzt auch durch *Transferenz*) nicht immer in einheitlicher und miteinander konformer Weise verwendet werden (cf. hierzu auch Tesch 1978: 31-41).

(1) **Kontrast.** Zwei Sprachen L1 und L2 weisen in einer systematischen - und auf verschiedenen linguistischen Ebenen gesondert durchzuführenden - Gegenüberstellung ihrer Elemente stets spezifische Gemeinsamkeiten und Unterschiede auf. In den Unterschieden manifestiert sich ein *Kontrast* zwischen L1 und L2. Mit dieser Feststellung, welche die Ebene des Sprachsystems als Bezugsgröße heranzieht, sind zunächst noch keinerlei Aussagen über das Sprachverhalten der Sprecher von L1 und L2 getroffen. Eine der zentralen Annahmen der Kontrastivhypothese ist es jedoch nun, daß ein unmittelbarer Zusammenhang zwischen den auf der Ebene des Sprachsystems identifizierten Kontrasten und den bei nicht-muttersprachlichen Sprechern zu beobachtenden Fehlern der L2-Produktion besteht.

(2) **Transfer.** Der Begriff *Transfer* stammt ursprünglich aus der (Lern-) Psychologie und bezeichnet hier einen Einfluß eines bereits erlernten Materials auf das Erlernen eines folgenden (cf. Fröhlich 1997: 414). Transfer in dieser allgemeinen Bedeutung meint also zunächst die *Übertragung* von bekannten Wissensbeständen oder Fähigkeiten auf Neues, Unbekanntes, und damit einen Prozeß, der bei jeder Art von Lernen eine wichtige Rolle spielt. Weinreich (1953) führte den Begriff in einer eng daran angelehnten Bedeutung in die Sprachkontaktforschung ein: Hier wird eine Übernahme von linguistischen Merkmalen aus einer Sprache L1 (i.d.R. die Muttersprache) in eine Sprache L2 als Transfer bezeichnet. So werden beim Erlernen einer Fremdsprache die aus der eigenen Muttersprache geläufigen Strukturen herangezogen und auf die Fremdsprache transferiert. Ein solcher Transfer vollzieht sich nicht nur auf der hier im Mittelpunkt stehenden Ebene der Aussprache (Phonetik), sondern auch auf lexikalischer, morphologischer, syntaktischer oder sogar pragmatischer Ebene.

Grundsätzlich kann Transfer in allen Lernsituationen, so auch beim Erlernen einer Fremdsprache, dem Lernerfolg zuträglich sein. Allgemein ist dies dann der Fall, wenn der Lernstoff dem bereits bekannten Material *ähnlich* ist. Dies wird auch für den L2-Erwerb und dessen Beeinflussung durch die vorhandene Kenntnis einer L1 angenommen: In solchen Bereichen, in denen sich die beteiligten Sprachen hinreichend ähnlich sind, kann der Transfer die Aneignung der Zielsprache erleichtern. So ist beispielsweise die Aussprache einiger Kernkonsonanten und -vokale in vielen europäischen Sprachen zumindest ähnlich, so daß der Sprecher die entsprechenden Muster aus seiner Muttersprache übernehmen kann: Will z.B. ein Sprecher des Deutschen die Aussprache des spanischen Phonems

/a/ erlernen, so kann er das entsprechende deutsche Phonem /a/ als Muster heranziehen und gelangt damit (vorläufig) zu einem positiven Lernergebnis. In diesem Fall liegt daher *positiver Transfer* vor (cf. Hammarström 1966: 98).

Während positiver Transfer dem Lernprozeß förderlich ist, kommt es bei strukturellen Kontrasten zwischen L1 und L2 oftmals zu einer fehlerhaften Anwendung oder Generalisierung der aus der Muttersprache bekannten Muster. Ein solcher *negativer Transfer* resultiert generell aus Diskrepanzen zwischen vorhandenem Wissen und neuem Lernmaterial. Die transferierten Muster sind in diesem Fall inadäquat für das neue Material und führen daher zu Abweichungen vom gewünschten Lernergebnis, d.h. zu Fehlern.

(3) **Interferenz.** Resultat dieses Transfers ist eine sich im Sprechakt manifestierende „Störung“ der sprachlichen Norm von L2. Diese Störung wird schließlich als *Interferenz* bezeichnet. Auch das Konzept der Interferenz ist aus der Lernpsychologie übernommen. Tritt ein negativer Transfer von Lerninhalten auf, so kommt es zu einer sog. *proaktiven Hemmung*, d.h. einer Beeinträchtigung des Erlernens eines neuen Lerninhaltes durch die Einwirkung eines früher gelernen Inhaltes. Das Resultat ist eine Überlagerung und Vermengung von neuen und alten Inhalten und damit eine Störung des Lernerfolgs.

Der von Weinreich (1953, 1957) in einen sprachwissenschaftlichen Kontext eingeführte Begriff der Interferenz bildete einen zentralen Bestandteil der „Kontrastivhypothese“ und ist trotz späterer Kritik bzw. der Revision des kontrastiven Modells noch heute als Element von Theorien der L2-Sprachproduktion wiederzufinden. Weinreich selbst gab folgende Begriffsbestimmung:

By phenomena of *interference* I mean those instances of deviation from the norms of a language which occur in the speech of a bilingual as the result of his familiarity with another language, i.e. as a result of language contact.
(Weinreich 1957: 1; Hvhbg. im Orig.)

Diese Definition ist noch sehr allgemein gehalten, da lediglich vom *Kontakt zweier Sprachen* die Rede ist, ohne die im Rahmen der Zweitspracherwerbsforschung später übliche Einschränkung, daß es sich stets (a) um die Erstsprache des Sprechers sowie (b) eine zu erlernende Zweitsprache handelt, welche vom Sprecher nur unvollkommen beherrscht wird.

Tatsächlich betont Weinreich an anderer Stelle die Irrelevanz von Begriffen wie „Muttersprache“, „Erstsprache“, „Zweitsprache“ für die Bestimmung sprachlicher Interferenz (cf. Weinreich 1953: 74). Somit wird auch eine *wechselseitige* Beeinflussung der Sprachen nicht ausgeschlossen: Der Einfluß verläuft nicht notwendigerweise von der Erstsprache zur Zweitsprache, sondern kann prinzipiell auch in umgekehrter Richtung verlaufen (psychologisch gesprochen handelt

es sich um eine sog. *retroaktive Hemmung*, wenn neue Lerninhalte auf bereits vorhandene Wissensbestände negativ einwirken).

2.2.2 Interferenz und Fehler im Fremdsprachenerwerb

In Weinreichs Bestimmung finden sich wichtige Elemente des Interferenzbegriffs, welche auch in späteren Modifikationen und Neudefinitionen erhalten geblieben sind. Dazu gehören folgende Aussagen:

- (a) Die Interferenz manifestiert sich in Form von *Abweichungen von den Normen* einer Sprache;
- (b) sie tritt in der *Rede von bilingualen Sprechern* auf (wobei man dem Kontext von Weinreichs Zitat entnehmen kann, daß der Begriff *bilingual* hier nicht im Sinne von „vollständig zweisprachig“ verstanden wird);
- (c) ihre Ursache liegt in der Kenntnis und Vertrautheit (*familiarity*) des Sprechers mit einer anderen Sprache und
- (d) sie tritt generell in einer Situation des *Sprachkontaktes* auf.

Gegenüber den noch recht allgemein gefaßten Begriffsbestimmungen Weinreichs, mit denen versucht wurde, die begriffliche Trichotomie *Kontrast - Transfer - Interferenz* für die linguistische Forschung fruchtbar zu machen, weisen viele im Rahmen der Kontrastivhypothese publizierten Arbeiten einen explizit didaktischen Hintergrund auf.

So trägt etwa das einflußreiche Werk von C. Fries (1945) den Titel *Teaching and Learning English as a Foreign Language*; Robert Lados Hauptwerk *Linguistics across Cultures* (1957) hat den Untertitel *Applied Linguistics for Language Teachers*. Der kontrastive Ansatz war also zumeist eng mit dem praktischen Ziel verbunden, methodische sprachpädagogische Verbesserungen aus linguistischen Analysen abzuleiten. Fehleranalyse, Fehlerkorrektur und sogar Fehlerprognose sollten aus einer vergleichenden Gegenüberstellung von Sprachsystemen hergeleitet werden. Die Leitthese dieses Ansatzes formuliert z.B. Lado (1957) wie folgt:

[...] individuals tend to transfer the forms and meanings, and the distribution of forms and meanings of their native language and culture to the foreign language and culture - both productively when attempting to speak the language and to act in the culture, and receptively when attempting to grasp and understand the language and the culture as practiced by natives.
(Lado 1957: 2)

Dieser Ansatz wurde dabei für alle linguistischen und z.T. auch darüber hinausgehenden Ebenen herangezogen. So ist Lados Werk *Linguistics Across Cultures*,

das als repräsentativ für den hier dargestellten Ansatz gelten kann, in sechs Hauptkapitel gegliedert, von denen fünf mit „*How to compare two ...*“ betitelt sind. Diese Methode eines paarweisen Vergleichs mit anschließender Analyse von Fehlerquellen wendet Lado in gleicher Weise auf *Sound Systems*, *Grammatical Structures*, *Vocabulary Systems*, *Writing Systems* und schließlich auf *Cultures* an. Die kontrastive Analyse erstreckt sich also von Lautsystemen über verschiedene höhere linguistische Ebenen wie Syntax, Lexikon und Schrift bis hin zu ganzen Kulturen, die mittels des Verfahrens zunächst in diskrete Elemente aufgespalten, einander gegenübergestellt und schließlich bezüglich der aus ihnen resultierenden Lernschwierigkeiten ausgewertet werden.

2.2.3 Beispiel einer kontrastiven Analyse

Das Verfahren soll anhand eines Beispiels aus dem hier zur Diskussion stehenden Bereich der Phonetik und Phonologie verdeutlicht werden: Stellt man beispielsweise die Phonemsysteme des Deutschen und des Englischen vergleichend gegenüber, so treten Gemeinsamkeiten und Differenzen hervor. Die folgende Tabelle zeigt eine solche Gegenüberstellung am Beispiel der Konsonantenphoneme des Deutschen und Englischen:

Konsonanten	DEUTSCH	ENGLISCH
<i>Plosive</i>	p b t d k g	p b t d k g
<i>Affrikaten</i>	p̄f̄ t̄s̄	t̄ʃ̄ d̄ʒ̄
<i>Frikative</i>	f v s z ʃ ʒ ² ç x h	f v θ ð s z ʃ ʒ h
<i>Nasale, Liquide</i>	m n ŋ r l	m n ŋ r l
<i>Approximanten</i>	j	w j
Anzahl Phoneme	23	24

Tabelle 2.1: Konsonantenphoneme des Deutschen und Englischen

Festzustellen ist zunächst, daß der Umfang (Anzahl der Phoneme) beider Konsonanten-Inventare annähernd gleich ist. Die zeilenweise Gruppierung der Phoneme nach ihrer Artikulationsart zeigt, daß dabei viele Übereinstimmungen, aber auch vereinzelte Abweichungen bestehen: Während z.B. die Klasse der Plosive in beiden Sprachen identisch ist, treten bei den Affrikaten, Frikativen und Appro-

² Das Phonem /ʒ/ tritt im Deutschen nur in Fremdwörtern auf, z.B. in *Genie*, *Etage*, *Regime*.

ximanten erkennbare Unterschiede auf. So weist das Deutsche gegenüber dem Englischen zwei zusätzliche Affrikaten $/\widehat{pf}/$ und $/\widehat{ts}/$ auf, während das Englische über $/\widehat{tʃ}/$ und $/\widehat{dʒ}/$ sowie einen zusätzlichen Approximanten $/w/$ verfügt. Deutliche Unterschiede treten bei den Frikativen auf, wo im Englischen die zusätzlichen Interdental-Laute $/\theta/$ und $/\delta/$ auftreten, das Deutsche hingegen über die im Englischen nicht vorhandenen Phoneme $/x/$ und $/ç/$ verfügt³.

Die zur Diskussion stehenden Konsonantensysteme lassen sich demzufolge auf der Dimension *Ähnlichkeit-Kontrast* beschreiben, zunächst mit dem allgemeinen Ergebnis, daß es (a) Elemente gibt, die in *beiden* Systemen *identisch* sind (Ähnlichkeit), und (b) Elemente, die nur in *einem der beiden Systeme* auftreten (Kontrast). Um nun genauer zu spezifizieren, an welchen Stellen der Systeme Ähnlichkeit und Kontrast vorliegen, können die fraglichen Elemente zweier Sprachen in Form zweier Mengen L1 und L2 dargestellt werden, zwischen denen Überlappungen und Differenzen bestehen:

$L1 = \{P_1, P_2 \dots P_n\}$	Menge der Elemente in L1
$L2 = \{P_1, P_2 \dots P_n\}$	Menge der Elemente in L2
$L1 \cap L2$	Schnittmenge von L1 und L2
$L1 \setminus L2$	Differenzmenge: Elemente aus L1, die nicht in L2 enthalten sind
$L2 \setminus L1$	Differenzmenge: Elemente aus L2, die nicht in L1 enthalten sind

Wendet man eine solche Darstellung in Form von Mengen nun auf das oben genannte Beispiel der Konsonantensysteme des Deutschen und Englischen an, so erhält man folgende Mengen von Phonemen:

ENGLISCH = $\{ p, b, t, d, k, g, \widehat{tʃ}, \widehat{dʒ}, f, v, \theta, \delta, s, z, ʃ, ʒ, h, m, n, \eta, r, l, w, j \}$

DEUTSCH = $\{ p, b, t, d, k, g, \widehat{pf}, \widehat{ts}, f, v, s, z, ʃ, ʒ, ç, x, h, m, n, \eta, r, l, j \}$

DEUTSCH \cap ENGLISCH = $\{ p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, h, m, n, \eta, r, l, j \}$

DEUTSCH \setminus ENGLISCH = $\{ \widehat{pf}, \widehat{ts}, ç, x \}$

ENGLISCH \setminus DEUTSCH = $\{ \theta, \delta, w, \widehat{tʃ}, \widehat{dʒ} \}$

³ $/x/$ und $/ç/$ werden hier als gesonderte Phoneme aufgeführt (der Analyse von Ungeheuer (1969/77: 74-75) folgend). Tatsächlich ist ihr phonemischer Status umstritten, und es wurden taxonomische Argumente angeführt, sie als Allophone eines einzigen zugrundeliegenden Phonems $/ç/$ bzw. $/x/$ zu werten (cf. z.B. Meinhold & Stock 1982: 133).

Hieraus kann nun eine gegenüber der oben gezeigten Tabelle modifizierte Darstellungsform abgeleitet werden, in der die gemeinsamen und die kontrastierenden Elemente deutlich sichtbar hervortreten:

Konsonanten	DEUTSCH \ ENGLISCH	DEUTSCH \cap ENGLISCH	ENGLISCH \ DEUTSCH
<i>Plosive</i>		p b t d k g	
<i>Affrikaten</i>	\overline{pf} \overline{ts}		$\overline{tʃ}$ $\overline{dʒ}$
<i>Frikative</i>	ζ x	f v s z ʃ ʒ h	θ δ
<i>Nasale, Liquide</i>		m n ŋ r l	
<i>Approximanten</i>		j	w

Tab 2.2: Konsonanteninventare des Deutschen und Englischen: Überschneidungen und Differenzen

Die schattierten Bereiche in der zweiten bzw. vierten Spalte enthalten diejenigen Elemente, bei denen ein Kontrast zwischen L1 und L2 vorliegt, während die mittlere Spalte die identischen Elemente zeigt. Ergebnis dieser Analyse wäre demnach, daß die Differenzmenge DEUTSCH \ ENGLISCH die Phoneme \overline{pf} , \overline{ts} , ζ und x umfaßt, während die Menge ENGLISCH \ DEUTSCH die Elemente θ , δ und w sowie $\overline{tʃ}$ und $\overline{dʒ}$ enthält. Bei Plosiven, Nasalen und Liquiden besteht vollständige Übereinstimmung der Phonem-Inventare, wohingegen bei den Affrikaten die Schnittmenge leer ist, d.h. keine Übereinstimmung besteht.

Genau an diesem Sachverhalt setzt nun die Kontrastivhypothese in ihrer Beschreibung möglicher Fehler des L2-Lernenden an. Danach treten Fehler vorrangig dort auf, wo ein Kontrast zwischen den Sprachsystemen L1 und L2 vorliegt (d.h. im Bereich der Differenzmengen). Während ähnliche bzw. identische Elemente der beteiligten Sprachsysteme prinzipiell leicht erlernt werden können, da sie unverändert aus der L1 des Sprechers übernommen werden können, führen kontrastierende Elemente zu Interferenz zwischen den Sprachsystemen - und damit zu Fehlern. Elemente der Schnittmenge $L1 \cap L2$ sind, so die Hypothese, relativ problemlos zu erlernen, während die Elemente von $L1 \setminus L2$ bzw. $L2 \setminus L1$ mit tendenziell größeren Schwierigkeiten verbunden sind.

Bezogen auf das dargestellte Beispiel der Konsonantensysteme des Deutschen und Englischen würde dies z.B. bedeuten, daß für Muttersprachler des Englischen die Realisierung der deutschen Phoneme x und ζ prinzipiell fehlerträchtig ist, da diese im Englischen nicht auftreten. Für Segmente wie $ʃ$, g oder η wird dagegen - aufgrund der Überschneidung mit dem englischen Konsonantensystem - ein problemloser Transfer auf das Deutsche und damit eine weitgehend akkurate Aussprache prognostiziert.

2.2.4 Kritik der Kontrastivhypothese

Etwa seit den späten 1960er Jahren wurde die Gültigkeit der Kontrastivhypothese in der Form, wie sie bis hierher skizziert wurde, in Zweifel gezogen. Die Kritik bezog sich vor allem auf (1) den unzureichenden prognostischen Wert der kontrastiven Analyse, (2) die Betonung des negativen Transfers als ausschließlicher Fehlerquelle; (3) die Vernachlässigung von Befunden zum Erstsprachenerwerb, die sich auch für den Zweitsprachenerwerb als relevant erweisen könnten. Einige zentrale Aspekte dieser Kritik seien im folgenden zusammengefaßt.

Prognostischer Wert kontrastiver Analysen

Vielfach wurde der mangelnde prognostische Wert als ein wesentliches Defizit des kontrastiven Modells herausgestellt. Die durch kontrastive Analyse ermittelten Differenzbereiche zwischen L1 und L2 seien nicht geeignet, die in der Lernersprache tatsächlich auftretenden Fehler hinreichend genau vorherzusagen. Das Modell könne dem Anspruch, Fehler der L2-Produktion (z.B. Aussprachefehler) allein aufgrund der Kenntnis der linguistischen Kontraste zwischen L1 und L2 zuverlässig zu *prädizieren* letztlich nicht gerecht werden. Die Fehlerprognosen der kontrastiven Analyse, so resümiert McLaughlin (1989: 212) die Einwände, seien gleichzeitig zu stark und zu schwach: Zu stark deshalb, weil Fehler prognostiziert wurden, die in der Praxis nicht auftreten, zu schwach andererseits, weil viele der tatsächlich auftretenden Fehler durch die kontrastive Analyse *nicht* prognostiziert werden.

Dabei wurde jedoch der Erklärungsanspruch der kontrastiven Analyse von seinen Kritikern oftmals überzeichnet wiedergegeben (cf. hierzu auch James 1971). So ist z.B. in Lados Text zumeist nicht von einer *exakten Prognose* von Fehlern die Rede, sondern vielmehr von einer Identifizierung linguistischer Einheiten, deren Produktion für nicht-muttersprachliche Sprecher tendenziell mit größeren Schwierigkeiten verbunden ist. Lado spricht von "similarity and difference as determiners of ease and difficulty" (1957: 21, 59) sowie davon, daß Sprecher allgemein dazu *tendieren* (1957: 2, 11 und *passim*), Fehler insbesondere bei linguistisch kontrastierenden Einheiten zu produzieren. Zutreffend ist dennoch, daß *allein* eine kontrastive Analyse der Sprachsysteme L1 und L2 nicht geeignet ist, alle auftretenden Fehler nicht nur post factum als Resultat von Transfer und Interferenz zu *erklären*, sondern darüber hinaus auch zuverlässig zu *prädizieren*.

Die genannte Kritik des mangelnden prognostischen Wertes der Kontrastivhypothese speiste sich allerdings häufig aus konkreten didaktischen Bedürfnissen des Fremdsprachenunterrichts, die man nicht erfüllt sah. So findet man bei Cor-der (1967/74) folgende Einschätzung der kontrastiven Analyse:

Teachers have not always been very impressed by this contribution from the linguist for the reason that their practical experience has usually already shown them where [...] difficulties lie and they have not felt that [...] the linguist has provided them with significantly new information. They noted for example that many of the errors with which they were familiar were not predicted by the linguist anyway. (Corder 1967/74: 19)

Man kann im forschungsgeschichtlichen Rückblick jedoch die Frage aufwerfen, ob diese Kritik nicht allein auf einer grundlegenden Divergenz der Erkenntnisinteressen beruht: Während das Ziel linguistischer Untersuchungen darin besteht, leitende Prinzipien zu ermitteln, die in der Sprachproduktion relevant werden, ist das praktische Interesse des Fremdsprachendidaktikers ein anderes: Ihm geht es nicht zuletzt darum, dem Lernenden konkrete Strategien an die Hand zu geben, um die Sprachproduktion zu verbessern, d.h. Fehler zu vermeiden oder zu überwinden. Es ist aber nicht das primäre Ziel der Linguistik, solche praktischen Strategien bereitzustellen. Daher ist es sicher verfehlt, ihr einen grundlegenden theoretischen Mangel zu attestieren, weil sie dies nicht leistet.

Dennoch war es nicht zuletzt das Argument des unzureichenden prognostischen Potentials, das zu einer Reformulierung der Kontrastivhypothese führte. An die Stelle der sog. „starken“ Kontrastivhypothese (*strong claim*), wie sie in den 1950er und 1960er Jahren verfolgt wurde, trat zu Anfang der 1970er Jahre die „schwache“ Kontrastivhypothese (*weak claim*) als leitende Theorie. In dieser Reformulierung wurde der Transfer von L1-Elementen auf die Zielsprache L2 zwar weiterhin als Fehlerquelle anerkannt, allerdings mit zwei entscheidenden Modifikationen: (1) An die Stelle der *Vorhersage* von Fehlern trat der schwächere Anspruch, beobachtete Fehler nachträglich zu *erklären*; (2) es wurden neben dem L1-Transfer auch andere Fehlerquellen anerkannt (cf. Wardhaugh 1970).

Parallelen zwischen L1- und L2-Erwerb

Etwa zu Beginn der 1970er Jahre zeichnete sich jedoch ein Paradigmenwechsel und eine vollständige Aufgabe der Kontrastivhypothese ab. In einer relativ kurzen Forschungsphase, die besonders durch die Arbeiten von Dulay & Burt (1973, 1974) geprägt war, traten lingualer Transfer und daraus resultierende Interferenz als Ursachen für akzentgefärbte Aussprache ganz in den Hintergrund. Statt dessen wurden mögliche Parallelen zwischen dem Erlernen einer L2 und dem kindlichen Erstspracherwerb als Fehlerursache betont.

Die in der L2-Produktion auftretenden Fehler seien, so die Kernthese, in ihrer Art und ihrer Entwicklungsabfolge den Fehlern vergleichbar, die auch beim Erwerb der Zielsprache als Erstsprache zu beobachten sind. Produziert etwa ein Sprecher des Deutschen den fehlerhaften englischen Satz „*I not like that*“, so ist grundsätzlich auszuschließen, daß hier ein transferbedingter Fehler vorliegt, da

der Satzbau *weder* der Zielsprache Englisch *noch* der Ausgangssprache Deutsch entspricht⁴. Zu konstatieren ist hingegen, daß diese syntaktische Konstruktion von Kindern im Laufe des Erstsprachenerwerbs der L1 Englisch durchaus produziert wird. So besteht die zentrale These der aus derartigen Beobachtungen abgeleiteten *Identitätshypothese* darin, daß der Erwerb von L1 und das Erlernen einer L2 grundsätzlich ähnlichen bzw. identischen Prinzipien unterliegt und sich Fehler des L2-Lernenden auf der Grundlage dieser Prinzipien erklären lassen. In der Literatur wird diese theoretische Orientierung daher bisweilen mit der plakativen Formel *L1-Erwerb = L2-Erwerb* überschrieben.

Als Hauptvertreter dieser Richtung unterscheiden Dulay & Burt (1974: 115) grundsätzlich folgende im Zweitsprachenerwerb auftretende Fehlertypen:

- (i) **Interferenzartige Fehler**, in denen sich Merkmale der L1 des Sprechers widerspiegeln und die im Erstsprachenerwerb der Zielsprache *nicht* beobachtet werden können.
- (ii) **Entwicklungsbedingte Fehler** (*developmental errors*), die nicht auf Merkmale der L1 des Sprechers zurückgeführt werden können, jedoch im Erstsprachenerwerb der Zielsprache zu beobachten sind.
- (iii) **Mehrdeutige Fehler** (*ambiguous errors*), bei denen nicht eindeutig entscheidbar ist, ob sie Prozesse des Erstsprachenerwerbs reflektieren oder aus der L1 des Sprechers erklärt werden können.
- (iv) **Idiosynkratische Fehler**, die weder in Daten zum Erstsprachenerwerb beobachtet werden können, noch in der Struktur der L1 des Sprechers zu finden sind.

In einer Studie über spanische Sprecher, die Englisch als L2 erlernten, ermittelten die Autoren die relativen Häufigkeiten dieser Fehlertypen in ihren Sprachdaten und gelangten zu dem überraschenden Ergebnis, daß (nach der Eliminierung mehrdeutiger Fehler) 85% aller Fehler dem Typus (ii) zugerechnet werden können und nur 3% der Fehler auf Erstsprachen-Interferenz zurückzuführen seien (Dulay & Burt 1973). Sprachlicher Transfer, wie er von der kontrastiven Analyse als Hauptfehlerquelle angenommen wurde, tritt hier also zugunsten von entwicklungsbedingten Einflußfaktoren völlig zurück.

Obgleich solch eindeutige Ergebnisse in späteren Studien nicht repliziert werden konnten, wurde die Erklärungsadäquatheit der kontrastiven Analyse durch diese Resultate zunächst in erheblichem Maße in Frage gestellt. Jedoch wurden nach relativ kurzer Zeit methodologische Mängel der Studien von Dulay & Burt offenkundig, die Anlaß zu weiteren Folgeuntersuchungen gaben. Die Ergebnisse konnten nicht repliziert werden; die der „Identitätshypothese“ zugrundeliegenden Annahmen konnten somit nicht verifiziert werden (cf. McLaughlin 1989:

⁴ Dieses Beispiel ist eine Variation eines Beispiels von Dulay & Burt (1974: 112).

213). Jedoch behält ein im Rahmen der Identitätshypothese vorgebrachtes Argument weiterhin seine Gültigkeit: Fehler in der L2-Produktion lassen sich *nicht ausnahmslos* als Folge negativen Transfers erklären. Für bestimmte Fehlertypen scheidet diese Erklärung schon dadurch aus, daß keine entsprechende Struktur in L1 existiert, aus der sich der Fehler ableiten ließe.

Fehlerquellen außerhalb der Interferenz

Verschiedene Studien zu Beginn der 1970er Jahre lieferten empirische Indizien dafür, daß weitere Fehlerursachen neben dem sprachlichen Transfer aus der Muttersprache zu berücksichtigen sind. Für den Bereich der Phonetik und Phonologie gehören hierzu beispielsweise:

- die Arbeiten von Brière (1968), der eine aus arabischen, französischen und vietnamesischen Elementen zusammengesetzte artifizielle Mischsprache verwendete und ihre Aussprache durch Sprecher des (amerikanischen) Englischen untersuchte;
- eine umfangreiche Studie von Johansson (1973) für die L2 Schwedisch mit Sprechern neun verschiedener Muttersprachen sowie
- die empirischen Untersuchungen, die Nemser (1971b, 1971c) mit ungarischen Sprechern für die Zielsprache Englisch durchführte.

Zwar entsprachen einige Aussprachefehler, die in diesen Studien von Sprechern produziert wurden, durchaus den Prognosen einer kontrastiven Analyse. Jedoch produzierten die untersuchten Sprecher für bestimmte zielsprachliche Laute auch Ersatzlaute, die im Inventar ihrer L1 *nicht* vorkamen (cf. Brière 1968: 65-68). Derartige Beobachtungen fallen grundsätzlich aus dem Erklärungsrahmen der starken Kontrastivhypothese heraus, die ja von der dichotomischen Annahme ausgeht, daß *entweder* ein negativer Transfer von L1-Elementen stattfindet *oder* die L2-Struktur korrekt realisiert wird. Mögliche unvollkommene Approximationen von L2-Lauten oder auch eine Substitution durch Laute, die weder L1 noch L2 zugeschrieben werden können, sind hierbei nicht vorgesehen. Tatsächlich weisen die Lernaltersprachen zum Teil artikulatorische Tendenzen auf, die unabhängig von der L1 des Sprechers sind: So besteht bei der Substitution von L2-Lauten eine generelle Tendenz, extreme Artikulationspositionen in Richtung einer neutralen Stellung der Artikulatoren zu modifizieren, etwa durch die Zentralisierung von Vokalen (cf. Johansson 1973).

Darüber hinaus konnte festgestellt werden, daß nicht immer ein direkter Zusammenhang zwischen lautlicher Ähnlichkeit oder Fremdartigkeit und dem damit verbundenen Schwierigkeitsgrad für den Lerner besteht (wie es von der Kontrastivhypothese vorhergesagt wurde): Die Produktion mancher L2-Laute, die ein phonetisch sehr ähnliches Gegenstück in der L1 des Sprechers hatten, bereiteten größere Schwierigkeiten, während andere, die keinem L1-Laut ähnlich waren,

weitgehend problemlos artikuliert werden konnten (cf. Johansson 1973). Diese letzte Beobachtung wird weiter unten unter dem Terminus *Äquivalenzklassifizierung* aufgegriffen.

2.3 Interlanguage und approximative Systeme

2.3.1 Lernersprachen als eigenständige Sprachsysteme

Infolge grundsätzlicher Kritik an der Kontrastivhypothese verlagerte sich das Interesse der L2-Forschung in den frühen 1970er Jahren von den transferbedingten Einflüssen auf die Zielsprache hin zu einer Betrachtung der Lernersprachen als eigenständige Sprachsysteme. Dies kommt insbesondere in den Konzepten der *Interlanguage* (Selinker 1972) sowie der *approximativen Systeme* (Nemser 1971a) zum Ausdruck. Beide Entwürfe sind vor allem von dem Bestreben geprägt, die Lernersprachen nicht länger als reines Transfer- und Interferenzprodukt zu betrachten, sondern als einen Untersuchungsgegenstand *sui generis* zu etablieren, ohne dabei jedoch die Idee eines zugrundeliegenden und der Untersuchung prinzipiell zugänglichen Sprachsystems aufzugeben. Selinker bezeichnete dieses System als *Interlanguage* (verschiedentlich mit dt. *Interimsprache* übersetzt), die er definiert als ein “[...] separate linguistic system based on the observable output which results from a learner's attempted production of a TL [target language] norm” (Selinker 1972: 35).

Im Modell der Interlanguage (IL) werden zwar einerseits Elemente des kontrastiven Ansatzes (wieder) aufgegriffen, jedoch wendet man sich in einem wesentlichen Punkt davon ab: Nicht mehr die *Ähnlichkeit* oder *Verschiedenheit* (Kontrast) von L1 und L2 und der sich daraus ergebende Einfluß auf die Sprachproduktion von Lernenden steht im Mittelpunkt, sondern die IL als ein eigenständiges Sprachsystem. Der Begriff „Transfer“ erfährt im Rahmen dieses Modells zwar neuerliches Interesse, allerdings mit einer Reihe von Modifikationen gegenüber früheren Theorie-Ansätzen: So wird beispielsweise auch das Vermeidungsverhalten von L2-Lernenden (sprachliche Merkmale von L2, die in L1 nicht vorhanden sind, werden gar nicht erst produziert) als Transferphänomen behandelt (cf. McLaughlin 1989: 215). Weiterhin steht – im Sinne des *weak claim* der Kontrastivhypothese – nicht länger die *Vorhersage* von Fehlern im Mittelpunkt, sondern deren nachträgliche *Erklärung*.

Selinker (1972: 35) unterscheidet dabei fünf „zentrale Prozesse“ (*central processes*), welche die Oberflächenform von Interlanguage-Äußerungen bestimmen.

Dabei ist die durch negativen Transfer bedingte Interferenz nur einer dieser Faktoren.

- (1) **Sprachlicher Transfer** (*language transfer*). Hierbei handelt es sich um Transfer und daraus folgende Interferenz im traditionellen Sinne der kontrastiven Analyse.
- (2) **Übungstransfer** (*transfer of training*). Die in der Lernsituation angebrachten Beispielformen werden von den Lernern als repräsentativ für die gesamte Zielsprache angesehen und daher übergeneralisierend angewendet.
- (3) **Lernerstrategien** (*strategies of second language learning*). Der Sprecher verändert die Struktur des zu erlernenden Materials im Laufe des Aneignungsprozesses; die Form seiner L2-Äußerungen reflektiert diese Veränderungen. Dies resultiert beispielsweise in einer Vereinfachung zielsprachlicher Strukturen („Es ist großes Haus“; „Ich habe schreiben Brief“) oder sogar ihrer vollständigen Vermeidung.
- (4) **Kommunikationsstrategien** (*strategies of second language communication*). Strukturen der Zielsprache werden vom Sprecher vereinfacht, um die Kommunikationssituation nicht unnötig zu belasten. Hierbei werden v.a. solche Elemente ausgelassen, die dem Sprecher als nicht zwingend notwendig zur Erreichung des aktuellen Kommunikationsziels erscheinen.
- (5) **Übergeneralisierung** (*overgeneralization*). Erlernte linguistische Regeln werden vom Sprecher in unzulässiger Weise ausgedehnt auf neues L2-Material, für das die Regeln nicht gelten.

Neben diesen fünf zentralen Prozessen identifiziert Selinker weitere maßgebliche Faktoren, z.B. den Einfluß der muttersprachlichen Schriftaussprache auf die Aussprache von L2, Hyperkorrektheit oder holophrastisches Lernen (ibid., 41).

Ganz unabhängig aber von der Frage nach der Vollständigkeit oder Gültigkeit der von Selinker vorgeschlagenen Fehlertaxonomie liegt die bedeutendere Weiterentwicklung gegenüber dem kontrastiven Ansatz sicher darin, daß gegenüber der monokausalen Erklärung (Fehler als Resultat von Transfer) nun ein komplexeres Wirkungsgefüge bei der Entstehung von *Interlanguages* angenommen wird, bei dem verschiedene linguistische, psychologische und kommunikative Ursachen zugleich berücksichtigt werden.

2.3.2 Approximationen und Übergangsstufen

Ein der *Interlanguage* verwandtes Modell lieferte Nemser (1971a) mit der Einführung des Modells *approximativer Systeme*. Ausgangspunkt bildet auch bei Nemser eine Kritik an der Kontrastivhypothese. In ihrer starken Form - d.h. verbunden mit dem Anspruch einer Prädiktion von Fehlern auf der Grundlage

eines Sprachvergleichs – sei diese Hypothese problematisch, da sich (1) je nach theoretischer Grundlage unterschiedliche Voraussagen ergeben können, (2) die getroffenen Voraussagen oftmals mehrdeutig seien und (3) verschiedene linguistische Ebenen wie Phonologie, Morphologie, Lexik oder Syntax stets Interdependenzen aufweisen, so daß eine kontrastive Analyse, die nur eine dieser Ebenen isoliert untersucht, als verkürzt gelten müsse. Aber auch mit der schwachen Form der Kontrastivhypothese, welche die Fehler des Lernenden nicht voraussagen, sondern lediglich *erklären* will, seien oft beträchtliche Probleme verbunden, weil für einen Fehlertyp zumeist verschiedene konkurrierende Erklärungen in Frage kommen, die sich in Abhängigkeit von Analyseverfahren und leitender Theorie erheblich unterscheiden können (Nemser 1971a: 121).

Diese Probleme lassen sich Nemser zufolge nur überwinden, wenn man die von den Lernenden angewandten Sprachsysteme nicht primär als aus L1 oder L2 abgeleitete Systeme betrachtet, sondern als unabhängige, eigenständige Systeme. Es handle sich dabei um vorübergehende Sprachsysteme, die sich der Lernende von der Zielsprache aufgebaut hat und die als sog. *approximative Systeme* einen eigenen, von den Sprachsystemen L1 und L2 verschiedenen Untersuchungsgegenstand konstituieren:

An *approximative system* is the deviant linguistic system actually employed by the learner attempting to utilize the target language. Such approximative systems vary in character in accordance with proficiency level; variation is also introduced by learning experience (including exposure to the target language script system), communication function, personal learning characteristics, etc. (Nemser 1971a: 116).

Der Sprachlernprozeß lasse sich dabei als eine Abfolge von approximativen Systemen beschreiben, die sich zunehmend dem Sprachsystem L2 annähern. Die Phasen des Sprachlernprozesses unterscheiden sich nicht nur nach der Stärke, sondern auch nach Art der Fehler. Bestimmte Fehler treten typischerweise in den frühen Phasen auf, während andere meist in späteren Phasen beobachtet werden können (ibid., 119/120).

Auch Nemser unterstellt dabei eine *Systemhaftigkeit* der Lernaltersprachen. Zwar handle es sich bei approximativen Systemen um temporäre, vom jeweiligen L2-Kennntnisstand des Lernenden geprägte Übergangsstadien, andererseits weisen jede dieser Übergangsstufen eine strukturelle Autonomie auf: Die Sprache des Lernenden sei “structurally organized, manifesting the order and cohesiveness of a system” (ibid., 116). Diese Systematizität äußere sich auch darin, daß bei der Kombination einer bestimmten Ausgangs- und Zielsprache *typische* approximative Systeme auftreten, insofern die approximativen Systeme von Lernenden, die auf der gleichen Lernstufe stehen, oftmals bestimmte Ähnlichkeiten aufweisen („*roughly coincide*“). So seien beispielsweise ähnliche Fehlertypen bei Lernenden

mit gleichem muttersprachlichen Hintergrund zu beobachten, die auf einer Lernstufe befindlich sind (ibid., 117).

Ein wesentlicher Unterschied zwischen regulären Sprachen und approximativen Systemen bestehe jedoch darin, daß Sprecher, die eine Fremdsprache erlernen, keine genuinen Sprachgemeinschaften bilden. Es fehle daher die konstante Verstärkung von außen durch andere Sprecher. Dennoch bildeten sich auch in Lerngemeinschaften mitunter Gruppen, in denen sich bestimmte Merkmale approximativer Systeme stabilisieren und langfristig erhalten bleiben (ibid., 120). Ein besonders starkes Indiz für die Existenz einheitlicher approximativer Systeme liefern dabei sogenannte „stabile Systeme“ (*stable systems*), zu denen beispielsweise Einwanderersprachen gehören. Ihre Sprecher haben allgemein ein hohes Niveau der Beherrschung der jeweiligen Zielsprache erreicht; dennoch weist ihre Sprache einzelne Merkmale auf, in denen muttersprachliches Niveau in der Zielsprache nicht erreicht wird. Diese Merkmale sind einerseits relativ stabil, insofern sie sich einer weiteren Annäherung an die zielsprachliche Norm zu widersetzen scheinen; andererseits erreichen sie eine relativ hohe interindividuelle Konstanz, d.h., sie treten in ähnlicher Weise in der gesamten Sprechergruppe auf.

2.4 Perzeptive Assimilation und Äquivalenzklassifizierung

2.4.1 Einfluß der Perzeption auf die Produktion

Schon in frühen Arbeiten zur allgemeinen Phonetik und Phonologie findet man die Überlegung, daß beim Erwerb einer Fremdsprache und deren Lautunterscheidungen das Phonemsystem der eigenen Muttersprache eine prägende Rolle einnimmt, welche bereits bei der *Perzeption* von L2-Lauten wirksam wird. So findet sich bereits im Rahmen des *Cercle Linguistique de Prague* bei Polivanov (1931) der Hinweis, daß die Laute einer Fremdsprache stets in Kategorien akustisch wahrgenommen werden, welche durch das muttersprachliche Phonemsystem vorgeprägt sind. In die gleiche Richtung zielt Trubetzkoy (1939/89) in seiner Beschreibung des muttersprachlichen Lautsystems als „Filter“, durch welchen alle L2-Laute wahrgenommen und kategorisiert werden. Neue Sprachlaute werden demzufolge den bereits bekannten Lautkategorien (der L1) untergeordnet.

Die Idee, die hier zum Ausdruck kommt, ist die einer *Vorgängigkeit der Perzeption* von L2-Lauten gegenüber ihrer Produktion: Wenn fremdsprachliche Laute vom Lernenden bereits in gewisser Weise „verfälscht“ aufgenommen werden, so ist grundsätzlich anzunehmen, daß dies auch Auswirkungen auf die Produkti-

on dieser Laute hat. Ein solcher Einfluß wird noch *vor* allen potentiellen artikulatorisch-phonetischen Hindernissen bei der Produktion von L2-Lauten wirksam.

Diese Annahme hat seit den 1980er Jahren wieder zunehmendes Interesse auf sich gezogen und wurde seitdem in verschiedenen experimentellen Untersuchungen geprüft. Die zentrale Frage hierbei ist die nach einer möglichen *Korrelation* einerseits und nach einer *kausalen Beziehung* von Perception und Produktion andererseits. Diese Beziehungen können unterschiedlich verlaufen, wobei prinzipiell folgende Relationen auftreten können:

- (1) Die Perception geht der Produktion voraus;
- (2) perzeptive und produktive Fertigkeiten verlaufen parallel;
- (3) die Produktion geht der Perception voraus.

Die Frage nach dem Verhältnis von Perception und Produktion fremdsprachlicher Laute ist dabei nicht nur von theoretischem Interesse, sondern mit unmittelbaren fremdsprachendidaktischen Konsequenzen verbunden: Wäre beispielsweise nachweisbar, daß bestimmte Lautdistinktionen der Zielsprache von Lernenden bereits auf perzeptiver Ebene nur mit Schwierigkeiten getroffen werden können, so sollte ein gezieltes Training solcher Fertigkeiten in den Unterricht einbezogen werden (zum positiven Einfluß eines perzeptiven Trainings auf die L2-Produktion cf. Logan et al. 1991; Yamada 1993; Jamieson 1995).

Die bisherigen Ergebnisse sind jedoch nicht eindeutig interpretierbar. Eine Reihe von Studien deutet auf eine ausgeprägte Korrelation von perzeptiven und produktiven Fähigkeiten im Erwerb von L2-Lauten hin. So konnte Barry (1989) in einer Untersuchung der Produktion von englischen Vokalen durch deutsche Muttersprachler zeigen, daß die Bildung stabiler perzeptiver Kategorien von L2-Phonemen oftmals mit einer akkuraten (d.h. weniger akzentbehafteten) Produktion dieser Phoneme einhergeht. In dieselbe Richtung weisen Ergebnisse einer Studie von Rochet (1995), in der mit Hilfe synthetisch erzeugter Laute die perzeptive Kategorisierung französischer Vokale durch Sprecher des (kanadisch) Englischen und (brasilianisch) Portugiesischen ermittelt wurden. Die bei diesen Sprechergruppen beobachteten systematisch auftretenden Produktionsfehler – französisches [y] wurde von den kanadisch-englischen Sprechern oft als [u], von den brasilianisch-portugiesischen Sprechern dagegen als [i] realisiert – reflektierten die im Perzeptionstest aufgetretenen Kategorisierungsfehler.

Ergebnisse wie diese können als Indiz für eine positive Korrelation perzeptiver und produktiver Fähigkeiten interpretiert werden. Dennoch stellt sich die Sachlage nicht eindeutig dar. So kann eine *Vorgängigkeit* perzeptiver Fertigkeiten gegenüber der Produktion fremdsprachlicher Laute – im Sinne einer kausalen Ursache-Wirkungs-Beziehung – nicht ohne weiteres daraus abgeleitet werden. Beispielsweise konnten Sheldon & Strange (1982) in einer Studie mit japanischen

Muttersprachlern für die Zielsprache Englisch zeigen, daß in der untersuchten Sprechergruppe die Distinktion der englischen Phoneme /l/ und /r/ zwar artikulatorisch-phonetisch hinreichend akkurat produziert wurde, dieselbe Distinktion jedoch auf perceptiver Ebene nicht mit der gleichen Genauigkeit vollzogen werden konnte. In diesem Fall gehen also die produktiven Fertigkeiten der Sprecher über die perceptiven hinaus; die These einer *generellen* Vorgängigkeit der Perception erscheint demnach auf der Basis derartiger Ergebnisse fraglich.

In den späten 1980er Jahren erfuhr die Frage nach dem Verhältnis von Perception und Produktion von L2-Lauten jedoch wichtige Präzisierungen. Eine wesentliche Differenzierung wurde durch die Einführung von sprecher- bzw. sprechergruppenspezifischen Variablen erreicht. Die zugrundeliegende Frage richtet sich damit nicht länger auf das *generelle* Verhältnis von Perception und Produktion, sondern darauf, ob dieses Verhältnis möglicherweise in Abhängigkeit von der untersuchten Sprechergruppe variieren kann.

So fanden Bohn & Flege (1990), daß der Grad der Erfahrung bzw. Vertrautheit mit der Zielsprache nicht nur als eine generelle Einflußgröße auf die Fertigkeiten und die Akkuratheit der L2-Produktion wirksam wird, sondern sich möglicherweise auch das Verhältnis von perceptiven und produktiven Fertigkeiten in Abhängigkeit vom Lernstadium verändern kann: Während in einem frühen Stadium des L2-Erwerbs ein generell starker perceptiver Einfluß phonemischer Kategorien der Muttersprache auftritt, durch den die Kategorisierung von L2-Lauten (und in der Folge auch ihre Produktion) gelenkt wird, ist diese Art der Korrelation bei erfahrenen L2-Lernenden nicht in gleichem Maße gegeben. Vielmehr scheint sich ein ausgiebiger Kontakt mit der Zielsprache in späteren Lernstadien zwar positiv auf die *produktiven* Fertigkeiten auszuwirken, jedoch ist dies nicht notwendigerweise von zunehmenden *perceptiven* Fertigkeiten begleitet. In Abhängigkeit vom jeweiligen Lernstadium des Sprechers ist also möglicherweise von unterschiedlichen Zusammenhängen zwischen Perception und Produktion auszugehen (cf. Flege 1988b).

2.4.2 Äquivalenzklassifizierung und Stabilisierung von Lautkategorien

In den 1990er Jahren wurden verschiedene Modelle entworfen, in denen die oben erwähnte Vorstellung vom muttersprachlichen Phonemsystem als „Filter“ der Sprachwahrnehmung weitergeführt, präzisiert und in Teilen auch experimentellen Untersuchungen zugeführt wurde. Zu diesen Modellen gehören etwa das *Perceptual Assimilation Model* (Best 1994) oder die *Perceptual Magnet Theory* (Kuhl & Iverson 1995). Trotz im Detail unterschiedlicher Konzeptionen beruhen sie auf der Grundidee, daß auf der Grundlage des muttersprachlichen Phonemsystems eine perceptive „Umorganisation“ und Assimilation einlaufender Sprachlaute vollzogen wird. Hierdurch werden fremdsprachliche Laute auf einer bestimmten Stufe der perceptiven und kognitiven Verarbeitung an das L1-System

stimmten Stufe der perzeptiven und kognitiven Verarbeitung an das L1-System angeglichen. *Wahrnehmung* und *Kategorisierung* sind demnach entscheidende Einflußgrößen bei der Entstehung von fremdsprachlichen Akzenten. McAllister (1997) spricht in diesem Zusammenhang von einem „perceptual foreign accent“, welcher – z.B. in Form einer reduzierten Fähigkeit zur phonetischen Diskrimination fremdsprachlicher Laute – auch bei Sprechern mit guten oder sehr guten Kenntnissen der Zielsprachen auftreten kann. Dieser manifestiert sich u.a. darin, daß Nicht-Muttersprachler in Situationen, in denen der Kommunikationskanal beeinträchtigt ist (z.B. durch Störgeräusche), meßbar größere Verständnisschwierigkeiten haben als Muttersprachler.

Eines der am weitesten ausgearbeiteten Modelle der Produktion und Perzeption von L2-Lauten wurde von J. E. Flege mit seinem *Speech Learning Model* (Flege 1988b; 1992; 1995) entworfen. Grundsätzlich wird auch im Rahmen dieses Modells das Verhältnis von Perzeption und Produktion als ein zentraler Aspekt der Genese fremdsprachlicher Akzente betrachtet. Als ein entscheidender Faktor wird hierbei der *Grad der Stabilität* des muttersprachlichen Lautsystems wirksam. Je weiter die perzeptiven Kategorien dieses Lautsystems sich im Laufe der Ontogenese stabilisieren, desto schwieriger ist es für den Sprecher/Hörer, neue Kategorien aufzubauen. Der Grad der Verfestigung, Stabilisierung und Undurchlässigkeit des muttersprachlichen Lautsystems wird damit zu einem wichtigen Indikator für die Performanz des Sprechers bei der Produktion fremdsprachlicher Laute.

Ein wichtiges Konzept im Rahmen dieses Modells ist die sog. „Äquivalenzklassifizierung“ (*equivalence classification*; cf. Flege 1987). Hiermit ist der Sachverhalt gemeint, daß L2-Laute von einem Sprecher dauerhaft in Kategorien wahrgenommen werden, die durch das Phonemsystem seiner L1 vorgegeben sind. Obgleich akustische Differenzen zwischen fremdsprachlichen und muttersprachlichen Lauten bestehen, werden die fremdsprachlichen Laute mit Lauten identifiziert, welche einer L1-Kategorie zugehörig sind. Fleges Modell beinhaltet auch spezifische Hypothesen darüber, *welche* L2-Laute in *welcher* Phase des L2-Erwerbs einer Äquivalenzklassifizierung unterworfen sind. Grundlegend dabei ist die Unterscheidung zwischen (a) „identischen“, (b) „neuen“ und (c) „ähnlichen“ L2-Lauten, die wie folgt verläuft (cf. Flege 1995: 572f. und 1997: 16f.):

(a) **Identische Laute.** Einige Laute im L2-Inventar sind phonetisch identisch mit L1-Lauten. Ihre Produktion bildet für den Lernenden in der Regel kein Problem, da hier ein positiver Transfer erfolgen kann, der in einer authentischen Aussprache der L2-Laute resultiert.

(b) **Ähnliche Laute.** Diese L2-Laute sind zwar von L1-Lauten verschieden, weisen dabei jedoch eine ausgeprägte phonetische Ähnlichkeit mit bestimmten L1-Kategorien auf.

(c) **Neue Laute.** Einige L2-Laute sind akustisch so verschieden von den im L1-Inventar enthaltenen Lauten, daß sie keiner bekannten Kategorie zugeordnet werden können.

Eine zentrale Hypothese im Rahmen dieses Modells ist nun, daß der Sprecher bzw. Hörer für die „neuen“ Laute eigene phonetische Kategorien etablieren kann, die *dauerhaft neben den vorhandenen muttersprachlichen Kategorien existieren* können. So wird z.B. ein Sprecher des Deutschen zumeist imstande sein, für die englischen Interdentale [θ] und [ð] neue Kategorien zu bilden, da diese Laute im Inventar des Deutschen nicht auftreten und keine phonetischen Äquivalente, d.h. deutsche Laute mit ausgeprägter Ähnlichkeit existieren.

Anders verhält es sich dagegen bei den „ähnlichen“ Lauten. Aufgrund ihrer akustisch-phonetischen Nähe zu vorhandenen muttersprachlichen Lautkategorien sind sie oftmals einer Äquivalenzklassifizierung unterworfen, so daß die vorhandenen phonetischen Differenzen bereits perceptiv nivelliert werden. Die Etablierung einer eigenen perceptiven Kategorie gelingt für diese Laute oftmals nicht. Beispielsweise wird ein Sprecher des Französischen, der das Deutsche erlernt, die zwei Realisierungsvarianten des deutschen Phonems /p/

- (1) als aspirierter Plosiv [p^h] und
- (2) als nicht-aspirierter Plosiv [p]

oftmals bereits perceptiv nicht unterscheiden können. Zum einen sind beide Laute akustisch sehr ähnlich, zum zweiten aber liegt hier eine allophonische Differenzierung vor, die im Französischen nicht auftritt. Der Sprecher wird dazu neigen, beide Laute der Phonemkategorie /p/ des Französischen zuzuordnen, die lediglich die nicht-aspirierte Realisierungsvariante [p] aufweist. Somit wird die Distinktion zwischen [p^h] und [p] vom Sprecher als nicht relevant erachtet und in der Folge auch nicht produziert.

Die Voraussagen über die lautliche Beschaffenheit von Akzenten, die sich aus dem Modell einer solchen „Äquivalenzklassifizierung“ ableiten lassen, laufen früheren Annahmen z.T. diametral entgegen. Der traditionelle kontrastive Ansatz, wie er an früherer Stelle in diesem Kapitel skizziert wurde, ging noch davon aus, daß L2-Laute, die über ein Äquivalent (also ein phonetisch sehr ähnliches Gegenstück) in L1 verfügen, von einem L1-Sprecher ohne Schwierigkeiten zu realisieren sind, während insbesondere die stark abweichenden Laute problematisch hinsichtlich ihrer Produktion sind. Es läge demnach zumeist eine *positive Korrelation* zwischen phonetischem Kontrast und Akzentstärke vor. Genau diese Annahme wird mit Fleges Modell gewissermaßen auf den Kopf gestellt: Statt dessen

wird nun angenommen, daß gerade in der phonetischen *Ähnlichkeit* zwischen L1- und L2-Lauten eine wesentliche Ursache für akzentbehaftete Aussprache liegt – die Korrelation zwischen phonetischem Kontrast und Akzentstärke ist demnach oftmals negativ.

Fleges Modell beinhaltet darüber hinaus Hypothesen zur Wirksamkeit des Prinzips der Äquivalenzklassifizierung in verschiedenen Phasen der sprachlichen Ontogenese. Als Ursache der Äquivalenzklassifizierung nimmt Flege eine *perzeptive Stabilisierung* des muttersprachlichen Phonemsystems an. Grundsätzlich sei eine Kategorisierung einlaufender Sprachsignale und eine perzeptive Festigung des Phonemsystems als ein wichtiger und notwendiger Teil der sprachlichen Entwicklung anzusehen. Hierdurch werde es erst möglich, von der phonetischen Vielfalt der umgebenden Sprache zu abstrahieren und die sprachlich relevanten Lauteigenschaften zu fokussieren (cf. Flege 1987: 34). Die Phase der Stabilisierung von Phonemkategorien dauert Flege (1997: 14) zufolge etwa bis zum Alter von 5 bis 6 Jahren an. Während dieser Phase sei das System noch in der Entwicklung befindlich und weithin offen für eine stabile und dauerhafte Etablierung von *zusätzlichen* bzw. *eigenständigen* Kategorien für L2-Laute. Mit zunehmender Festigung des muttersprachlichen Systems nehme diese Fähigkeit jedoch ab, wodurch zugleich eine Verstärkung von Effekten der Äquivalenzklassifizierung auftrete.

Hinsichtlich der Herausbildung von Akzenten und dem Erreichen einer authentischen Aussprache in einer L2 läßt sich hieraus die Hypothese ableiten, daß der Erwerb einer gesprochenen Fremdsprache während dieser Phase – d.h. bis zum Alter von 5 bis 6 Jahren – auf optimale Voraussetzungen trifft. Bei einem späteren Erwerb der L2 steigt die Wahrscheinlichkeit eines kaum überwindbaren Akzents. Flege interpretiert dies jedoch ausdrücklich *nicht* im Sinne einer „kritischen Periode“ (siehe unten Kap. 2.5), d.h. einer biologisch determinierten Reifungsphase, nach deren Ablauf bestimmte sprachliche Fähigkeiten nicht mehr erworben werden können.

Vielmehr tritt bei Flege an die Stelle einer hypothetisierten „kritischen Periode“ die Hypothese der „phonologischen Übersetzung“ (*phonological translation hypothesis*; cf. Flege 1981). Ihr zufolge besteht grundsätzlich eine Kontinuität in der Entwicklung und Ausbildung lautsprachlicher Fähigkeiten. So können phonetische Kategorien für „neue“ L2-Laute durchaus auch noch im Erwachsenenalter etabliert werden (cf. Flege 1997: 42). Jedoch verhindere die Ausbildung eines stabilisierten Phonemsystems die Bildung neuer Kategorien für solche Laute, die den bereits bekannten Kategorien ähnlich sind. Die „ähnlichen“ L2-Laute bilden daher oftmals einen Problembereich hinsichtlich einer authentischen Aussprache, da sie eine perzeptive „Übersetzung“ in Richtung des muttersprachlichen Phonemsystems erfahren.

2.5 Überwindung von Akzenten und „Fossilisierung“

Eines der wiederkehrenden und höchst umstrittenen Kernthemen der Sprachwissenschaft ist die Frage nach den biologischen Voraussetzungen, unter denen Sprache erworben wird. Verbunden damit ist die Frage nach möglichen angeborenen kognitiven Strukturen, welche unserer Sprachfähigkeit zugrunde liegen, dem Einfluß des Lernens auf die Sprachentwicklung innerhalb der Ontogenese sowie möglichen Wechselwirkungen zwischen beiden Domänen. Teil dieser Frage ist auch, ob und wie weit die generelle Fähigkeit zum Erwerb von Sprache mit spezifischen biologischen „Reifungsprozessen“ korreliert und somit z.B. an ein bestimmtes Lebensalter gebunden ist.

Eine der prägnantesten Formulierungen dieser These findet sich in Eric Lennebergs „Critical Period Hypothesis“ (cf. Lenneberg 1972: 196-219), derzufolge sich der Spracherwerb innerhalb einer bestimmten Phase der Ontogenese – der sogenannten „kritischen Periode“ – mit relativ geringer bewußter Anstrengung vollzieht, während nach Ablauf dieser Phase solche mühelosen Lernerfolge nicht mehr erzielt werden können. Lenneberg datiert diese Periode etwa auf den Zeitraum vom zweiten Lebensjahr bis zur Pubertät und vertritt die These, daß die Existenz einer solche Periode auf spezifische biologisch-neurologische Voraussetzungen des Menschen zurückzuführen sei. Während Lenneberg selbst sich jedoch kaum der Phonetik und ihrem Erwerb widmete, übertrugen andere Autoren dieses Modell auch auf den Erwerb fremdsprachlicher Aussprache. So wurde von einigen Autoren unter Berufung auf eine *Critical Period Hypothesis* die These vertreten, daß eine perfekte Aussprache einer Fremdsprache nach dem Eintritt der Pubertät grundsätzlich nicht mehr zu erreichen sei (z.B. Scovel 1969) – entsprechende Instruktionsbemühungen wären demzufolge vergeblich.

Die *Critical Period Hypothesis* ist hinsichtlich ihrer ausgeprägt biologischen Fundierung sowie der angenommenen oberen Altersgrenze stark umstritten (für eine Zusammenfassung dieser Diskussion cf. Long 1990; Scovel 2000; Snow 1988; Major 1997). Die generelle Aussage jedoch, daß der Erwerb von Sprache – sowohl der Muttersprache als auch einer Fremdsprache – in einem bestimmten, relativ frühen Lebensalter auf die optimalen Voraussetzungen trifft und sich in einer späteren Lebensphase weniger leicht vollzieht, dürfte bei Vertretern opponierender Lager gleichermaßen Zustimmung finden. Bei der Entwicklung phonetisch-phonologischer Fähigkeiten zeigt sich dies etwa darin, daß Kinder, die frühzeitig eine Fremdsprache erlernen, diese später oftmals weitgehend akzentfrei sprechen, während „Spätlerner“ zumeist einen deutlich hörbaren Akzent behalten.

Im Bereich der Phonetik bzw. Aussprache erscheint der Einfluß des Lebensalters beim Erwerb der Fremdsprache besonders ausgeprägt. Scovel (1988) gelangt sogar zu der Einschätzung, daß die *Critical Period Hypothesis* ausschließlich im

Bereich der Phonologie Gültigkeit besitzt, während sie auf anderen linguistischen Ebenen wie z.B. Lexikon/Wortschatz oder Syntax nicht nachgewiesen werden könne. Ein oft zitiertes – anekdotisches – Beispiel hierfür liefert der Autor Joseph Conrad, der aus seiner Heimat Polen nach Großbritannien emigrierte. Obgleich er mit Werken, die von ihm in der Fremdsprache Englisch verfaßt wurden, zu einem anerkannten Schriftsteller wurde, dem man eine meisterhafte Beherrschung des Englischen attestierte, behielt er bis zuletzt einen sehr stark ausgeprägten polnischen Akzent. Scovel (1969) prägte in Anlehnung daran den Ausdruck „Joseph-Conrad-Phänomen“ für eine derartige Diskrepanz zwischen einzelnen sprachlichen Teilfähigkeiten – einer perfekten Beherrschung der fremdsprachlichen Syntax, Semantik und des Wortschatzes einerseits, andererseits aber einer nur unzulänglichen, stark von der Muttersprache geprägten Beherrschung der L2-Aussprache.

Die Beobachtung, daß ein fremdsprachlicher Akzent nur äußerst schwierig bzw. letztlich nicht vollständig zu überwinden ist, gab Anlaß zu verschiedenen Hypothesen und Studien. Selinker (1972: 41) führte in diesem Zusammenhang den Begriff der *Fossilisierung* (engl. *fossilization*) ein: Von fossilisierten Elementen spricht Selinker dann, wenn die produktiven Fähigkeiten in einer zu erlernenden Zielsprache trotz nachhaltiger Instruktion und anhaltendem Kontakt mit der Zielsprache auf einem bestimmten Niveau stehen bleiben. Ein Beispiel für ein solches fossilisiertes Element der L2-Aussprache ist die dauerhafte Beibehaltung des uvularen „r“-Lautes [ʀ] bei französischen Sprechern des Englischen, welche ein hervorstechendes Merkmal eines französischen Akzents bildet. Charakteristisch für die Fossilisierung sei dabei, so Selinker (1972: 41/42), daß Lerner, die dieses Stadium im Prinzip bereits überwunden haben, in bestimmten Situationen (z.B. bei begrenzter Aufmerksamkeit) wieder in eine frühere Stufe zurückfallen.

Tarone (1987: 80-83) stellt zusammenfassend verschiedene Hypothesen einander gegenüber, die vorgeschlagen wurden, um das dauerhafte Verbleiben bestimmter L2-Fertigkeiten auf einem unvollkommenen oder fragmentarischen Niveau zu erklären. Hierzu gehören:

- **Motorisch-physiologische Erklärungen** wie etwa ein „Einfrieren“ der erforderlichen Artikulatoren in einem bestimmten Lebensalter. Hierbei handelt es sich jedoch um eine alltagsweltliche Erklärung, für die noch kein wissenschaftlicher Beleg gefunden werden konnte.
- Erklärungen, die auf einer Variante der **Critical Period Hypothesis** beruhen. Diese Erklärungen gründen sich zumeist auf biologische Argumente (und knüpfen damit unmittelbar an Lenneberg an); sie wurden jedoch zum Teil auch mit entwicklungspsychologischen Erweiterungen versehen, die über die rein biologisch-neurologische Ontogenese hinausweisen.

- **Veränderungen der Lernstrategien:** Während kindlicher Spracherwerb sich weitgehend unbewußt vollzieht, tritt später ein bewußtes Lernen an diese Stelle. So beruht die sog. *Monitor-Theorie* (Krashen 1981) auf der Annahme, daß es sich bei Erwerb (*acquisition*) und Erlernen (*learning*) um zwei grundverschiedene Aneignungsmodi handelt, die zu entsprechend unterschiedlichen Leistungen beim Erwerb der L2 führen.
- **Psychologische Gewohnheitsbildung** (*habit formation*) und dadurch begründeter Transfer muttersprachlicher Muster auf die Zielsprache. Ein solcher Transfer kann auch darin bestehen, daß die in der Lernsituation aufgetretenen Beispiele in unangemessener Weise in die Zielsprache übertragen werden.
- **Das Fehlen affektiver Bindungen** bei erwachsenen L2-Lernern, welche im kindlichen Spracherwerb als wichtige Motivationsquelle wirksam werden. Hierbei handelt es sich letztlich um außerlinguistische Einflüsse, die aber stets als potentielle Einflußgrößen zu berücksichtigen sind.

Wenn auch bisher keine Erklärung als alleinige Ursache einer „Fossilisierung“ identifiziert werden konnte, so finden sich dennoch empirische Hinweise, die bestimmte Entwicklungslinien zumindest plausibel erscheinen lassen. Vieles deutet z.B. darauf hin, daß die Grundlagen für das spätere Vorhandensein eines Akzents bereits in frühester Kindheit gelegt werden. Das Phonemsystem der Muttersprache übt offenbar schon früh einen prägenden Einfluß aus, der letztlich auch zu Lasten der Fähigkeit zur Verarbeitung fremdsprachlicher Laute gehen kann. So konnten Werker & Tees (1984) zeigen, daß Kinder, die Englisch als Muttersprache erlernen, bereits im Alter von 6 bis 12 Monaten eine Abnahme ihrer Fähigkeit zur perceptiven Distinktion fremdsprachlicher Laute aufweisen. Zurückgeführt wird dies auf den Einfluß des muttersprachlichen Lautsystems, welches in dieser Altersstufe wenn auch noch nicht produktiv, so doch bereits *perzeptiv* wirksam ist (die generelle Idee einer perzeptuellen Reorganisation lautsprachlicher Wahrnehmung auf der Basis des muttersprachlichen Phonemsystems soll weiter unten in diesem Kapitel in einem anderen Kontext noch einmal aufgenommen werden).

Zugleich scheint auch die vorsprachliche *Produktion* von Lauten bei Säuglingen erste sprachspezifische phonetische Merkmale der sie umgebenden Muttersprache aufzuweisen. Beispielsweise fanden Boysson-Bardies et al. (1986, 1989) in einer Reihe von Studien, daß die Lautproduktion von Säuglingen bereits im Alter von 8 Monaten über charakteristische spektrale Merkmale verfügt, die auch bei erwachsenen Sprechern der betreffenden Muttersprache vorzufinden seien. Darüber hinaus seien erwachsene Sprecher in der Lage, anhand von Aufnahmen von Kindern dieses Alters zu unterscheiden, ob es sich um Kinder ihrer eigenen Muttersprache – in diesem Falle Französisch – oder um arabische oder chinesische Kinder handelt (cf. Boysson-Bardies et al. 1984).

Die genannten Schwierigkeiten bei der Überwindung eines Akzents gaben auch Anlaß zu Untersuchungen, in denen nach Trainingsmethoden zur Verbesserung der phonetischen Performanz in einer Fremdsprache gesucht wurde. Beispielsweise konnte Neufeld (1977) mittels einer Instruktionmethode, bei der die Versuchspersonen gezwungen waren, sich vollständig und ausschließlich auf die Lautung der untersuchten Zielsprache(n) zu konzentrieren (d.h. unter Ausklammerung der Bedeutung) in seiner Versuchsgruppe fast muttersprachliche Fertigkeiten der Aussprache erreichen. Jedoch ist dabei zu berücksichtigen, daß für diese Studie eine artifizielle Lernsituation etabliert wurde, welche (1) sowohl semantische Inhalte und pragmatische Einbettungen der Äußerungen ausblendet und in der (2) mit einem stark begrenzten Vokabular operiert wurde. Die Übertragbarkeit der Ergebnisse auf reale Lernsituationen erscheint daher fraglich.

Obgleich einige zunächst vielversprechende Ergebnisse der 1970er und 1980er Jahre nicht dauerhaft in den festen Bestand der Lehrmethoden eingegangen sind, deuten Einzelergebnisse darauf hin, daß das genannte „Joseph-Conrad-Phänomen“ tatsächlich durch Anwendung neuer Methoden zumindest abgeschwächt werden kann. Das gezielte Erlernen einer akkurateren Produktion von L2-Lauten erscheint demnach durchaus möglich. Neuere Studien, die in diese Richtung weisen, sind z.B. Bongaerts et al. (1997), Nikolov (2000) und Neufeld (2001). Ein Erfolg derartiger Methoden würde letztlich auch auf theoretische Debatten zurückwirken, da die *Critical Period Hypothesis* in ihrer starken Form um so fragwürdiger erscheinen muß, je größer das Potential solcher Methoden ist.

2.6 Weitere Einflüsse

Die zuletzt genannte Diskussion um eine Erlernbarkeit der L2-Aussprache blendet einen wesentlichen Aspekt des Erwerbs von Fremdsprachen aus. So erscheint es fraglich, ob eine dauerhafte Verbesserung der L2-Aussprache tatsächlich allein auf der Basis *linguistischer* Methoden erfolgen kann: Der Erwerb einer Fremdsprache ist ein Prozeß, bei dem nicht nur alle linguistischen Ebenen von der Phonetik bis zur Pragmatik involviert sind, sondern der darüber hinaus auch durch außerlinguistische Faktoren entscheidend mitgeprägt wird. So sind beispielsweise verschiedene affektive und motivationale – also im weitesten Sinne psychologische – Faktoren auf Seiten des Lernenden stets als wichtige Einflußgrößen zu berücksichtigen. Krashen (1981, Kap. 2) nennt hier zusammenfassend vor allem:

- **Wille zur Integration.** Das Bestreben und die Motivation des Lernenden, sich in die Sprachgemeinschaft der zu erlernenden Sprache zu integrieren;

- **Utilitaristische Motivation.** Das Bestreben, die zu erlernende Fremdsprache aus praktisch-ökonomischen Gründen zumindest so weit zu erlernen, daß bestimmte Ziele in der Sprachgemeinschaft verfolgt und erreicht werden können;
- **Persönlichkeitsmerkmale.** Bestimmte Persönlichkeitsmerkmale scheinen positiv zu korrelieren mit einer hohen Erfolgsrate beim Erwerb einer L2. Hierzu zählen u.a. Merkmale wie Selbstvertrauen – als Oberbegriff von Angstfreiheit, Extravertiertheit (cf. Busch 1982) und hohem Selbstwertgefühl (cf. Heyde 1979) – sowie die generelle Fähigkeit zur Empathie.

Auch in verschiedenen experimentellen Studien wurden derartige Aspekte aufgegriffen und der spezifische Einfluß ausgewählter psychologischer Variablen auf die L2-Aussprache untersucht. Guiora et al. (1972) untersuchten beispielsweise den Einfluß der *Empathie* auf die phonetische Performanz in einer Zielsprache. Durch eine künstliche experimentelle Erhöhung des Empathie-Niveaus konnte eine tendenzielle Verbesserung der Performanz in Richtung einer weniger akzentbehafteten Aussprache erreicht werden. Den Einfluß des *Angstniveaus* sowie des *Grades der Formalität* der aktuellen Sprechsituation auf die L2-Performanz untersuchte Stølen (1987) in einer begrenzten Studie (mit nur einer Versuchsperson). Dabei wurden (a) die syntaktisch-grammatischen und (b) die phonologischen Fähigkeiten einer einzelnen Sprecherin der Fremdsprache Dänisch in Sprechsituationen untersucht, die durch verschiedene Grade von Formalität gekennzeichnet waren. Während die syntaktischen Fähigkeiten in stark formalen Situationen auf einem hohen Niveau verblieben, wirkte sich das hohe Erregungsniveau negativ auf einzelne phonetisch-phonologische Fertigkeiten (d.h. die Aussprache) der Sprecherin aus.

Wenn sich auch aus den bisherigen experimentellen Arbeiten noch kein eindeutiges Bild des Einflusses spezifischer psychologischer Merkmale auf die L2-Aussprache ergibt, so kann man dennoch einen anderen, allgemeineren Schluß aus ihnen ziehen: Die Studien zeigen, daß keineswegs von einer *intra-individuell stabilen* L2-Performanz als Normalfall ausgegangen werden kann, sondern daß diese stets gewissen situativ bedingten Schwankungen und Alternationen unterworfen ist. Hierzu gehören einerseits Einflüsse durch variierende Persönlichkeitsdispositionen des Sprechers, andererseits aber auch Variablen wie die aktuelle Redesituation oder die kommunikative Absicht in einem zielorientierten Dialog. Auch ist es denkbar, daß der jeweilige *Kommunikationspartner* und seine Kompetenz der betreffenden Zielsprache einen Einfluß auf die L2-Performanz des Sprechers ausübt: Handelt es sich um einen muttersprachlichen Sprecher der Zielsprache, so wird der Lernende möglicherweise größere Aufmerksamkeit auf die korrekte Produktion seiner Äußerungen legen als im Fall der Kommunikation mit einem anderen Nicht-Muttersprachler, bei der die Zielsprache lediglich als gemeinsame Verständigungsbasis dient (diese situative Angleichung an den Part-

ner kann freilich nur so weit gehen, wie die L2-Produktion der bewußten Kontrolle des Sprechers unterliegt).

Insofern ist Dickerson (1975) zuzustimmen, wenn sie (auf der empirischen Basis einer Langzeitstudie über japanische Sprecher mit der Zielsprache Englisch) die Auffassung vertritt, es handle sich bei Lernersprachen prinzipiell um *variable Systeme*, die sich in Abhängigkeit vom Kontext der Sprachverwendung (z.B. freies Sprechen vs. Ablesen von Texten; monologische vs. dialogische Situation etc.) verändern können. Intra-individuelle Variation ist demnach keinesfalls als ein bloßer Störfaktor beim Auffinden möglicher Regularitäten von Lernersprachen zu betrachten, sondern als ihr fester Bestandteil. So weist auch Beebe (1987: 173) darauf hin, daß die in der sprachwissenschaftlicher L2-Forschung verbreitete Ansicht, Lernersprachen seien nicht in gleichem Maße einer soziolinguistischen Variation unterworfen wie natürliche Sprachen, ein Irrglaube sei; tatsächlich seien sie in hohem Maße situativ variabel, und diese Variabilität beginne darüber hinaus bereits in sehr frühen Lernstadien.

Eine forschungspraktische Konsequenz dieses Umstands ist darin zu sehen, daß jegliche Untersuchung von Lernersprachen erschwert wird, da man stets mit einem komplexen Wechselverhältnis zwischen einer internen Systematizität bzw. Regelmäßigkeit der Sprachproduktion einerseits und einer durch situative und motivationale Faktoren bedingten Variabilität andererseits zu rechnen hat. Dies würde u.a. bedeuten, daß, bevor Systematizität in einem bestimmten Aspekt der Lernersprachen überhaupt konstatiert werden kann, zu überprüfen ist, ob und wie weit diese Systematizität *intra-individuell konstant* ist. In der Forschungspraxis läßt sich das Einflußpotential intra-individueller Variation jedoch nicht immer im gewünschten Maße kontrollieren oder a priori zuverlässig abschätzen. Jedoch sollte dort, wo versucht wird, eine Systematik der Lernersprachen zu entwerfen, die Existenz eines solchen Einflusses zumindest bewußt bleiben.

2.7 Implikationen für die vorliegende Arbeit

2.7.1 Systematizität der L2-Produktion

Eine einfache, aber für den vorliegenden Zusammenhang zentrale Quintessenz, die man aus der bis hierher skizzierten Forschung zum Zweitsprachenerwerb ziehen kann, ist der Sachverhalt, daß nicht-muttersprachliche Aussprachevarianten, obwohl sie sich oftmals durchaus auf spezifische linguistische Ursachen zurückführen lassen, in ihrer genauen Beschaffenheit *nicht exakt prädizierbar* sind. Keines der vorgestellten Modelle ist imstande, eine genaue Vorhersage über die

Aussprachevarianten oder -fehler zu treffen, welche von einem spezifischen Sprecher in einer spezifischen Situation produziert werden.

Während die frühen Theorien und Studien im Rahmen der „starken Kontrastivhypothese“ noch davon ausgingen, daß eine Gegenüberstellung zweier Phonemsysteme Aussprachefehler nicht nur erklären, sondern sogar präzisieren kann, wurde diese Position allmählich aufgegeben und spielt heute als Leittheorie der L2-Sprachproduktion keine Rolle mehr. Ein wesentliches Versäumnis des kontrastiven Ansatzes – das sich jedoch tendenziell auch in späteren Modellen fortsetzte – kann darin gesehen werden, daß eine Differenzierung zwischen der Ebene des *Sprachsystems* und des individuellen *Sprechaktes* nicht oder nicht in ausreichendem Maße vorgenommen wurde: Sprecherverhalten kann nicht unmittelbar aus der linguistischen Analyse des Sprachsystems abgeleitet werden. Zwar kann die Ermittlung von strukturellen Kontrasten zwischen L1 und L2 durchaus ein sinnvolles Hilfsmittel sein, um Produktionsfehler (z.B. Aussprachefehler) von Lernenden vorherzusagen. Es ist dabei aber stets zu bedenken, daß damit lediglich eine Prognose *möglicher* oder *wahrscheinlicher* Aussprachefehler, keinesfalls aber die Ermittlung notwendiger und hinreichender Bedingungen für das Zustandekommen spezifischer Fehler ermöglicht wird. *Potentielle* und *tatsächlich auftretende* Interferenz sind also klar voneinander zu trennen (cf. hierzu auch Rein 1983: 94).

Sowohl das Modell der *Interlanguage* als auch das der *approximativen Systeme* markieren hier einen wichtigen Übergang innerhalb der Entwicklung der Zweitsprachenerwerbs-Forschung. Dieser ist erstens gekennzeichnet durch die Anerkennung der Lernaltersprachen als eigenständige Systeme, die nicht allein aus L1 oder L2 ableitbar sind, und zweitens durch die wichtige Einsicht, daß für das Zustandekommen von Fehlern des Lernenden ein ganzes Ursachengefüge in Frage kommt, welches weit über die rein linguistische System-Interferenz hinausweist. Dennoch fällt auf, daß auch diese Modelle in weiten Teilen an der Vorstellung festhalten, daß die Lernaltersprachen letztlich als kohärente *Sprachsysteme* zu betrachten seien. Die Systemhaftigkeit des Untersuchungsobjektes, die für den Sprachforscher augenfällig wird, ist aber für sich genommen noch kein Beweis dafür, daß der Sprecher tatsächlich auf der Basis eines solchen Systems operiert, wenn er sprachliche Äußerungen in einer Fremdsprache produziert. Wenn beispielsweise Nemser (1971a: 117) von einer “systematic nature of the stages of foreign language acquisition” spricht oder Selinker (1972: 33) eine “latent language structure” in die unmittelbare Nähe einer “latent psychological structure” stellt, so ist stets zu bedenken, daß hier oftmals ein relativ unmittelbares und möglicherweise problematisches Verhältnis von sprachlichen Strukturen einerseits und den dem sprachlichen Handeln zugrundeliegenden psychologischen Vorgängen andererseits angenommen wird.

Das Verhältnis zwischen individuellem Sprechakt und einer (möglichen) Systemhaftigkeit der Sprachproduktion innerhalb einer spezifischen Sprechergruppe bleibt auch in dieser Arbeit stets als Grundproblem erhalten: Eine Systematizität und Stabilität der Lernersprachen, die über den individuellen, konkreten Sprechakt hinausweist, wird dabei zunächst unterstellt, denn dies ist eine notwendige Arbeitshypothese und Grundbedingung der entwickelten Methode: Eine regelbasierte Modellierung von linguistischen Merkmalen erscheint nur dann sinnvoll, wenn innerhalb der Sprechergruppe, deren Verhalten modelliert werden soll, interindividuell konstante Merkmale identifiziert werden können.

Statt nun aber die Frage nach der Systemhaftigkeit ausschließlich theoretisch zu erörtern und auf dieser Basis zu einer Entscheidung zu gelangen, ob eine Modellierung der betreffenden Aussprachevarianten überhaupt zweckmäßig, machbar oder linguistisch adäquat ist, wird im weiteren Verlauf dieser Arbeit ein anderer Weg verfolgt: Die regelbasierte Modellierung wird entwickelt, ohne den Versuch zu unternehmen, die Frage schon im Vorfeld umfassend zu beantworten. Statt dessen wird die entwickelte Methode abschließend einer Evaluierung unterzogen, wobei vor allem ermittelt wird, ob und in welchem Maße tatsächliches Sprecherverhalten damit adäquat modelliert werden kann. Diese Evaluierung soll vor allem überprüfen, ob und wie weit einzelne Merkmale akzentbehafteter Aussprache sprecherübergreifend auftreten. Damit gibt sie aber zugleich Auskunft darüber, in welchem Maße die untersuchten *Interlanguages* eine innere Systematizität aufweisen: Denn die regelbasierte Modellierung kann grundsätzlich nur dann erfolgreich sein, wenn eine solche Systematizität zumindest teilweise vorliegt. Die zunächst nur als Arbeitshypothese formulierte Annahme einer Systematizität von Lernersprachen kann somit zumindest in einigen Aspekten verifiziert oder falsifiziert werden.

2.7.2 Direkte Implikationen für die Modellbildung

Einige der in diesem Kapitel skizzierten Modelle sind unmittelbar in das entwickelte Verfahren eingeflossen. So ist die oben dargestellte, von Nemser formulierte Idee einer sequentiellen Stufung approximativer Systeme als Grundlage in das Modell der prototypischen *Akzentstufen* eingegangen, das eine wichtige Basis für das Konzept der regelbasierten Modellierung akzentbehafteter Aussprache bildet. In Kapitel 4.4 wird deutlich werden, in welcher Weise dieses Modell eine geeignete Basis hierfür bildet.

Auch aus Fleges Modell einer „Äquivalenzklassifizierung“ ergeben sich spezifische Hypothesen, die für die vorliegende Arbeit bedeutsam sind. Wenn nämlich die Herausbildung von Kategorien für „neue“ L2-Laute grundsätzlich leichter erreicht werden kann als für „ähnliche“ Laute, so kann man hieraus die Vorhersage treffen, daß auch erfahrene Sprecher einer Fremdsprache Schwierigkeiten

bei der authentischen Produktion „ähnlicher“ Laute haben. Trotz zunehmend authentischer L2-Aussprache bleiben die „ähnlichen“ Laute als ein dauerhafter Problembereich erhalten. Eine Modellierung von Aussprachefehlern dieser Sprechergruppe sollte sich dann darauf konzentrieren, phonetische Äquivalente zu identifizieren, d.h. solche Laute der Muttersprache der Sprecher, die nur sehr geringfügige Abweichungen von entsprechenden zielsprachlichen Lauten aufweisen. Das Modell der Akzentstufen berücksichtigt derartige Aussprachefehler, die auch in den untersuchten Sprachdaten vielfach auftraten. Jedoch wird dabei (entgegen der gängigen Praxis) versucht, das Modell der Äquivalenzklassifizierung und die Vorhersagen der Kontrastivhypothese nicht als diametral entgegengesetzte Positionen zu behandeln, sondern sie komplementär zu betrachten. In das Modell der Akzentstufen gehen daher sowohl Lautersetzungen ein, wie sie im Rahmen der Kontrastivhypothese vorausgesagt wurden, als auch solche, die dem Modell der Äquivalenzklassifizierung entsprechen.

Es erscheint für den vorliegenden Zusammenhang generell nicht zweckmäßig, sich ausschließlich an einer der „großen Hypothesen“, wie sie in diesem Kapitel erörtert wurden, als Leittheorie zu orientieren. Es soll statt dessen der Versuch unternommen werden, aus dem derzeitigen Stand des Wissens über die Genese und die phonetische Beschaffenheit fremdsprachlicher Akzente diejenigen Aspekte zu extrahieren, die sich für eine regelbasierte Modellierung als wesentlich und zweckmäßig erweisen. Zur Beantwortung darüber hinausgehender theoretischer Grundfragen können die in dieser Arbeit erhobenen Sprachdaten zwar durchaus herangezogen werden, jedoch liegt hierin nicht das primäre Ziel der vorliegenden Studie. Das folgende Kapitel vollzieht daher einen Perspektivenwechsel, bei dem das Thema fremdsprachlicher Akzente nun unter dem spezifischen Gesichtspunkt der *Sprachtechnologie* erörtert wird.

3 Varietäten und Akzente in der Sprachtechnologie

Die Motivation zu dieser Arbeit ergibt sich aus aktuellen Forschungsfragen im Bereich der maschinellen Sprachverarbeitung. Letztere hat unter anderem zum Ziel, in Schriftform vorliegenden Text akustisch auszugeben (automatische Sprachsynthese, *Text-to-Speech*, *TTS*) oder gesprochene Sprache zu erkennen und in einen anderen Code umzusetzen (automatische Spracherkennung, *Automatic Speech Recognition*, *ASR*). Obwohl Synthese und Erkennung unterschiedliche Ziele verfolgen, ist das Verstehen, Strukturieren und Systematisieren der Variabilität natürlicher Sprechereignisse unabdingbare Voraussetzung für eine erfolgreiche Systemgestaltung. So hängt die Leistungsfähigkeit beider Systemtypen entscheidend davon ab, wie gut es gelingt, lautsprachliche Äußerungen in natürlichen Sprech- bzw. Hörsituationen für den technischen Generierungs- bzw. Erkennungsvorgang zu modellieren.

Im folgenden Kapitel werden exemplarisch Arbeiten skizziert, die sich (a) allgemein mit der Modellierung von Aussprachevariation in der automatischen Spracherkennung und Sprachsynthese sowie (b) speziell mit einer Modellierung fremdsprachlich akzentgefärbter Varianten in der Sprachtechnologie (vor allem in der automatischen Spracherkennung) befassen. Dabei stehen weniger die technischen Details der jeweils verfolgten Adaptionsmethoden im Vordergrund, sondern vielmehr die grundlegenden Verfahren, welche hier zum Einsatz kommen. Von besonderem Interesse sind hierbei Methoden der sog. *lexikalischen Adaption*, da diese auch in dieser Arbeit verwendet werden. Ihre Grundlagen werden daher detaillierter dargestellt. Die in Kap. 7 dargelegte Modellierungstechnik greift auf diese Grundlagen zurück.

Weiterhin wird in diesem Kapitel – in der Form eines Exkurses – dargestellt, welche linguistischen Besonderheiten bei *Eigennamen* auftreten, wie sie im Rahmen dieser Arbeit als primäres sprachliches Untersuchungsmaterial herangezogen werden. Einige dieser Besonderheiten erfordern eine spezielle Behandlung auch im Rahmen sprachtechnologischer Forschungs- und Entwicklungsarbeit. Es werden daher auch Studien dargestellt, die sich in einem sprachtechnologischen Kontext mit dieser speziellen Wortklasse auseinandergesetzt haben.

3.1 Automatische Spracherkennung

3.1.1 Grundprobleme

Automatische Spracherkennung (*Automatic Speech Recognition*, ASR) – also die automatische Überführung gesprochener Sprache in einen Text, welcher das Gesprochene wiedergibt – ist ein sprachtechnologisches Anwendungsfeld, in dem das Phänomen der Aussprachevariation von grundlegendem Interesse ist. Abstrahiert man von allen technischen Einzelproblemen, die in der aktuellen Forschung zur automatischen Spracherkennung bearbeitet werden, so kann man sagen, daß die Variabilität menschlicher Lautsprache schlechthin *das* Kernproblem der automatischen Spracherkennung bildet: Würde es sich nämlich bei der lautlichen Realisierung einer Äußerung durch einen Sprecher um eine stets gleichförmige und präzifizierbare Aktualisierung einer Phonemsequenz handeln, so wäre das Problem der automatischen Spracherkennung bereits mittels eines einfachen Mustervergleichs zu lösen.

Tatsächlich aber stellt sich die Situation anders dar. Trotz erheblicher technischer Fortschritte in den vergangenen Jahren sind viele Teilprobleme keineswegs vollständig gelöst. Einige Beispiele seien hier genannt:

- **Vokabulargröße und kontinuierliche Sprache.** Während die Erkennung isolierter Wörter eines relativ begrenzten Gesamtvokabulars weitgehend zuverlässig erreicht werden kann, sinkt mit zunehmender Vokabulargröße und dem Übergang zu kontinuierlicher Sprache (d.h. vollständige Äußerungen im Satzkontext anstelle von Einzelwörtern) die Erkennungsleistung deutlich ab (cf. Rabiner 1999; Lippmann 1997). Für eine Erkennung unrestringierter „natürlicher Sprache“ – und damit auch für eine breitere Akzeptanz der Technologie – muß das langfristige Ziel darin bestehen, derartige Beschränkungen bezüglich des Vokabularumfangs und der Sprechweise aufzuheben.
- **Sprecherunabhängigkeit.** Während sog. *sprecherabhängige* Systeme, welche an die Sprechweise eines einzelnen Sprechers optimal adaptiert sind, oft eine zufriedenstellende Erkennungsleistung erzielen (etwa in kommerziell erhältlichen automatischen Diktiersystemen), ist die Leistungsfähigkeit von *sprecherunabhängigen* Systemen – von Spracherkennern also, deren Nutzer nicht a priori bekannt sind und die somit für eine prinzipiell unbegrenzte Gruppe von Sprechern konzipiert sind – deutlich eingeschränkt. Hier wird die *interindividuelle* Variabilität der Aussprache als Einflußgröße wirksam, und es gilt daher, Systeme möglichst robust gegenüber derartigen Einflüssen zu gestalten.
- **Sprechstile.** Während die Erkennung bei „formeller“ Sprechweise (z.B. Rundfunknachrichten) bzw. gelesener Sprache auch bei verhältnismäßig großen

Vokabularien bereits mit guten Ergebnissen realisierbar ist, bildet sog. „Spontansprache“, d.h. informelle Sprechweisen im weitesten Sinne, noch ein erhebliches Problem für die automatische Erkennung (cf. Furui 2001).

Bei Strik & Cucchiarini (1999) wird der Versuch unternommen, die im Rahmen sprachtechnologischer Fragestellungen und Problemen auftretenden Formen sprachlicher Variation bzw. Aussprachevariation zu klassifizieren. Aus dem Text läßt sich folgende Taxonomie ableiten:

Variationstyp	Beispiele
(1) Intraspeaker variation	
a) stylistic variation	<i>Sprechgeschwindigkeit, Deutlichkeit</i>
b) free variation	
(2) Interspeaker variation	
a) social/group-specific	
- <i>regional</i>	<i>Dialekt, regionale Besonderheiten</i>
- <i>social (class, group)</i>	<i>Soziolekt, gruppenspezifische Sprache</i>
- <i>mother tongue</i>	<i>Akzent</i>
b) physical/individual-specific	
- <i>sex</i>	<i>Stimme: Grundfrequenz</i>
- <i>age</i>	<i>Stimmqualität</i>
- <i>physiognomy</i>	<i>Stimmqualität</i>

In der vorliegenden Arbeit werden nicht alle der hier genannten Quellen von Variation berücksichtigt. So werden Parameter wie Sprechgeschwindigkeit oder Stimmqualität nicht untersucht. Dies bedeutet freilich nicht, dass derartige Merkmale grundsätzlich keine untersuchenswerten Variablen akzentbehafteter Aussprache bilden: Beispielsweise ist eine gegenüber Muttersprachlern veränderte Sprechgeschwindigkeit (verbunden mit Pausen, Abbrüchen, Verzögerungen und dergleichen) in akzentbehafteter Sprache durchaus zu erwarten, und auch eine solche Besonderheit kann im Rahmen der automatischen Spracherkennung den Einsatz spezifischer Adaptionstechniken erforderlich machen.

Die unter (2a) genannten Formen der sprecherübergreifenden Variation (*interspeaker variation*) bilden hingegen einen zentralen Bereich der vorliegenden Untersuchung. Vor allem bildet die Frage nach dem Grad der sprecherübergreifenden Variabilität ein zentrales Problem der Deskription und der Modellierung akzentbehafteter Sprache: Im ungünstigsten Falle ist diese Variabilität so ausge-

prägt, daß keinerlei gemeinsame (=sprecherübergreifende) Merkmale identifiziert werden können. Eine Adaption an ganze Sprechergruppen wäre in diesem Fall kaum aussichtsreich, denn es könnten lediglich sprecherspezifische, individuelle Eigenarten erfaßt werden. Das Vorhandensein sprecherübergreifender Invarianten bildet also eine Grundbedingung für das hier verfolgte Vorhaben.

Erschwert wird die Situation weiterhin dadurch, daß die in der Taxonomie genannten Quellen von Variation in akzentbehafteter Aussprache auf vielfältige Weise zusammenwirken können. Daß die unter (2a) genannte Muttersprache des Sprechers einen wie auch immer gearteten Einfluß auf die phonetische Produktion in einer Fremdsprache ausübt, erscheint trotz aller Kritik an der Kontrastivhypothese grundsätzlich plausibel. Weitaus schwieriger zu beantworten ist jedoch die Frage, *in welcher Weise* dieser Einfluß verläuft und *wie stark* er ist. Die im vorausgehenden Kapitel skizzierten Arbeiten zur kontrastiven Phonetik und ihre Kritik zeigen dies deutlich.

Weniger offenkundig ist auch, in welcher Weise Variablen wie *Alter*, *Zugehörigkeit zu einer sozialen Gruppe* oder auch der *dialektale Hintergrund* des Sprechers als Einflußgrößen für die phonetische Performanz eines Sprechers wirksam werden. Beispielsweise finden sich in den Beiträgen in James & Kettemann (1983) Überlegungen und Studien über den Einfluß des dialektalen Hintergrunds der Sprecher auf die phonetische Performanz in einer Fremdsprache. Dabei wird u.a. die Möglichkeit erörtert, daß durch den Dialekt des Sprechers besondere Interferenz-Probleme der L2-Aussprache aufgeworfen werden (Interferenz zwischen dialektalen Lauten und L2-Lauten). Ebenso ist aber denkbar, daß derartige Interferenzen aufgrund des dialektalen Hintergrunds des Sprechers gerade *nicht* auftreten, weil der betreffende Dialekt - im Gegensatz zur Hochlautung - Laute aufweist, die mit den zu produzierenden zielsprachlichen Lauten identisch sind. Eine akkurate L2-Aussprache würde in diesem Falle durch den Dialekt des Sprechers nicht behindert, sondern erleichtert. So kann eine auf den ersten Blick wenig relevante Variable wie der dialektale Hintergrund des Sprechers einen deutlichen Einfluß auf die lautliche Beschaffenheit von Akzenten nehmen.

Die Untersuchung akzentbehafteter Aussprache wird zudem dadurch erschwert, daß zwischen den in der Taxonomie genannten Bereichen (1) *intraspeaker variation* und (2) *interspeaker variation* stets eine Wechselwirkung besteht: Wie bereits im vorausgehenden Kapitel dargelegt, verläuft die Produktion von L2-Lauten durch einen Sprecher nicht stets gleichartig, sondern kann in Abhängigkeit von der aktuellen Kommunikationssituation variieren. Zuverlässige Aussagen zur *sprecherübergreifenden* Konstanz innerhalb einer Sprechergruppe setzen daher voraus, daß zunächst ermittelt wird, ob und wie weit die L2-Produktion der untersuchten Sprecher das Kriterium *individueller* Konstanz erfüllt. Damit ist allgemein die Frage nach dem möglicherweise idiolektalen Cha-

rakter fremdsprachlicher Akzentaussprache angesprochen, die im Laufe dieser Arbeit noch mehrfach in verschiedenen Zusammenhängen wiederkehren wird.

3.1.2 Adaptionstechniken

Es ist eine Grundbedingung für einen erfolgreichen Einsatz automatischer Spracherkennung, daß phonetische Variation in allen oben genannten Formen von den betreffenden Systemen toleriert wird. Wie bereits in der Einleitung dieser Arbeit angesprochen, ist ein solcher Toleranzbereich bei heutigen Spracherkennungssystemen durchaus gegeben. Phonetische Variation führt daher zu keiner grundsätzlichen Leistungsbeeinträchtigung dieser Systeme, sofern sie innerhalb eines gewissen Spektrums verbleibt. Für eine Reihe von Sprechstilen – von denen einige oben bereits genannt wurden – gilt dagegen, daß sie dieses Spektrum überschreiten. In diesen Fällen sind Modifikationen eines Spracherkennungssystems erforderlich, die in möglichst spezifischer Weise die Besonderheiten des betreffenden Sprechstils erfassen. Grundsätzlich können zu diesem Zweck verschiedene Methoden herangezogen werden. Im zitierten Text von Strik & Cucchiaroni (1999) wird eine Übersicht und Klassifizierung der bisher erprobten Methoden vorgestellt. Demzufolge können Modifikationen prinzipiell an unterschiedlichen Systemkomponenten vorgenommen werden:

(i) **Adaption auf der Ebene des Lexikons (lexikalisch-phonetische Adaption).** Das phonetische Lexikon des Spracherkenners, in dem die kanonische Aussprache (also die Standardaussprache) der Worteinheiten angegeben ist, wird hierbei um alternative Aussprachevarianten ergänzt. Dies kann in Form von statischen Auflistungen der bekannten Varianten oder einer regelbasierten Ableitung der Varianten aus einer kanonischen phonetischen Oberflächenform oder einer phonemischen Basisrepräsentation erfolgen. Die Varianten wiederum werden (a) *datenbasiert*, d.h. durch Analyse von Sprachkorpora, oder (b) *wissensbasiert*, d.h. auf der Grundlage linguistischen Wissens gewonnen. Dieses Adaptionsverfahren ist dasjenige, welches in der vorliegenden Arbeit zum Einsatz kommt. Es wird daher an späterer Stelle ausführlicher erörtert.

(ii) **Adaption auf der Ebene der akustischen Modelle.** Die akustischen Basiseinheiten eines automatischen Spracherkenners sind in der Regel sog. Phonem-Modelle, d.h. akustische Modelle der lautlichen Basiseinheiten, aus denen sich die zu erkennende Sprachsequenz zusammensetzt. Dabei handelt es sich nicht um *Phoneme* im linguistischen Wortsinn, d.h. kleinste funktional verschiedene (=bedeutungsunterscheidende) Einheiten, sondern vielmehr um Modelle *konkreter Realisierungen* von Phonemen. Jedes dieser Phonem-Modelle deckt dabei einen spezifischen akustischen Merkmalsraum ab.

Gewonnen werden diese Modelle durch ein sog. *Training*, bei dem anhand von umfangreichen Sprachdaten (d.h. digitalisierter Audio-Aufzeichnungen realer Sprecher) die verschiedenen möglichen Realisierungen eines Phonems in einem statistischen Modell zusammengefaßt werden. Daraus folgt, daß die Beschaffenheit der für einen Spracherkenner verwendeten Phonem-Modelle in hohem Maße vom Trainingsmaterial abhängt: Wird beispielsweise zum Training ausschließlich eine Sprechergruppe des Dialekts X herangezogen, so sind die daraus resultierenden Modelle möglicherweise ungeeignet zur Erkennung von Sprache eines anderen Dialekts Y.

Eine Adaption an Sprecher des Dialekts Y kann in diesem Falle durch ein erneutes Training der Modelle ausschließlich mit Sprachdaten des jeweiligen Dialekts erfolgen. Hieraus resultierten spezialisierte, dialektsspezifische Modelle – zwangsläufig jedoch mit einem entsprechend restringierten Einsatzbereich. Eine Variante dieser Methode besteht darin, mehrere akzent- oder dialektsspezifische Phonem-Modelle zu einem einzigen Modell zu bündeln, indem das Neutraining anhand eines möglichst repräsentativen Querschnitts verschiedener Sprechergruppen erfolgt. Die daraus resultierenden Modelle weisen einen expandierten akustischen Merkmalsraum auf und sind damit – zumindest theoretisch – zur Erkennung eines breiteren Spektrums von Dialekten (bzw. allgemeiner: Sprechstilen) geeignet. Ob dies in der Praxis die bevorzugte Methode ist, hängt u.a. vom geplanten Einsatzbereich des Erkenners ab, sowie auch davon, wie viele voneinander verschiedene Sprechweisen sinnvollerweise zu einem Modell komprimiert werden können (zu dieser Problematik cf. van Compernelle 1999).

(iii) **Adaption auf der Ebene des Sprachmodells.** Das sog. *Sprachmodell* eines Spracherkenners für kontinuierliche Sprache (oft auch als ‚Grammatik‘ bezeichnet) ist formuliert als eine Menge von N-Grammen (d.h. Modelle linearer syntaktischer Kombinationswahrscheinlichkeiten), aus denen die statistische Wahrscheinlichkeit von Satzthesen abgeleitet werden kann. Auch das Sprachmodell wird aus umfangreichen Beständen sprachlicher Daten mittels eines „Trainings“ gewonnen. Jedoch sind es hier – im Gegensatz zum Training der akustischen Modelle – *Textdaten*, welche zur Generierung des Modells herangezogen werden. Auch Sprachmodelle sind in hohem Maße domänenspezifisch: Ein Sprachmodell beispielsweise, das anhand von sorgfältig gelesenen Texten (z.B. Rundfunknachrichten) gewonnen wurde, ist kaum geeignet zur Modellierung und Erkennung spontaner Äußerungen, da diese hinsichtlich der möglichen und wahrscheinlichen Wortsequenzen vollkommen anders beschaffen sind. Auch hier ist daher ggf. ein Neutraining des Modells mit entsprechendem Textmaterial erforderlich, das die lexikalische und syntaktische Spezifik der betreffenden sprachlichen Domäne repräsentiert.

Modifikationen des Sprachmodells werden mitunter auch dann erforderlich, wenn zuvor das Lexikon um Varianten erweitert wurde (siehe oben), da die

Kombinationswahrscheinlichkeiten der Worteinheiten an die Auftretenshäufigkeit einzelner Aussprachevarianten angepaßt werden müssen (cf. Wester et al. 2000). So sind Aussprachevarianten mit ausgeprägten Lautverschleifungen und -elisionen vor allem in spontaner Sprache wahrscheinlich, während derartige Erscheinungen in formeller Sprechweise weniger häufig auftreten.

3.1.3 Spracherkennung und fremdsprachliche Akzente

Fremdsprachliche Akzente sind als eigenständiges Thema erst seit verhältnismäßig kurzer Zeit in den Blickpunkt sprachtechnologischer Forschung gelangt. Ausgangspunkt für dieses wachsende Interesse ist der in vielen Studien bestätigte Sachverhalt, daß die bei Nicht-Muttersprachlern zu beobachtenden Abweichungen von der Normaussprache die Performanz automatischer Spracherkennung erheblich beeinträchtigen (für eine Übersicht cf. van Compernelle 1999).

Auf den ersten Blick könnte die Problemlösung lediglich als ein Spezialfall der Adaption an Dialekte o.ä. aufgefaßt werden, der mit einer oder mehrerer der oben skizzierten Adaptionstechniken zu behandeln wäre. Tatsächlich treten jedoch bei fremdsprachlichen Akzenten eine Reihe zusätzlicher Schwierigkeiten auf: Während intra-linguale Varianten (z.B. Dialekte) oft durch eine relativ systematische Verschiebung der normsprachlichen Lautdistinktionen gekennzeichnet sind, sind fremdsprachliche Akzente lautlich äußerst heterogen. So bilden Nicht-Muttersprachler keinesfalls eine einheitliche Sprechergruppe; sie sind (a) unterschiedlicher muttersprachlicher Herkunft, (b) unterschieden sich stark in ihren Kenntnissen und Fertigkeiten der zielsprachlichen Phonetik/Aussprache und ziehen (c) in unterschiedlichem Maße Drittsprachenkenntnisse zur Aussprache der Zielsprache heran. Aufgrund dieser und anderer Besonderheiten erfordern fremdsprachliche Akzente spezielle Adaptionsmethoden.

Derzeit besteht jedoch noch kein Konsens darüber, welche Adaptionstechniken optimal geeignet sind. Die (relativ wenigen) sprachtechnologisch orientierten Untersuchungen auf diesem Gebiet weisen noch einen deutlich explorativen Charakter auf. Dabei lassen sich die erprobten Methoden prinzipiell nach der oben skizzierten Taxonomie klassifizieren (Adaption auf lexikalischer, akustischer oder Sprachmodell-Ebene). Die bisherigen Ergebnisse lassen noch keine Rückschlüsse auf eine optimale Methode bzw. Kombination von Methoden zu; sie sind z.T. sogar widersprüchlich hinsichtlich der Bewertung der Eignung einzelner Methoden. Einige Beispiele für derartige Studien seien im folgenden skizziert.

Adaption der akustischen Modelle

Ein gängiges Adaptionsverfahren für Spracherkennungssysteme besteht, wie oben dargelegt, in einem Neutraining akustischer Modelle, d.h. einer gezielten Generierung von akustischen Referenzmustern, welche die phonetischen Besonderheiten der gewählten Sprechergruppe möglichst gut repräsentieren. Ein solches Neutraining setzt jedoch grundsätzlich zweierlei voraus: (1) Es werden Sprachdaten (d.h. digitalisierte Aufnahmen realer Sprecher) in erheblichem Umfang benötigt und (2) die Sprechergruppe sollte in ihrem Sprechstil und Akzent repräsentativ für die spätere Benutzergruppe sein. Beide Anforderungen werfen bei einer Adaption an nicht-muttersprachliche Sprecher erhebliche Probleme auf.

Das Problem der erforderlichen großen *Menge* der Sprachdaten, das bereits bei Systemen für muttersprachliche Sprecher auftritt, potenziert sich im Falle von nicht-muttersprachlichen Sprechern um einen kaum mehr überschaubaren Faktor: Soll z.B. ein für das Deutsche ausgelegtes Spracherkennungs-System so angepaßt werden, daß es auch mit englischem Akzent gesprochenes Deutsch zuverlässig erkennt, so kann dies zwar prinzipiell mittels eines Neustrainings mit entsprechenden akzentbehafteten Sprachdaten (d.h. Aufzeichnungen von deutschsprechenden Muttersprachlern des Englischen) erfolgen. Damit ist jedoch aus einer Vielzahl möglicher Akzentgruppen nur *eine* erfaßt. Sollen aufgrund neuer, veränderter Einsatzbedingungen des Programms darüber hinaus auch z.B. französische, italienische und ungarische Sprecher mit ihren jeweils spezifischen Akzenten erfaßt werden, so wäre das Neustraining drei weitere Male zu wiederholen – einschließlich der vorausgehenden Gewinnung von Sprachdaten in großem Umfang. Eine solche vergleichsweise geringfügige Erweiterung des potentiellen Einsatzbereiches des Spracherkennungs-Systems kann also nur mit erheblichem Aufwand erreicht werden. Dieser steigt leicht in kaum mehr vertretbare Dimensionen, wenn weitere Sprechergruppen bzw. Akzente hinzutreten. Das Interesse, ein Neustraining zu umgehen, ist daher aus schon technisch-pragmatischen Gründen hoch. Darüber hinaus ist diese Methode aus sprachwissenschaftlicher Sicht relativ unergiebig, da sie wenig verwertbare Generalisierungen über die behandelten Sprechstile hervorbringt – Generalisierungen, die letztlich auch für eine Weiterentwicklung sprachtechnologischer Systeme nutzbar gemacht werden könnten.

Die Ergebnisse von Studien zum *Neustraining* als Adaptionstechnik für nicht-muttersprachliche Sprecher lassen bisher keine eindeutigen Schlüsse zu. Sie zeigen aber wichtige Tendenzen auf. Teixeira et al. (1997) untersuchten beispielsweise die Erkennungsleistung eines Spracherkennungssystems für die englische Sprache mit Sprechern unterschiedlicher muttersprachlicher Herkunft (Dänisch, Deutsch, Spanisch, Italienisch und Portugiesisch). Zwei Methoden wurden dabei erprobt: Der Einsatz von datenbasiert gewonnenen alternativen phonetischen Transkriptionen (d.h. eine Variante der lexikalisch-phonetischen Adaption) führte zu einer Performanzsteigerung des Spracherkenners. Dagegen konnte durch ein

Neutraining der akustischen Modelle mit nicht-muttersprachlichem Material keine deutliche Verbesserung der Erkenner-Performanz erzielt werden. Für das Neutraining wurden dabei Sprachdaten von Sprechern verschiedener Muttersprachen zugleich verwendet. Die Autoren vermuten, daß die Variabilität innerhalb dieser Trainingsgruppe zu hoch ist, um daraus geeignete akustische Modelle generieren zu können. Die große Bandbreite der Aussprachevarianten im Trainingsmaterial führt demnach zu akustischen Modellen, die zu unspezifisch sind.

Der entgegengesetzte Weg - anstelle eines Neutrainings mit Sprechern *verschiedener* Muttersprachen werden nur Daten von Sprechern *einer* Muttersprache herangezogen - erwies sich dagegen in einer Studie als aussichtsreicher: Uebler & Boros (1999) konnten eine Verbesserung der Erkennung fremdsprachlich akzentgefärbter Sprache mit unterschiedlicher muttersprachlicher Herkunft durch das Neutraining mit Sprachmaterial aus nur *einer* Fremdsprache (Italienisch) erzielen. Dieses Ergebnis kann dahingehend interpretiert werden, daß Akzente verschiedener muttersprachlicher Herkunft eine Reihe lautlicher Gemeinsamkeiten aufweisen, so daß eine Bündelung von gemeinsamen akustischen Modellen für eine Gesamtgruppe der nicht-muttersprachlichen Sprecher möglich erscheint.

Witt & Young (1999) untersuchten eine weitere Variante der Adaption von Phonem-Modellen, bei der ein Neutraining umgangen werden kann. Durch die Kombination vorhandener akustischer Modelle aus beiden Sprachen - d.h. der Muttersprache der Sprecher *und* der Zielsprache - konnte eine Leistungsverbesserung eines Spracherkenners für die englische Sprache bei japanischen Muttersprachlern erzielt werden. Van Leeuwen & Orr (1999) schließlich führten Versuche mit niederländischen Muttersprachlern in der Zielsprache Englisch durch und versuchten dabei das Neutraining der akustischen Modelle zu umgehen, indem lediglich ein Wechsel der Phonemmodelle (niederländische Modelle anstelle amerikanisch-englischer Modelle) vorgenommen wurde. Diese Adaptionstechnik, die später auch mit der umgekehrten Sprachrichtung (Ausgangssprache Englisch, Zielsprache Niederländisch) wiederholt wurde, führte jedoch bei keiner der beiden Sprachrichtungen zu einer Performanzsteigerung (cf. van Leeuwen & Wijngaarden 2000).

Adaption des Sprachmodells

Tomokiyo (2000a, 2000b) untersuchte verschiedene Adaptionstechniken für Muttersprachler des Japanischen mit der Zielsprache Englisch und gelangte u.a. zu dem Ergebnis, daß eine lexikalische Adaption (manuelle Hinzufügung alternativer Aussprachevarianten) keine Verbesserung der Performanz bewirkt. Eine Adaption des Sprachmodells resultierte hingegen in einer höheren Erkennungsleistung.

Tomokiyo untersuchte dabei allerdings vorrangig den Einfluß von Vorlesefehlern und Fehlern in Spontansprache auf die *kontinuierliche* Spracherkennung, d.h. die der Erkennung zusammenhängend gesprochener Sätze. Hieraus läßt sich möglicherweise der relativ starke Einfluß des Sprachmodells auf die Erkennungsleistung erklären: Die in einem solchen Szenario produzierten Fehler der japanischen Sprecher manifestieren sich vor allem in abweichender Vokabularselektion (aufgrund eines gegenüber englischen Muttersprachlern geringeren aktiven Vokabulars) und abweichenden Wortsequenzen, die durch syntaktische Fehler, Auslassungen, Abbrüche oder Pausen verursacht werden (zu diesen Besonderheiten akzentbehafteter Sprache siehe auch Raux & Eskenazi 2004). Es ist in diesem Falle nicht überraschend, wenn eine Adaption auf der Ebene des Sprachmodells sich für diesen spezifischen Sprechstil als die effizienteste erweist, denn derartige Abweichungen können prinzipiell *nur hier* modelliert werden. Somit wird die Übertragbarkeit dieser Ergebnisse auf andere Erkennungsaufgaben (z.B. isoliert gesprochene Wörter) erschwert.

Lexikalisch-phonetische Adaption

Für die vorliegende Untersuchung sind insbesondere Arbeiten zur lexikalisch-phonetischen Adaption relevant, in denen phonologische Prozesse wie Assimilation, Elision, Substitution oder Epenthese in Form von Regeln (sog. *re-write rules*) modelliert werden.

Auf dem Gebiet der automatischen Spracherkennung wurden derartige Verfahren zunächst für verschiedene Arten von innersprachlicher Variation erprobt, d.h. für verschiedene Sprechstile, in denen entsprechende phonetisch-phonologische Prozesse verstärkt auftreten. Angewendet wurde diese Adaptionmethode u.a. von Wester & Kessens (1998) sowie Cremelie & Martens (1995) für das Niederländische; Ferreiros et al. (1998) für das Spanische sowie Flach (1995), Wesenick (1996), Kemp (1996) und Lehtinen & Safra (1998) für das Deutsche. Trotz Unterschieden im Detail ist diesen Arbeiten gemeinsam, daß alternative Aussprachevarianten (z.B. für Spontansprache oder regionale Varietäten) maschinell durch phonologische Regeln generiert werden. Typischerweise bildet die kanonische Lautung (Standardlautung) die Eingabe für die variantengenerierenden Regelsätze. Wester et al. (2000) konnten durch die Modellierung von wortinterner und wortübergreifender Variation mittels nur weniger phono-

logischer Regeln eine Verbesserung der Performanz des verwendeten Spracherkenners erzielen. Ferreiros et al. (1998) erreichten mittels phonologischer Regeln, die einem regulären Graphem-nach-Phonem-Umsetzer für Spanisch nachgeschaltet waren, eine signifikant höhere Erkennungsrate bei kontinuierlicher Erkennung, konnten jedoch für Einzelworterkennung keine Verbesserung feststellen. Flach (1995) modellierte wortübergreifende Koartikulations- und Assimilationserscheinungen bei Spontansprache mittels Regeln und konnte damit gleichfalls eine Verbesserung der Erkennungsleistung erzielen.

Es wurde jedoch bei den zuvor genannten Studien und ihren in der Grundtendenz positiven Ergebnissen zunächst nur *intralingual* auftretende Variation modelliert. Eine Anwendung lexikalischer Adaptionmethoden auf die Domäne fremdsprachlicher Akzente erfolgte in einer Reihe von Pilotstudien der jüngeren Vergangenheit. So konnten Ikeno et al. (2003) durch die lexikalisch-phonetische Modellierung eines charakteristischen Aussprachefehlers spanischer Sprecher im Englischen (fehlende Vokalreduktion) die Performanz eines Spracherkenners verbessern. Livescu & Glass (2000) erreichten durch lexikalische Modellierung mittels einer Phonem-Vertauschungsmatrix ebenfalls eine geringfügige Verbesserung der Erkennungsrate. Bei dieser Methode wurden aus vorliegenden Sprachdaten Wahrscheinlichkeiten abgeleitet, mit denen eine spezifische Phonemverwechslung auftritt. Das Wissen hierüber kann helfen, entsprechende Fehler als wahrscheinliche Aussprachevarianten bereits beim System-Design zu berücksichtigen und somit die Erkennungsleistung zu steigern. Amdal, Korkmazskiy & Surendran (2000) wendeten eine ähnliche Methode an, in der sie die *Assoziationsstärke* zwischen zwei Phonemen ermitteln und aus dieser automatisch phonetische Regeln erstellen lassen. Der Begriff „Assoziationsstärke“ (*association strength*) bezeichnet dabei die Häufigkeit des Eintretens bestimmter nicht-muttersprachlicher Ersatzlaute für bestimmte Phoneme der Standard-Transkription. Angewendet in einem automatischen Spracherkenner konnte auch mit dieser Methode eine Verbesserung der Erkennungsleistung erzielt werden.

3.2 Generierung von Varietäten in der Sprachsynthese

Auch für den Bereich der maschinellen *Sprachsynthese* liegen verschiedene Studien vor, in denen die Generierung sprachlicher Varietäten oder Akzente thematisiert wird (z.B. Miller 1998; Fitt & Isard 1999). Dabei handelt es sich jedoch stets um innersprachliche Varietäten wie z.B. Dialekte oder spezifische Sprechstile. Wissenschaftliche Studien, die sich explizit mit einer Generierung fremdsprachlicher Akzente in der Sprachsynthese beschäftigen, sind hingegen weit

weniger zahlreich. Ein Grund für das vergleichsweise geringere wissenschaftlich-technische Interesse an Akzenten in der Sprachsynthese ist sicher darin zu sehen, daß im Gegensatz zur automatischen Spracherkennung bisher keine unmittelbare Notwendigkeit bestand, vorhandene Systeme beispielsweise mit einer akzentbehafteten synthetischen Stimme auszustatten. Während in der Spracherkennung eine Adaption an möglichst viele verschiedene Sprechstile und phonetische Varietäten langfristig unabdingbar ist, um eine gleichbleibende Performanz solcher Systeme über breite Nutzergruppen hinweg zu gewährleisten, stehen bei der Sprachsynthese vor allem Faktoren wie Verständlichkeit, Natürlichkeit und Akzeptabilität als Qualitätsmaßstab an vorderster Stelle (cf. Gibbon, Moore & Winski 1997: 506f.). Legt man diese Maßstäbe zugrunde, so ist nicht unmittelbar erkennbar, worin der Nutzen einer gezielten automatischen Generierung eines Sprechstils liegen könnte, dessen Verständlichkeit sogar *herabgesetzt* ist, wie es ja bei einem Sprachsynthese-System mit einem „Akzent“ möglicherweise der Fall wäre.

Mittel- oder langfristig dürfte sich das Interesse jedoch durchaus auch auf Aspekte der Modellierung spezifischer Sprecher bzw. Sprechergruppen richten, wobei dialektale oder fremdsprachlich akzentbehaftete Sprache als Komponente sogenannter „personalisierter Stimmen“ (*personalized voices*) eine wichtige Rolle spielen könnte. Mit zunehmendem Einsatz von Sprachsynthesystemen wird eine Modellierung sprecher- bzw. sprechergruppenspezifischer Merkmale wachsendes Interesse auf sich ziehen. Im Zusammenhang damit wird sich in der Folge auch die Frage stellen, ob und wie weit derartige Merkmale möglicherweise die *Akzeptanz* solcher Systeme fördern können. Erste Studien deuten beispielsweise darauf hin, daß nicht-muttersprachliche Nutzer von automatischen Sprachdialogsystemen eine Präferenz für (synthetische oder natürliche) Stimmen haben, deren Sprechstil ihrem eigenen Akzent gleicht und diese einer kanonischen Standard-Aussprache vorziehen (cf. Dahlbäck et al. 2001).

Obwohl aber die aktuellen Forschungsaktivitäten, die explizit auf eine Modellierung von Akzenten für die Sprachsynthese abzielen, weniger ausgeprägt sind als im Bereich der Spracherkennung, kann Sprachsynthese prinzipiell auch von Arbeiten profitieren, die im Rahmen der Spracherkennung durchgeführt wurden. Denn prinzipiell können die zuvor geschilderten Techniken der lexikalischen Adaption von ASR-Systemen mit nur geringfügigen Modifikationen auch für die Sprachsynthese nutzbar gemacht werden: Aussprachelexika bilden eine wichtige Komponente beider Systemtypen und gleichen sich in ihrem Aufbau. Daher können z.B. phonologische Regelsysteme, mit denen Varianten für die automatische Spracherkennung modelliert werden sollen, in wenig veränderter Form auch für die Sprachsynthese angewandt werden. Die Aufgabe einer lexikalischen Modellierung von Akzenten für die Sprachsynthese ist sogar grundsätzlich einfacher zu lösen als für die Spracherkennung, da in der Sprachsynthese die Entscheidung

über Art und Stärke des zu generierenden Akzents a priori vom Entwickler getroffen werden kann, wohingegen in der Spracherkennung sämtliche potentiellen Varianten zu berücksichtigen sind.

Zum Zeitpunkt der Entstehung dieser Arbeit sind vor allem zwei Arbeiten zu nennen, die sich der Untersuchung akzentbehafteter Aussprache widmen und dabei explizit die Sprachsynthese als Anwendungsziel nennen. So untersuchte Kim in seiner Dissertation (2000) die Aussprache des Englischen durch koreanische Muttersprachler und entwickelte ein Computerprogramm, das auf der Basis der *Optimalitätstheorie* (Prince & Smolensky 1993) charakteristische Aussprachevarianten koreanischer Sprecher generiert. Dieser Ansatz unterscheidet sich insofern von gängigen Arbeiten mit sprachtechnologischem Hintergrund, als hier im festen Bezugsrahmen einer linguistischen Theorie operiert wird, die spezifische Vorhersagen über die Beschaffenheit der Aussprachevarianten macht. Es handelt sich somit um eine der relativ wenigen Arbeiten, die eine Schnittstelle zwischen linguistischer Theoriebildung und sprachtechnologischer Praxis darstellen.

Einen vorwiegend empirisch ausgerichteten Ansatz verfolgten dagegen Eklund & Lindström (1999, 2001), die umfangreiche Untersuchungen zur Verwendung von fremden Sprachlauten (von den Autoren als „*Xenophones*“ bezeichnet) im Schwedischen durchführten. Einen Schwerpunkt der Untersuchungen (mit mehr als 400 Sprechern des Schwedischen) sind englischsprachige Namen, die durch die Massenmedien in Schweden einen hohen Bekanntheitsgrad erlangt haben (*James Bond, Margaret Thatcher*). Dabei konnte beispielsweise festgestellt werden, daß vollständige Assimilationen der Ziellaute - d.h. eine vollständige Angleichung englischer Phoneme an das schwedische Lautinventar - relativ selten auftreten. Überraschend ist weiterhin der Befund, daß eine Reihe von fremdsprachlichen Lauten von den Sprechern *nicht* produziert wurden, obgleich sie im schwedischen Phonemsystem existieren (Eklund & Lindström 1999). Eine wichtige Implikation ihrer Studien sehen die Autoren darin, daß die Annahme eines homogenen, in sich abgeschlossenen Phonem-Inventars einer Sprache in der Praxis nicht tragfähig ist, da mit zunehmender Internationalisierung stets lautliche Einflüsse aus anderen Sprachen wirksam werden.

Das Phänomen einer solchen „Sprachmischung“ ist dabei nicht allein auf die Ebene der Lautung beschränkt. Sowohl für die Spracherkennung als auch für die Sprachsynthese muß grundsätzlich die Möglichkeit *gemischtsprachiger* Äußerungen in Betracht gezogen werden (etwa im deutschen Satz: „Er *surfte* im *Internet*“). Im Bereich der Sprachsynthese wurde die Möglichkeit solcher fremdsprachiger Inklusionen im Rahmen eines gemischtsprachigen Systems bereits technisch realisiert (cf. Pfister & Romsdorfer 2003). Bezüglich der potentiellen Aussprachevarianten derartiger Inklusionen sind dabei zweifellos noch viele Fra-

gen offen, insbesondere die Frage nach einer möglichen Angleichung von L2-Lauten an das muttersprachliche Laut-Inventar.

3.3 Exkurs: Eigennamen als sprachtechnologisches Problem

3.3.1 Linguistische Besonderheiten von Eigennamen

Innerhalb der Sprachwissenschaft ist eine begriffliche Trennung von appellativischem Wortschatz (*Nomen appellativum*, Gattungsnamen) und Eigennamen (*Nomen proprium*) seit langem etabliert. Diese Sonderstellung des Eigennamens spiegelt sich unter anderem auch in einer disziplinären Spezialisierung der *Onomastik* als demjenigen Teilgebiet der Sprachwissenschaft, welches sich mit der Erforschung, Entstehung und dem Gebrauch von Eigennamen beschäftigt. Die Behandlung von Eigennamen als gesonderte Wortklasse ist ursprünglich vor allem motiviert durch semantische und grammatische Besonderheiten, die jedoch hier nicht diskutiert werden sollen (für eine detaillierte Darstellung cf. z.B. Koß 2002). In onomastischen Studien weitgehend unberücksichtigt geblieben sind dagegen eine Reihe von morphologischen, graphematischen und phonologischen Merkmalen von Eigennamen, die im vorliegenden Zusammenhang als wichtige Einflußgrößen beim Zustandekommen von Aussprachefehlern (bei Nicht-Muttersprachlern, aber auch bei muttersprachlichen Sprechern) in Erscheinung treten. Einige dieser Merkmale werden im folgenden skizziert.

(1) **Morphologische Intransparenz.** Auch für Sprecher mit durchschnittlichen bis guten Kenntnissen der Zielsprache bildet der morphologische Aufbau von Eigennamen häufig ein größeres Problem als der appellativische Wortschatz. Grund dafür ist ihre morphologische Intransparenz, d.h. ihr formaler Aufbau aus morphologischen Konstituenten, die im Lexikon der betreffenden Sprache nicht bzw. nicht mehr als produktive Wortbildungselemente vorkommen. Für diese Konstituenten wurde die Bezeichnung *onomastische Morpheme* oder *onomastische Konstituenten* vorgeschlagen (cf. Fleischer 1992), um der Tatsache Rechnung zu tragen, daß sie ausschließlich als innerhalb von Eigennamen auftreten. Hierzu gehören einerseits sogenannte unikale Elemente wie *Dresden*, *Klaus*, *Teigler*, andererseits aber auch regelmäßig auftretende onomastische Suffixe, wie etwa *-stedt*, *-rode* oder *-ing(en)* bei Ortsnamen oder *-sen*, *-hoff*, *-meyer* bei Familiennamen. Oftmals treten derartige Konstituenten auch in Kombination mit

lexikalischen Morphemen auf, wodurch es zu Mischformen kommt (z.B. *Königsborn*). Eine Klassifizierung der Strukturtypen von Namen im Deutschen findet sich bei Fleischer (1992) sowie – aus der spezifischen Perspektive der automatischen Sprachsynthese – bei Belhoula (1996, Kap. 7).

Im vorliegenden Zusammenhang ist vor allem von Bedeutung, daß die (der Phonemisierung vorausgehende) Analyse von Namen durch solche spezifisch onomastischen Wortbildungselemente für Nicht-Muttersprachler erschwert wird. In Kap. 6.5 werden Beispiele dafür gegeben, in welcher Weise ein solcher morphologischer Aufbau auf die von Sprechern angewandten Strategien des Lesens und der phonetischen Realisierung von Namen einwirkt.

(2) **Irreguläre Graphem-Phonem-Zuordnungen.** In Eigennamen sind oftmals orthographische oder phonetisch-phonologische Merkmale erhalten geblieben, welche in der heutigen Sprachform nicht mehr auftreten. Dabei handelt es sich z.B. um Spuren älterer Sprachstufen bzw. anderer Sprachen, mit denen die Sprache im Laufe ihrer historischen Entwicklung in Kontakt gekommen ist. Als eine Erscheinungsform solcher Relikte findet man in einigen Namen beispielsweise Laut-Buchstaben-Zuordnungen vor, die von den heutigen Regeln der betreffenden Hochlautung abweichen. So weisen beispielsweise regionale Eigenarten wie das sog. „westfälische Dehnungs-e“ (etwa in Ortsnamen wie *Coesfeld*, *Soest*) irreguläre Graphem-Phonem-Zuordnungen auf, die oftmals auch bei Muttersprachlern des Deutschen zu Aussprachefehlern führen ([kø:sfelt], [zø:st]).

Es wurde bereits in einer Reihe von Studien beobachtet, daß dieses Merkmal von Eigennamen sich als ein Problem für sprachtechnologische Anwendungen erweisen kann. So wurde bei der Entwicklung automatischer Graphem-Phonem-Umsetzer – von Programmen also, die eine orthographische Eingabe in eine korrekte phonetische Transkription umwandeln – vielfach festgestellt, daß für die Behandlung von Eigennamen zusätzliche oder spezialisierte Regeln erforderlich sind (cf. Dutoit 1997: 125-26; Belhoula 1996; Gustafson 1994; Jannedy & Möbius 1997; Mengel 1993). Aus denselben Gründen ist es wahrscheinlich, daß die Aussprache einiger L2-Eigennamen für Nicht-Muttersprachler problematisch ist: Die hier auftretenden, von der L2-Norm abweichenden Schriftaussprache-Konventionen sind selbst Sprechern mit guten bis sehr guten Kenntnissen der Zielsprache meist nicht geläufig.

(3) **Exonyme und phonetisch assimilierte Endonyme.** Für eine Reihe von Städte- und Ortsnamen – zumeist Großstädte mit internationaler Bedeutung – existieren in einer oder mehreren Fremdsprachen standardisierte *Exonyme*, d.h. orthographisch und phonetisch assimilierte und lexikalisierte Varianten. Der offiziellen Begriffsbestimmung der Vereinten Nationen zufolge handelt es sich bei einem Exonym um einen

[...] name used in a specific language for a geographical feature situated outside the area where that language has official status, and differing in its form from the name used in the official language or languages of the area where the geographical feature is situated. (Kadmon 1998)

Der üblichen Verwendungsweise des Terminus zufolge handelt es sich dabei um solche Namen, die orthographisch und lautlich vollständig an die aufnehmende Sprache assimiliert sind, etwa frz. *Londres* (\approx engl. *London*), dt. *Mailand* (\approx ital. *Milano*), *Kopenhagen* (\approx dän. *København*) oder frz. *La Haye* (\approx niederl. *s'Gravenhage*). Oftmals sind solche Exonyme auch in regulären Wörterbüchern der aufnehmenden Sprachen zu finden, was als Indiz für eine weitgehend abgeschlossene Lexikalisierung betrachtet werden kann. Zu diesen genuinen Exonymen sind auch Namen zu zählen, deren endonymische Form nur geringfügig an die orthographischen und phonetischen Gegebenheiten der aufnehmenden Sprachen assimiliert wurde, etwa span. *Hamburgo* (\approx dt. *Hamburg*), dt. *Brügge* (\approx fläm. *Brugge*) oder ital. *Vurtemberga* (\approx dt. *Württemberg*).

Im Kontext sprachtechnologischer Anwendungen sind derartige Formen stets als mögliche Aussprachevarianten zu berücksichtigen, wenn fremdsprachige Benutzer vorgesehen sind. So besteht etwa bei einem automatischen Reiseauskunfts-Dialogsystem eine hohe Wahrscheinlichkeit, daß ein nicht-muttersprachlicher Benutzer für die Bezeichnung eines Zielortes ein Exonym verwendet, sofern ein solches existiert. Die Realisierung des deutschen Satzes *Ich möchte nach München fahren* durch einen englischen Muttersprachler kann somit auch durchaus die Form *Ich möchte nach Munich fahren* annehmen. Die phonetische Realisierung [mju:nɪk] ist daher prinzipiell als eine mögliche Aussprachevariante von *München* zu berücksichtigen.

Die oben zitierte Definition des Terminus *Exonym* läßt freilich offen, ob die genannten Abweichungen der Form des betreffenden Namens sich auf die orthographische, die lautliche, oder auf beide Formen beziehen. Denn neben den zuvor genannten genuinen Exonymen existiert eine weitere Gruppe von Ortsnamen, die man als *phonetische Exonyme* bezeichnen kann: Hierbei handelt es sich um Namen, zu denen zwar kein vollständig lexikalisiertes Exonym existiert, deren Aussprache aber dennoch in der aufnehmenden Sprache relativ standardisiert ist. Beispiele hierfür sind engl. [ʃɜ:bʊəg] für frz. *Cherbourg*, deutsch [gø:təbøɔk] für schwed. *Göteborg* oder deutsch [bœɔmɪŋhɐm] für engl. *Birmingham*. Kennzeichnend ist dabei, daß lediglich mehr oder weniger starke phonetische Assimilationen an die aufnehmende Sprache auftreten, dabei jedoch die orthographische Form des Namens unverändert bleibt. Weiterhin ist kennzeichnend, daß die lautliche Assimilation zu einer sprecherübergreifend relativ einheitlichen, stabilen und annähernd konventionalisierten Form führt.

Generell gilt aber, daß unter den zahllosen Toponymen einer Sprache diejenigen Namen, zu denen genuine oder ‚phonetische‘ Exonyme existieren, die Ausnahme bilden. Die Regel bildet die große Zahl verhältnismäßig unbekannter Namen ohne fremdsprachliches Äquivalent. Gerade in diesen Fällen ist zu erwarten, daß die Aussprache durch Nicht-Muttersprachler in hohem Maße variabel und von sprecherspezifischen Einflüssen geprägt sein wird, so daß eine Vielzahl möglicher Aussprachevarianten zu berücksichtigen ist.

(4) **Intralinguale Variation der Aussprache.** Gegenüber dem Standard-Wortschatz einer Sprache, für den eine kodifizierte Hochlautung existiert, die z.B. auch im Fremdsprachenunterricht als Referenzform herangezogen werden kann, weist die Lautung von Eigennamen stets ein gewisses Spektrum *intralingualer Variabilität* auf. Die oben erwähnte Irregularität der Schriftaussprache, wie sie bei einer Reihe von Ortsnamen vorzufinden ist, ist eine mögliche Quelle phonetischer Variation, die bereits innerhalb des betreffenden Sprachraums auftreten kann. Hierbei ist grundsätzlich von der Möglichkeit auszugehen, daß auch ein Muttersprachler die regional korrekte Aussprache nicht zweifelsfrei bestimmen kann. Zu den Namen mit normabweichenden Graphem-Phonem-Zuordnungen tritt weiterhin der Sonderfall *regionaler* Aussprachevarianten. Diese Varianten, die parallel zur Hochlautung des betreffenden Names bestehen, werden nur innerhalb eines spezifischen Dialektraums bzw. auf lokaler Ebene gebraucht und sind überregional nicht bekannt. Beispiele hierfür sind deutsch *Stuttgart* als [ʃtu:ɪgɑ:t] oder engl. *Appletreewick* als [æptɹɪk]. Derartige Fälle zeigen beispielhaft, daß bei der Aussprache von Namen bereits innersprachlich mit Variabilität zu rechnen ist.

(5) **Unübersetzbarkeit von Eigennamen.** Schließlich sei noch auf eine semantisch-pragmatische Besonderheit im sprachenübergreifenden Gebrauch von Eigennamen hingewiesen, welche im vorliegenden Zusammenhang bedeutsam ist: Ortsnamen sind *nicht übersetzbar*. Namen, die über ein Exonym verfügen, können hier zwar als Ausnahmefall betrachtet werden (in gewisser Weise kann frz. *Londres* als „Übersetzung“ von engl. *London* angesehen werden), grundsätzlich jedoch sind Namen nicht in der gleichen Weise in eine Fremdsprache übertragbar, wie es bei appellativischem Wortschatz der Fall ist (siehe hierzu auch die Diskussion aus semantischer Perspektive bei Lyons 1977: 215ff.).

Dies führt zu der Situation, daß Namen innerhalb übersetzter Texte stets als fremdsprachliche Elemente erhalten bleiben („*Lewis* reiste weiter nach *Hempstead*“). Daher kann in bestimmten Zusammenhängen auch ein Sprecher *ohne* jegliche Kenntnisse der Zielsprache in die Situation gelangen, Eigennamen der betreffenden Sprache phonemisieren zu müssen. Ein Beispiel hierfür ist ein Dialog im Rahmen einer Reiseauskunft (z.B. telefonischer Fahrkartenverkauf):

Selbst wenn ein solcher Dialog grundsätzlich in der Muttersprache des Benutzers geführt wird, besteht bei Ortsnamen und anderen Namen die Notwendigkeit, diese in der jeweiligen Zielsprache zu artikulieren. Verfügt der Sprecher in diesen Fällen über keinerlei Kenntnisse der Zielsprache und kann nicht auf ein Exonym zurückgreifen, sind erhebliche Abweichungen von der regulären Aussprache zu erwarten. Dieser Umstand ist z.B. für das Design automatischer Sprachdialogsysteme mit integrierter Spracherkennung relevant: Selbst wenn ein solches System prinzipiell multilingual ist – sich also der Muttersprache des Benutzers anpassen kann (cf. Schultz & Waibel 2001; Zue et al. 1996; Uebler 2000) –, so bleiben Eigennamen aufgrund ihrer Unübersetzbarkeit als fremdsprachliche Residualelemente erhalten. Entsprechend sind hier gravierende Abweichungen von der Norm-Aussprache zu erwarten.

3.3.2 Sprachtechnologische Studien zu Eigennamen

Einige der genannten Spezifika von Eigennamen haben dazu geführt, daß auch im Rahmen sprachtechnologischer Forschung Einzelstudien durchgeführt wurden, die ausschließlich der Wortklasse der Eigennamen gewidmet waren. Hinsichtlich der Aussprache von Eigennamen besteht ein Kernproblem bereits darin, zuverlässige phonetische Referenzformen in großem Umfang verfügbar zu machen. So war es beispielsweise das Ziel des in den 1990er Jahren durchgeführten europäischen Kooperationsprojekts ONOMASTICA, umfangreiche computerlesbare Aussprachelexika ausschließlich für Eigennamen (Personennamen, Ortsnamen, Straßennamen, Firmennamen) für eine Vielzahl europäischer Sprachen zu erstellen (Trancoso & Viana 1995; Gustafson 1994). Im Rahmen dieses Projektes entstanden Aussprachelexika für 11 Sprachen mit einem Umfang von bis zu einer Million Einträgen pro Sprache.

Die zentrale Datensammlung dieses Projekts wird ergänzt durch das sogenannte *Onomastica Interlanguage Pronunciation Lexicon* (Onomastica Consortium 1995). Hierbei handelt es sich um ein Aussprachelexikon, welches pro Sprache eine Untermenge von 1.000 Einträgen (Ortsnamen, Straßennamen und andere Toponyme) enthält. Diese wurden für jede der verbleibenden 10 Sprachen mit sogenannten *Interlanguage*-Aussprachevarianten versehen, d.h. mit potentiellen Varianten, wie sie ein Muttersprachler der jeweiligen Sprache produzieren könnte. Dabei wurde jedem Eintrag pro Muttersprache i.d.R. eine Variante hinzugefügt. Somit enthält das Lexikon insgesamt 110.000 Einträge.

Obwohl die in der vorliegenden Arbeit verfolgte Grundidee dem *Interlanguage Lexicon* durchaus verwandt ist, wird in einem wesentlichen Punkt ein anderer Ansatz verfolgt: Da nicht-muttersprachliche Aussprachevarianten in hohem Maße von sprecherspezifischen Eigenschaften (etwa L2-Kompetenz, Alter, Bildung etc.) geprägt sind, erscheint die Annahme plausibel, daß bereits innerhalb einer Spre-

chergruppe mit gleichem muttersprachlichen Hintergrund die *interindividuelle Variabilität* stark ausgeprägt sein wird. Es erscheint daher fragwürdig, ob nur *eine* potentielle Aussprachevariante pro Wort und Muttersprache der sprachlichen Realität gerecht werden kann. Leider wurden die Daten des ONOMASTICA-Projektes nicht veröffentlicht oder für weitere Forschungen zugänglich gemacht, so daß hier keine Aussagen über die im *Interlanguage Lexicon* enthaltenen Transkriptionen getroffen werden können. Jedoch deuten einige wenige in der Literatur verfügbare Hinweise (von Autoren, die aufgrund eigener Projektbeteiligung Zugang zu den Daten haben) darauf hin, daß die Transkriptionen z.T. artifizielle Aussprachevarianten repräsentieren. So bemerken Trancoso et al. (1999), daß viele der im *Interlanguage Lexicon* enthaltenen Transkriptionen französischer Namen mit „deutschem Akzent“ recht unnatürlich seien und weithin der Ausgabe eines automatischen Graphem-nach-Phonem-Umsetzers gleichen, welcher die Regeln seiner Basissprache in naiver und unmodifizierter Weise auf fremdsprachliche Namen anwendet.

Nicht zuletzt aus diesem Grund wurde für die vorliegende Studie die Entscheidung getroffen, die Größe des untersuchten Vokabulars deutlich zu limitieren und statt dessen einen Schwerpunkt (1) auf eine adäquatere Repräsentation möglicher Teilkenntnisse der zielsprachlichen Aussprache zu legen und (2) potentielle interindividuelle Variation der Aussprache stärker zu berücksichtigen.

Während das *Onomastica Interlanguage Pronunciation Lexicon* lediglich hypothetische Aussprachevarianten enthält, wurde in den Studien von Fitt (1995, 1998) untersucht, welche Aussprachevarianten von Sprechern tatsächlich produziert werden, wenn sie mit der Aufgabe konfrontiert sind, fremdsprachliche Ortsnamen auszusprechen. Untersucht wurde dabei die Aussprache von 70 Ortsnamen aus sechs europäischen Sprachen (Englisch, Französisch, Deutsch, Griechisch, Italienisch, Norwegisch) durch englische Muttersprachler. Hierbei wurden verschiedene Produktionsmodi (gelesen, nachgesprochen) einbezogen und miteinander verglichen. Eine für die vorliegende Studie wichtige Beobachtung Fitts ist die Tatsache, daß auch Sprecher mit keinen oder nur geringen Kenntnissen der Zielsprache nur selten die unmodifizierte (Schrift-) Aussprache ihrer Muttersprache auf zielsprachliche Namen übertragen. Oftmals verfügen die Sprecher zumindest über einige wenige Grundkenntnisse darüber, daß es sich bei bestimmten Graphem-Sequenzen um charakteristische Muster einer L2 handelt, gepaart mit elementarem Wissen über deren Aussprache, z.B. deutsches <sch> mit seiner Aussprache als [ʃ], (cf. Fitt 1998: 104 *passim*). Weiterhin konnte in den Studien ein hoher Einfluß durch Kenntnisse von Drittsprachen beobachtet werden, beispielsweise die Übertragung von deutschen Ausspracheregeln auf Sprachen wie Griechisch oder Norwegisch.

Eine Reihe von weiteren Einzelstudien mit sprachtechnologischem Forschungsinteresse ist der Aussprache von Eigennamen gewidmet. So untersuchten beispielsweise Cremelie & ten Bosch (2001) in einer Studie zur automatischen Erkennung von Personennamen aus verschiedenen Sprachräumen die parallele Verwendung von Graphem-nach-Phonem-Umsetzern für verschiedene Sprachen zur Generierung von alternativen Aussprachevarianten für einen automatischen Spracherkennner. Die Autoren berichten von einer verbesserten Erkennungsrate, die mittels dieser Methode erzielt werden konnte (zu dieser Studie siehe auch Kap. 7.3.4 dieser Arbeit). Eine verwandte Arbeit stellt van den Heuvel (1994) vor, der mittels eines Graphem-nach-Phonem-Umsetzers niederländisch gefärbte („*dutchised*“) Aussprachevarianten für fremdsprachliche Namen generieren ließ. Im Gegensatz zur genannten Studie von Cremelie & ten Bosch, in der die Graphem-nach-Phonem-Regeln verschiedener Sprachen unmodifiziert auf L2-Wortschatz angewendet wurden, wurde hier versucht, eine ‚Nativisierung‘ der Namen, d.h. eine Angleichung an die niederländische Aussprache zu modellieren. Entsprechend dem oben skizzierten Ansatz wird jedoch auch hier lediglich mit einer einzelnen, an das Niederländische adaptierten Aussprachevariante operiert (als „*ideal dutchised pronunciation*“ bezeichnet).

In der oben bereits erwähnten Pilotstudie von Trancoso et al. (1999) wurde die Aussprache deutscher Ortsnamen durch französische Sprecher sowie französischer Ortsnamen durch deutsche Sprecher untersucht. Dabei stützen sich die Autoren auf eigens zu diesem Zweck erhobenes Sprachmaterial von 28 deutschen und 150 französischen Sprechern, um hieraus phonologische Regeln für eine „gemischtsprachige“ Aussprache abzuleiten. Aufgrund der relativ beschränkten Datenbasis seien die erstellten Regeln jedoch, wie die Autoren selbst einräumen, lediglich als ein heuristischer Ausgangspunkt zu betrachten. Auch diese Studie stützt allerdings die weiter oben bereits gegebene Einschätzung, daß Sprecher nur selten die unmodifizierte Aussprache ihrer Muttersprache auf fremdsprachliche Namen anwenden.

3.4 Varietäten und Akzente in der Sprachtechnologie: Zwischenbilanz

Während eine gezielte Generierung von Akzenten oder Dialekten in der *Sprachsynthese* derzeit sicher kein drängendes Forschungsziel ist, ist die Behandlung akzentbehafteter Aussprache im Bereich der automatischen *Spracherkennung* ein Desiderat. Ein bei akzentgefärbter Aussprache zu konstatierender Anstieg der Fehlerrate eines Spracherkennungs-Systems um 600%, wie er in einem Fall be-

richtet wird (Gerosa & Giuliani 2004), ist sicher nicht akzeptabel⁵. Zwar handelt es bei derartigen Ergebnissen um Extremfälle; dennoch illustrieren sie letztlich das grundlegende Problem, das bereits zu Beginn dieser Arbeit umrissen wurde: *Jede stark ausgeprägte Abweichung von der sprachlichen Norm führt zu Leistungseinbußen sprachtechnologischer Systeme*. Extreme Abweichungen führen in einigen Fällen sogar dazu, daß entsprechende Systeme vollständig versagen, so daß an ihren praktischen Einsatz vorläufig nicht zu denken ist.

Ungeachtet möglicher Teilerfolge bleibt als vorläufige Quintessenz der genannten Studien zur automatischen Spracherkennung festzuhalten, daß eine technisch ideale Lösung für das Problem der Erkennung fremdsprachlicher Akzente bisher nicht identifiziert werden konnte. Gründe hierfür sind möglicherweise in den jeweils aufgabenspezifischen Erkennungsszenarien, in unterschiedlichen untersuchten Sprachenkombinationen oder in abweichenden technischen Rahmenbedingungen einzelner Untersuchungen zu suchen, die einen Vergleich und die Generalisierbarkeit der Ergebnisse nur bedingt zulassen.

Ein grundsätzlicheres Problem liegt möglicherweise auf einer anderen Ebene vor. Primär sprachtechnologisch-ingenieurwissenschaftlich ausgerichtete Studien zielen vorrangig auf eine unmittelbare Performanzsteigerung bei der automatischen Erkennung fremdsprachlich akzentgefärbter Sprache ab. Erreicht werden soll dies durch eine weitgehend unmodifizierte Übertragung gebräuchlicher Adaptionstechniken auf diese Domäne, wobei die praktisch-technische Effizienz des gewählten Lösungsansatzes eine primäre Rolle spielt.

Nur selten strebt man dabei explizit eine vorausgehende Analyse der Spezifik fremdsprachlich akzentgefärbter Aussprache an. Lediglich eine kleine Zahl von Untersuchungen zur Spracherkennung und -synthese (Trancoso et al. 1999; Tomokiyo 2000a; Lindström & Eklund 1999, 2001; Fitt 1995, 1998) stellen der Entwicklung und Erprobung von Adaptionstechniken zunächst eine detaillierte linguistische Analyse der auftretenden Varianten voran. Trotz noch fragmentarischer Ergebnisse erweisen sich derartige Studien jedoch als gewinnbringende Vorarbeit für eine Adaption sprachverarbeitender Systeme an die Besonderheiten fremdsprachlich akzentgefärbter Aussprache. So konnte beispielsweise gezeigt werden, daß Grundkenntnisse der zielsprachlichen Aussprache auch bei Sprechern vorhanden sind, die über keine formale Kenntnis der jeweiligen Sprache verfügen (Trancoso et al. 1999; Fitt 1995). So wird ein Muttersprachler des Englischen die deutsche Sequenz <sch> nur selten nach englischen Regeln als [sk] realisieren, sondern zumeist die korrekte deutsche Lautung [ʃ] realisieren. Auffäl-

⁵ Die Autoren der genannten Studie beziffern die Fehlerrate des von ihnen beschriebenen Spracherkenners mit 100%. Dies bedeutet faktisch, daß *keine* sprachliche Äußerung richtig erkannt wurde.

lig ist demnach, daß menschliche Sprecher, auch wenn sie nicht über Expertenwissen der zielsprachigen Phonetik verfügen, keineswegs in völlig „naiver“ Weise mit fremdsprachlichem Material umgehen – naiv im Sinne eines Graphem-nach-Phonem-Umsetzers, der zielsprachlichen Text nach einem rigiden Schema entsprechend den Regeln seiner Ausgangssprache phonemisiert. Vielmehr aktiviert der Sprecher partielle sprachliche Wissensquellen und verknüpft sie schließlich zu einer situativ angemessenen Strategie zur Aussprache der Worteinheiten. Die Leistungsfähigkeit von Adaptionstechniken steigt in dem Maße, wie es gelingt, dieses partielle Wissen explizit zu machen und zu modellieren.

4 Vorüberlegungen zu einer regelbasierten Modellierung

4.1 Grundproblem: Die Bestimmung von Zielformen

Jede regelbasierte oder mit anderen Mitteln realisierte Methode der Modellierung von Aussprache setzt voraus, daß zunächst ermittelt wird, welche lautlichen Charakteristika die zu modellierenden Varianten aufweisen. Dies gilt prinzipiell für *jede* Domäne lautsprachlicher Variation und betrifft gleichermaßen sprecher-spezifische, situationspezifische, stilistische, regionale und dialektale, aber auch fremdsprachlich akzentgefärbte Aussprache. Benötigt wird also stets eine *Zielform*, deren Beschaffenheit zu spezifizieren ist, bevor der Versuch unternommen werden kann, diese Form automatisch zu generieren.

Abhängig von der Domäne der Modellierung ist die Bestimmung dieser Referenzform unterschiedlich schwierig: Weitgehend unproblematisch ist dies für die *kanonische Standardaussprache*, wie sie beispielsweise mit automatischen Graphem-nach-Phonem-Umsetzern generiert werden soll. Hier existiert eine phonetische Referenz in Form einer kodifizierten „Hochlautung“, die in einzelsprachlichen Aussprachewörterbüchern nachzuschlagen ist (z.B. Mangold 1990; Wells 2001a; Warnant 1987) oder inzwischen auch in Form maschinenlesbarer Aussprachelexika verfügbar ist. Die Leistungsfähigkeit von Programmen, die eine solche kanonische Form generieren, läßt sich unmittelbar anhand dieser Referenz messen: Je genauer die Entsprechung zwischen automatisch generierter Form und Referenz ist, desto höher ist die Performanz des Programms.

Ein Problem ist die Bestimmung der angestrebten Zielform bereits im Falle einer Modellierung intra-lingualer (z.B. regionaler, dialektaler) Aussprachevariation. Hier existiert in der Regel keine eindeutig kodifizierte phonetische Basisform, die als Referenz herangezogen werden könnte. Dennoch wird eine Modellierung prinzipiell durch zwei Faktoren begünstigt: Erstens weisen dialektale Aussprachevarianten häufig eine sprecherübergreifend gültige *reguläre Entsprechung* zur Hochlautung auf, welche eine Beschreibung und Derivation z.B. mittels phonologischer Regeln prinzipiell ermöglicht. Zweitens ist das erforderliche linguistisch-phonetische Wissen hier oftmals relativ gut in Form dialektphonologischer Studien erschlossen und dokumentiert, die als Grundlage der Formulierung von Regeln herangezogen werden können. Darüber hinaus liegen vereinzelt auch Sprachdatenbanken für dialektale Varietäten vor, die zu diesem

Zweck genutzt werden können (für das Deutsche z.B. das Korpus *Regional Variants of German*⁶).

Im Falle fremdsprachlich akzentgefärbter Aussprache ist die Bestimmung der Zielform(-en) dagegen deutlich erschwert. Sprecher- und sprechergruppenspezifische Einflüsse (Muttersprache, L2-Kenntnisse, Alter, Bildung, Motivation u.v.m.) führen hier zu der Situation, daß eine Vielzahl möglicher Zielformen vorliegt, die zunächst gleichberechtigt nebeneinander stehen. Im folgenden Kapitel wird daher diskutiert, welche besonderen Schwierigkeiten bei einer Modellierung der hier auftretenden phonetischen Varianten auftreten und welche Methoden angewandt werden können, um dieser Situation gerecht zu werden. Aus dieser Diskussion geht das *Modell prototypischer Akzentstufen* hervor, welches in dieser Arbeit als Grundlage der Beschreibung und Modellierung akzentbehafteter Aussprache gewählt wurde. Es wird zum Ende des Kapitels skizziert. Seine technische Umsetzung im Rahmen der regelbasierten Modellierung von Aussprachevarianten, die sich unmittelbar aus dieser Darstellung ableitet, wird in Kap. 7 beschrieben.

Um die grundlegenden technischen Rahmenbedingungen der Diskussion abzustecken, soll zuvor die Methode der lexikalischen Adaption sprachtechnologischer Systeme, die bereits in Kap. 3.1 angesprochen wurde, in ihren Grundzügen dargestellt werden. Dies ist sinnvoll, da die nachfolgende Diskussion möglicher Modellierungstechniken immer wieder auf dieses grundlegende Verfahren Bezug nimmt. Möglichkeiten und Beschränkungen der im Rahmen dieser Arbeit gewählten Modellierungstechnik sind nicht zuletzt durch grundlegende Eigenschaften dieses Verfahrens bestimmt.

4.2 Das Verfahren der lexikalischen Adaption

Die in dieser Arbeit verwendete Methode bedient sich eines Verfahrens der sog. *lexikalischen Adaption* (eine Übersicht zu diesen Verfahren bietet Strik 2001). Dies bedeutet, daß das entwickelte Adaptions- bzw. Modellierungsverfahren allein auf der Ebene des *Lexikons* operiert. Um Mißverständnissen vorzubeugen, ist dabei vorauszuschicken, daß bezüglich des Begriffs *Lexikon* deutliche Unterschiede zwischen sprachwissenschaftlicher und sprachtechnologischer Terminologie bestehen: Ist im Rahmen sprachtechnologischer Arbeiten, in denen die Lautsprache behandelt wird, vom „Lexikon“ die Rede, sind dabei oftmals keinerlei semantische Aspekte oder Wortbedeutungen involviert, sondern lediglich eine Auflistung des Wortinventars, welches dem System zur Verfügung steht. Ein sol-

⁶ *Bayerisches Archiv für Sprachsignale*; <http://www.phonetik.uni-muenchen.de/Bas>

ches Lexikon ist fester Bestandteil heutiger Spracherkennungs- und Sprachsynthesysteme. In seiner einfachsten Form besteht es aus einer Liste von Worteinträgen in (1) ihrer orthographischen Form sowie (2) einer phonetischen Transkription des Worteintrages. Das folgende Beispiel zeigt einen Ausschnitt aus einem solchen (fiktiven) Lexikon⁷:

Alter	[ˈaltɐ]
Alteration	[altəʁatsjˈo:n]
alterieren	[altəʁˈi:ʁən]
altern	[ˈaltɛn]
alternativ	[altənətˈi:f]
Alternative	[altənətˈi:və]
Alternativlösungen	[altɛnatˈi:flø:zʊŋŋ]
alternieren	[altɛɹnˈi:ʁŋ]
Altersgrenze	[ˈaltɛsgɛntsə]

Dieses Lexikon bildet dabei die Schnittstelle zwischen der symbolischen und der akustischen Ebene der automatischen Sprachverarbeitung, insofern hier die orthographische Codierung des lexikalischen Bestands auf eine Lautfolge abgebildet wird. Letztere ist zwar ebenfalls symbolisch codiert, sie repräsentiert jedoch dabei letztlich eine Sequenz *akustischer* Ereignisse (Sprachlaute). Typischerweise handelt es sich bei der einem Worteintrag zugeordneten phonetischen Transkription um die *kanonische Aussprache*, d.h. die Standard-Aussprache („Hochlautung“) in der betreffenden Sprache. Dialektale, soziolektale oder stilistische Varianten sind daher zunächst nicht Bestandteil des Lexikons.

Genau hier setzen Verfahren der lexikalischen Adaption an. Die Grundidee ist dabei, dem Lexikon Aussprachevarianten hinzuzufügen, die typischerweise in bestimmten Sprechergruppen oder Sprechsituationen auftreten. Sind solche Varianten bekannt, so besteht die Möglichkeit, sie zusätzlich zur kanonischen Aussprache in das Lexikon aufzunehmen. Dabei ist es für das Verfahren selbst zunächst unerheblich, ob es sich um dialektale, sprecherspezifische, stilistische oder andere Varianten handelt. Wesentlich ist allein der Sachverhalt, daß eine Variante mit hinreichender Regelmäßigkeit im Sprachgebrauch auftritt. Ist beispielsweise bekannt, daß für das deutsche Wort *Post* neben seiner kanonischen Form [p^hɔst] die Variante [p^hɔʃt] existiert, so kann diese als alternative Aussprache dem Lexikon hinzugefügt werden. Eine solche Erweiterung wird für alle Einträge des Le-

⁷ Phonetische Transkriptionen sind hier und im weiteren Verlauf dieser Arbeit in der IPA-Lautschrift wiedergegeben. In der Praxis wird statt dessen eine computerlesbare Lautschrift (z.B. SAMPA, ARPABET) verwendet.

xikons durchgeführt, sofern sie über entsprechende Varianten verfügen. Für ein (ebenfalls fiktives) Miniatur-Lexikon mit den drei Einträgen *Post*, *hast* und *meist* würde das daraus resultierende erweiterte Lexikon wie folgt aussehen:

Wort	Kanonisch	Variante 1	Variante 2
<i>Post</i>	[p ^h ɔst]	[p ^h ɔʃt]	...
<i>hast</i>	[hast]	[haʃt]	...
<i>meist</i>	[maɪst]	[maɪʃt]	...

Bei der einfachsten Form lexikalischer Adaption erfolgt die Erweiterung des kanonischen Lexikons *manuell*. Damit ist gemeint, daß die bekannten phonetischen Varianten jedes Worteintrags dem Lexikon hinzugefügt werden. Da jedoch bei umfangreichen Lexika eine solche manuelle Erweiterung sehr arbeitsintensiv und fehleranfällig ist, sind Verfahren sinnvoll, mit denen die Erstellung von Varianten vereinfacht bzw. sogar automatisiert wird. Neben ihrem praktisch-technischen Nutzen bieten derartige Verfahren den Vorteil, daß linguistische Generalisierungen systematisch erfaßt und ggf. auf neues Vokabular übertragen werden können.

Hierzu muß jedoch eine Grundbedingung erfüllt sein: Zwischen der kanonischen Form und den zu erzeugenden Varianten muß eine *systematische* bzw. *reguläre Beziehung* bestehen. Handelt es sich dagegen um vollständig freie Variation ohne erkennbare Regelmäßigkeiten, so können Verfahren zur automatischen Derivation von Varianten nicht sinnvoll angewandt werden. Dieses Problem wird an späterer Stelle in dieser Arbeit noch einmal aufgegriffen. Hier soll zunächst als Ausgangspunkt der Überlegungen angenommen werden, daß eine solche linguistische Regelmäßigkeit konstatiert werden kann. Im genannten Beispiel der Lexikoneinträge *Post*, *hast*, *meist* und ihrer fiktiven Aussprachevarianten kann beispielsweise leicht festgestellt werden, daß zwischen der kanonischen Form und den Varianten eine reguläre Beziehung besteht, die für alle drei Lexikoneinträge gilt: Der Wandel von [s] zu [ʃ] vollzieht sich bei allen Wörtern gleichermaßen, wenn [s] vor dem Laut [t] auftritt. Für dieses Miniatur-Lexikon besteht demnach die prinzipielle Möglichkeit, die Varianten mittels einer allgemeinen Regel aus der kanonischen Form abzuleiten. Formal läßt sich diese Regel wie folgt darstellen:

$$[s] \rightarrow [t] / _ [ʃ]$$

Die Regel besagt, daß der Laut [s] vor dem Laut [t] in den Laut [ʃ] umgewandelt wird. Angewendet auf ein umfangreicheres Lexikon würde diese Regel eine solche Ersetzung für jedes Wort durchführen, in dem dieser Anwendungskontext

erfüllt ist. Auf diese Weise generiert die Regel ein dialekt spezifisches Lexikon für ein phonetisches Einzelphänomen des betreffenden Dialektes.

Unabhängig von den jeweiligen phonetischen Inhalten der Regeln bzw. der Art der zu modellierenden Variation verläuft der Weg von den sprachlichen Daten zu einer regelbasierten Modellierung bzw. lexikalischen Adaption prinzipiell wie folgt:

- (1) **Analyse.** Es werden zunächst *systematische Beziehungen* zwischen der (akzentfreien) Standard-Aussprache einer untersuchten Sprache und verschiedenen akzentbehafteten Varianten identifiziert. Eine wichtige Teilfrage ist dabei, ob sprecherübergreifende Konstanten identifiziert werden können, d.h. ob gleiche oder ähnliche phonetische Realisierungen bei verschiedenen Sprechern vorliegen.
- (2) **Formalisierung.** So weit die Identifizierung solcher Konstanten gelingt, werden diese Beziehungen in einem nächsten Schritt mittels *Ableitungsregeln* formalisiert. Diese werden zu zusammenhängenden Regelsätzen zusammengefaßt, wobei jeder der Regelsätze phonetische Besonderheiten innerhalb einer spezifischen Domäne (Sprechstil, Sprechergruppe usw.) repräsentiert.
- (3) **Programmtechnische Umsetzung.** Die so formalisierten Regeln werden in ein Programm integriert, welches seinerseits eine Komponente z.B. von Spracherkennungs- oder Sprachsynthesystemen bildet

Die Ausgabe des Programms sind potentielle Aussprachevarianten, welche typische phonetische Eigenschaften akzentbehafteter Aussprache aufweisen. Analog zum oben genannten Beispiel des Miniatur-Lexikons für *Post*, *hast* und *meist* soll in der vorliegenden Arbeit auf diese Weise ein Variantenlexikon für nicht-muttersprachliche Aussprache automatisch generiert werden. Ein solches Aussprachelexikon nimmt prinzipiell die folgende Form an (die verwendeten Aussprachevarianten dienen hier nur der Illustration):

Wort	kanonisch	englischer Akzent	französischer Akzent	italienischer Akzent	Akzent XY
<i>Post</i>	[p ^h ɔst]	[p ^h əʊst]	[pɔst]	[postə]	...
<i>hast</i>	[hast]	[hæst]	[ast]	[a:st]	...
<i>meist</i>	[ma:st]	[me:st]	[ma:st]	[ma:i:st]	...

Gelingt es, mittels dieser Methode charakteristische phonetische Eigenarten akzentbehafteter Aussprachevarianten adäquat zu erfassen und zu modellieren, so kann dieses Verfahren auf vielfältige Weise im Bereich der Spracherkennung und Sprachsynthese eingesetzt werden, um bestehende Systeme an neue Sprecher-

bzw. Nutzergruppen zu adaptieren. Im Rahmen der vorliegenden Arbeit wird ein solches Modell nicht nur entwickelt und programmtechnisch umgesetzt; es wird darüber hinaus auch ermittelt und quantifiziert, ob und in welchem Maße mittels dieser Methode das beobachtete Sprecherverhalten adäquat modelliert werden kann (siehe hierzu Kapitel 8).

4.3 Interindividuelle Variabilität als Modellierungsproblem

Wie bereits erwähnt wurde, besteht ein grundlegendes Problem einer regelbasierten Modellierung von Akzentaussprache darin, die mittels der Regeln zu generierenden *Zielformen* hinreichend genau zu definieren. Eine wesentliche Ursache für diese Schwierigkeit liegt in der Heterogenität der Sprechergruppe, deren Sprachverhalten modelliert werden soll.

Vorausgesetzt wird bei allen folgenden Ausführungen zunächst, daß es *den* fremdsprachlichen Akzent – d.h. einen einheitlichen und von der L1 des Sprechers unabhängigen Akzent – nicht gibt: Trotz berechtigter Kritik an der Kontrastivhypothese, wie sie in Kap. 2 dargestellt wurde, kann als sicher angenommen werden, daß die jeweilige Muttersprache des Sprechers Einfluß auf die Beschaffenheit seines Akzents ausübt. Es erscheint daher nicht sinnvoll, Sprecher verschiedener Muttersprachen zu einer Untersuchungseinheit zusammenzufassen und die hier auftretenden phonetischen Varianten zu analysieren. Das phonetische Variationsspektrum wäre in diesem Fall so erheblich, daß sinnvolle Verallgemeinerungen kaum mehr möglich wären. Es wird daher im folgenden stets von *Sprechergruppen gleicher muttersprachlicher Herkunft* als grundlegender Untersuchungseinheit ausgegangen. Das Variationsspektrum reduziert sich damit bereits deutlich, ist aber noch immer ausgeprägt genug, daß eine gesonderte Behandlung im Rahmen sprachtechnologischer Anwendungen erforderlich wird.

4.3.1 Die Situation bei Dialekten

Ein Blick auf dialektphonologische Erscheinungen hilft möglicherweise dabei, die Spezifika fremdsprachlicher Akzentvarianten deutlicher hervortreten zu lassen. Betrachtet man z.B. die folgenden phonetischen Varianten der Äußerung *Where's my coat?* in zwei fiktiven Varietäten A und B des Englischen,

- (1a) [weəz maɪ kəʊt] (Varietät A)
 (1b) [weəz mɪ kəʊt] (Varietät B)

so sind phonetische Gemeinsamkeiten und Differenzen festzustellen. Bei allgemein ähnlicher Lautung der Äußerungen finden sich Differenzen in der Aussprache von *my* und *coat*, die in Varietät A als [maɪ] und [kəʊt] und in Varietät B als [mɪ] und [kəʊt] realisiert werden. Auf phonetisch-segmentaler Ebene ist demnach eine Alternation [aɪ] ~ [ɪ] sowie [əʊ] ~ [aʊ] zu konstatieren. Zu beantworten ist nun die Frage, ob es sich hierbei um eine *regelmäßig auftretende Alternation* oder um eine bloß akzidentelle Erscheinung handelt. Hierzu können zunächst zwei Kriterien herangezogen werden:

(i) **Vokabularübergreifende Gültigkeit.** Die Alternation sollte in ähnlichen linguistischen Kontexten gleichermaßen wirksam werden. Bei der oben skizzierten Alternation [əʊ] ~ [aʊ] kann dann von vokabularübergreifender Gültigkeit gesprochen werden, wenn dieselbe Alternation bei strukturell ähnlichen Wörtern ebenfalls auftritt. Betrachtet man etwa die Äußerungen *Where's my boat?* und *Where's the road?* in denselben Varietäten A und B und beobachtet dabei folgende Realisierungen,

- (2a) [weəz maɪ bəʊt] (Varietät A)
 (2b) [weəz mɪ baʊt] (Varietät B)
 (3a) [weəz θə rəʊd] (Varietät A)
 (3b) [weəz θə raʊd] (Varietät B)

so ist festzustellen, daß hier dieselben Alternationen vorliegen wie in (1a) und (1b). Eine Korrespondenz von [əʊ] ~ [aʊ] scheint demnach (eine ähnliche phonetische Umgebung vorausgesetzt) wiederkehrend aufzutreten, womit ein wesentliches Kriterium für das Vorliegen einer dialektspezifischen Regelmäßigkeit erfüllt ist.

(ii) **Sprecherübergreifende Gültigkeit.** Das zweite wichtige Kriterium für das Vorliegen einer dialektalen „Regel“ liegt in der *interindividuellen Konstanz* phonetischer Varianten. Die beobachtete Variante bzw. Alternation sollte innerhalb einer spezifischen Sprechergruppe mit hoher sprecherübergreifender Regelmäßigkeit auftreten, um sich als dialektspezifische Variante zu qualifizieren. So sollten beispielsweise alle Realisierungen von *Where's my boat?* durch verschiedene Sprecher von Varietät B möglichst durchgängig die phonetische Variante [baʊt] aufweisen, wenn die Alternation [əʊ] ~ [aʊ] die Kriterien eines dialektspezifischen Merkmals erfüllen soll.

Sind beide Bedingungen (i) und (ii) erfüllt, so kann man sagen, daß für die betreffenden Laute eine *reguläre Korrespondenz* (cf. Petyt 1980: 23) zwischen den beobachteten Varietäten A und B vorliegt. Dialektale Varianten erfüllen in der Regel beide Kriterien; sie sind gekennzeichnet durch (i) relativ stabile und sprecherübergreifend konsistente Abweichungen von der Norm-Aussprache (Hochlautung), welche (ii) in ähnlichen phonetischen, graphematischen oder lexikalischen Kontexten regelmäßig auftreten.

Das Vorliegen einer regulären Korrespondenz zwischen Hochlautung und dialektaler Lautung eröffnet nun in einem nächsten Schritt die Möglichkeit, eine *Ableitungsbeziehung* (Derivation) zwischen beiden Varietäten zu etablieren, welche die Korrespondenz regelhaft beschreibt. So kann die oben skizzierte Alternation [əʊ] ~ [aʊ] zwischen Varietät A und Varietät B umformuliert werden in eine Derivationsbeziehung:

Varietät A	Derivationsregel	Varietät B
[kəʊt]	[əʊ] → [aʊ]	[kaʊt]

Auf der Basis einer so formulierten Derivationsregel ist es möglich, die zwischen Varietät A und B beobachtete reguläre Korrespondenz nicht nur in allgemeiner Weise zu formulieren, sondern auch für neue Lexikoneinträge vorherzusagen, die ein spezifisches phonetisches Merkmal aufweisen. Beispielsweise können auf diese Weise bzw. mit derselben Regel die Varianten [baʊt] und [raʊd] (Varietät B) aus den Varianten [bəʊt] und [rəʊd] (Varietät A) abgeleitet werden. Somit verfügt eine Regel der oben beschriebenen Art über ein gewisses prognostisches Potential, denn sie ermöglicht es, für ein Wort, welches in Varietät A den Diphthong [əʊ] enthält, die entsprechende Variante in Varietät B vorherzusagen. Tritt etwa ein neues Wort *load* [ləʊd] hinzu und wird in das Lexikon aufgenommen, so kann mittels der Regel dessen Variante [laʊd] in Varietät B abgeleitet werden.

Arbeiten der *generativen Dialektologie* (cf. z.B. Newton 1972) beruhen prinzipiell auf einem derartigen Verfahren, um regelhaft auftretende Lautvariation in der Domäne regionaler Varietäten von Einzelsprachen zu beschreiben. Dabei wird jedoch – im Gegensatz zur oben skizzierten Methode, bei der Alternationen zwischen phonetischen Oberflächenformen beschrieben wurden – davon ausgegangen, daß die Varietäten eine gemeinsame zugrundeliegende Repräsentation (*underlying representation*, UR) in der Gestalt einer abstrakten phonologischen Basisform aufweisen, aus denen sich beide Varietäten mittels jeweils dialekt-spezifischer phonologischer Regeln ableiten lassen.

4.3.2 Die Situation bei fremdsprachlichen Akzenten

Die hier skizzierte Methode einer Derivation von Aussprachevarianten erscheint prinzipiell auch für die Domäne fremdsprachlicher Akzente geeignet. Jedoch stellt sich das Verhältnis zwischen L2-Hochlautung und den zu erwartenden Aussprachevarianten im Falle fremdsprachlicher Akzente deutlich anders dar als bei Dialekten. Der zentrale Grund hierfür ist, daß die oben genannten Bedingungen für das Vorliegen einer regulären Korrespondenz – vokabularübergreifende und sprecherübergreifende Gültigkeit – nicht in demselben Maße gegeben sind. Während dialektale Varianten oft durch eine relativ systematische Verschiebung der normsprachlichen Lautdistinktionen gekennzeichnet sind, sind fremdsprachliche Akzente lautlich äußerst heterogen. Nicht-Muttersprachler bilden keine einheitliche Sprechergruppe, und auch bei Sprechern mit gleicher Muttersprache tritt interindividuelle Variation in hohem Maße auf. Die innerhalb einer Sprechergruppe mit gleichem muttersprachlichen Hintergrund auftretenden Varianten lassen sich beschreiben als ein Kontinuum, das sich von einer fast fehlerfreien zielsprachlichen Aussprache über verschiedene Zwischenstufen bis hin zu einer extrem stark akzentgefärbten Aussprache erstreckt.

Hinsichtlich des oben skizzierten Verfahrens der lexikalischen Adaption kann daraus bereits eine erste wichtige Folgerung gezogen werden: Eine Erweiterung eines Aussprachelexikons um lediglich *eine* zusätzliche Variante pro Worteinheit ist unzureichend, da sie das auftretende Variationsspektrum nicht abdecken kann. Andererseits erscheint es jedoch ebenfalls nicht praktikabel, bei einer symbolischen Modellierung *jede* potentielle nicht-muttersprachliche Aussprachevariante zu erfassen. Die gesuchte Modellierungstechnik sollte demzufolge weder zu spezifisch sein, so daß sie die auftretende interindividuelle Variation unberücksichtigt läßt, noch sollte sie darauf abzielen, *sämtliche* denkbaren Varianten zu generieren, da in diesem Fall für jede Worteinheit unzählige Varianten gleichermaßen in Frage kämen und das Modell damit seine Vorhersagekraft einbüßen würde.

4.3.3 Finite-State-Modelle

Methoden zur symbolischen Modellierung des gesamten Spektrums phonetischer Variation stehen durchaus zur Verfügung. Sogenannte *Finite-State-Modelle*, die sich in den vergangenen Jahren in Computerlinguistik und Sprachtechnologie etabliert haben (cf. Mohri 1997; Kaplan & Kay 1994), können genau dies leisten und verwenden dabei eine sehr ökonomische Darstellungs- und Beschreibungstechnik. Die folgende Abbildung zeigt die potentiellen Aussprachevarianten für den deutschen Ortsnamen *Aindling*, die in einer Gruppe von nicht-muttersprachlichen Sprechern des Deutschen (L1 Französisch)

sprachlichen Sprechern des Deutschen (L1 Französisch) aufgetreten sind, in Form eines solchen Finite-State-Modells (sog. „endlicher Automat“):

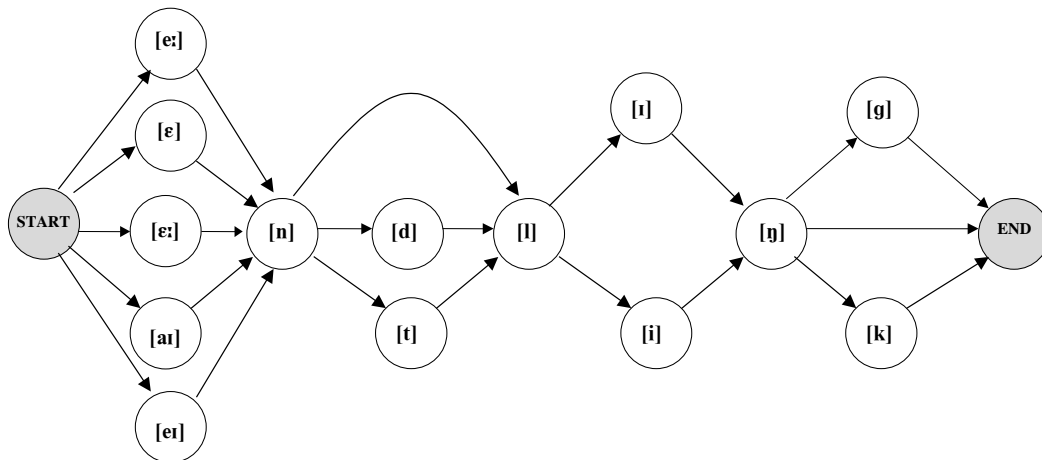


Abb. 4.1: Darstellung von Aussprachevarianten des deutschen Ortsnamens *Aindling* als Finite-State-Netzwerk

Es handelt sich hierbei um ein Übergangsnetzwerk, bei dem die Transitionen zwischen den einzelnen phonetischen Segmenten („Knoten“) mittels gerichteter Graphen („Kanten“) gekennzeichnet sind. Ausgehend von einem Zustand START sind verschiedene Übergänge möglich; jede einzelne Aussprachevariante bildet jeweils einen spezifischen Pfad durch das Netzwerk, welches durch einen Zustand END abgeschlossen wird. Dabei besteht nicht nur die Möglichkeit, daß an einer Position des Netzwerkes verschiedene alternative Segmente vorliegen können (etwa der initiale Vokal/Diphthong von *Aindling*, der durch [e:], [ɛ], [ɛ:], [e:], [aɪ] oder [eɪ] repräsentiert sein kann), sondern es können auch einzelne Segmente des Netzwerkes übersprungen werden, um so die Tilgung von Segmenten zu modellieren (in der Abbildung zum Beispiel der direkte Übergang von [n] zu [l]).

Derartige Netzwerke eignen sich sehr gut zur kompakten Beschreibung sämtlicher potentieller Varianten, d.h. des gesamten phonetischen Variationsspektrums für ein Wort bzw. eine Äußerung. So deckt das oben gezeigte Netzwerk bereits 60 Aussprachevarianten für *Aindling* ab, die hier zur Illustration des Beschreibungspotentials derartiger Finite-State-Modellen gezeigt sind:

(1)	amdɪŋg	(16)	endɪŋg	(31)	ɛ:nɪŋg	(46)	eɪnɪŋg
(2)	amdɪŋk	(17)	endɪŋk	(32)	ɛ:nɪŋk	(47)	eɪnɪŋk
(3)	amdɪŋ	(18)	endɪŋ	(33)	ɛ:nɪŋ	(48)	eɪnɪŋ
(4)	amdɪŋg	(19)	ɛnɪŋg	(34)	ɛ:nɪŋg	(49)	e:ndɪŋg
(5)	amdɪŋk	(20)	ɛnɪŋk	(35)	ɛ:nɪŋk	(50)	e:ndɪŋk
(6)	amdɪŋ	(21)	ɛnɪŋ	(36)	ɛ:nɪŋ	(51)	e:ndɪŋ
(7)	amɪŋg	(22)	ɛnɪŋg	(37)	eɪndɪŋg	(52)	e:ndɪŋg
(8)	amɪŋk	(23)	ɛnɪŋk	(38)	eɪndɪŋk	(53)	e:ndɪŋk
(9)	amɪŋ	(24)	ɛnɪŋ	(39)	eɪndɪŋ	(54)	e:ndɪŋ
(10)	amɪŋg	(25)	ɛ:ndɪŋg	(40)	eɪndɪŋg	(55)	e:nɪŋg
(11)	amɪŋk	(26)	ɛ:ndɪŋk	(41)	eɪndɪŋk	(56)	e:nɪŋk
(12)	amɪŋ	(27)	ɛ:ndɪŋ	(42)	eɪndɪŋ	(57)	e:nɪŋ
(13)	endɪŋg	(28)	ɛ:ndɪŋg	(43)	eɪnɪŋg	(58)	e:nɪŋg
(14)	endɪŋk	(29)	ɛ:ndɪŋk	(44)	eɪnɪŋk	(59)	e:nɪŋk
(15)	endɪŋ	(30)	ɛ:ndɪŋ	(45)	eɪnɪŋ	(60)	e:nɪŋ

Tabelle 4.1: Auflistung der mittels des Finite-State-Netzwerkes in Abb. 4.1 generierten Varianten

Die gezeigte Auflistung ist lediglich eine andere Darstellungsform des oben (4.1) abgebildeten Netzwerkes, bei der jeder alternative Pfad innerhalb des Netzwerkes zu einer vollständigen phonetischen Sequenz expandiert wurde. Bei längeren phonetischen Sequenzen mit einer entsprechend höheren Anzahl von Übergängen und alternativen Segmenten an einzelnen Positionen ergibt sich hieraus leicht eine Zahl von mehreren hundert Aussprachevarianten für ein einzelnes Wort. Mit einer solchen Aufzählung ist die Bandbreite potentieller Varianten zweifellos gut repräsentiert. Hinsichtlich einer *Abdeckung potentieller Varianten* in der Domäne fremdsprachlicher Akzente erscheint dies also auf den ersten Blick als eine attraktive Methode der symbolisch-phonetischen Deskription und Modellierung von Variation.

Dennoch sprechen verschiedene Argumente im vorliegenden Kontext gegen ein solches Verfahren. Zum einen kann eines der positiven Merkmale dieses Verfahrens - sein hohes Beschreibungspotential - zugleich auch ein Defizit sein: Die sehr hohe Anzahl von Varianten steht zunächst gleichberechtigt nebeneinander, ohne daß Informationen zu ihrem tatsächlichen Auftreten bzw. ihrer Auftrenshäufigkeit vorliegen. So enthält das Netzwerk in der oben dargestellten Form keinerlei Angaben darüber, ob einzelne Pfade wahrscheinlicher oder relevanter sind als andere. Dieses Defizit ließe sich jedoch prinzipiell mittels eines sog. *gewichteten* Übergangnetzwerkes beheben, bei dem die einzelnen Übergänge mit einer Übergangswahrscheinlichkeit versehen werden. In einem solchen probabilistischen Netzwerk wären einzelne Sequenzen aufgrund ihrer höheren Wahrscheinlichkeit anderen vorzuziehen.

Abgesehen von der grundsätzlichen Frage jedoch, auf welche Weise und nach welchen Kriterien die Information über die Wahrscheinlichkeit einzelner Übergänge zuverlässig ermittelt werden kann, ist in diesem Modell – auch nach einer Erweiterung durch probabilistische Informationen – ein weiterer Sachverhalt nicht repräsentiert, der kennzeichnend für fremdsprachliche Akzentvarianten ist. Bei Sprechern eines spezifischen L2-Kennntnisstandes ist häufig ein *gemeinsames Auftreten* bestimmter Aussprachefehler zu beobachten, die charakteristisch für die jeweilige Kenntnisstufe sind: So könnte man für das genannte Beispiel von *Aindling* postulieren, daß bei Muttersprachlern des Französischen mit geringen Deutsch-Kenntnissen häufig die Aussprache von <ai> als Monophthong [ɛ] auftritt, da hier eine Schriftaussprache-Regel des Französischen <ai> → [ɛ] zur Anwendung kommt. Zugleich tritt in dieser Sprechergruppe der Fehler auf, daß wortfinales <ng> phonetisch als [ŋk] realisiert wird. Für diese Sprechergruppe ist also ein Gesamtpfad durch das Netzwerk wahrscheinlich, der beide Fehler repräsentiert und z.B. eine Aussprachevariante [endlŋk] generiert.

Für die Modellbildung bedeutet dies, daß zwischen den genannten Aussprachefehlern <ai> → [ɛ] und <ng> → [ŋk] eine Beziehung besteht, die in das Modell eingehen sollte: Da beide Fehler innerhalb einer spezifischen Sprechergruppe (hier: L1 Französisch mit geringen L2 Deutsch-Kenntnissen) häufig gemeinsam auftreten, kann eine *implikative* Beziehung zwischen beiden Fehlern angenommen werden:

$$\langle ai \rangle \rightarrow [\epsilon] \quad \Leftrightarrow \quad \langle ng \rangle \rightarrow [\eta k]$$

Die Implikationsbeziehung (hier durch das Zeichen \Leftrightarrow dargestellt) ist dabei wechselseitig (bidirektional), sie verläuft also sowohl von Regel 1 zu Regel 2 als auch in umgekehrter Richtung, da keine der Regeln einen logisch vorgeordneten Status hat. Innerhalb eines Finite-State-Netzwerkes wäre derselbe Zusammenhang so darzustellen, daß *nicht-lokale Abhängigkeiten* zwischen einzelnen Übergängen bestehen, d.h. Abhängigkeiten zwischen zwei oder mehreren nicht benachbarten Segmenten.

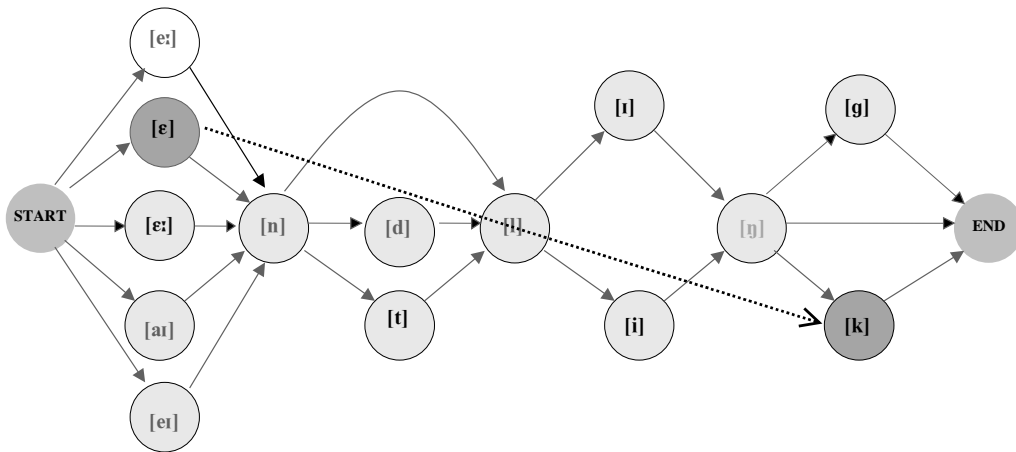


Abb. 4.2: Nicht-lokale implikative Beziehung in einem Finite-State-Netzwerk

Während aber im genannten Beispiel lediglich *eine* derartige Relation angenommen wurde, wird es in der Praxis eine Vielzahl solcher implikativer Relationen innerhalb eines Wortnetzes geben, welche letztlich zu einer höchst komplexen Repräsentation führen. Einer der Vorteile der Methode, nämlich ihre Beschreibungsökonomie, wäre damit jedoch weitgehend außer Kraft gesetzt.

Abgesehen jedoch von der grundsätzlichen Überlegung, ob das Ziel einer Modellbildung darin liegen sollte, lediglich sämtliche existierenden Phänomene aufzuzählen (wie es bei Finite-State-Modellen im Prinzip geschieht), resultiert ein weiterer Einwand gegenüber Finite-State-Modellen aus praktischen Erfordernissen bei der Entwicklung sprachtechnologischer Anwendungen. So ist z.B. bekannt, daß lexikalische Adaptionstechniken bei automatischen Spracherkennern nur bis zu einer bestimmten, meist relativ geringen Anzahl von Varianten pro Wort zu Verbesserungen der Erkennungsleistung führen. Bei Einführung einer höheren Anzahl von Varianten steigt gleichzeitig die Verwechselbarkeit von zu erkennenden Wörtern (cf. Wester & Fosler-Lussier 2000; Beulen et al. 1998; Amdal et al. 2000), so daß die Aufnahme neuer Aussprachevarianten stets mit Bedacht erfolgen und sich möglichst auf wichtige Varianten beschränken sollte. Auch unter diesem Gesichtspunkt erscheint also ein unrestringiertes Finite-State-Modell, wie es oben dargestellt wurde, nicht als optimale Lösung.

Erschwerend kommt hinzu, daß die einzelnen Varianten, welche durch das Modell repräsentiert bzw. generiert werden, zum Teil nur minimale phonetische Differenzen untereinander aufweisen, wie die Auflistung in Tabelle 4.1 zeigt. In modernen Spracherkennungssystemen werden derartige phonetische Differenzen jedoch in der Regel bereits mittels der akustischen Modelle erfaßt, so daß eine explizite Darstellung auf der Ebene des Lexikons redundant wäre.

Statt dessen sollte das gesuchte Modell so beschaffen sein, daß es einerseits wichtige, charakteristische Aussprachefehler innerhalb spezifischer Sprechergruppen berücksichtigt, andererseits aber dabei eine explizite oder implizite Klassifizierung vornimmt, mit der verschiedene Einzelvarianten zu repräsentativen Prototypen zusammengefaßt sind. Ein solches Modell wird im folgenden dargestellt.

4.4 Das Modell prototypischer Akzentstufen

Weiter oben war die Rede davon, daß die innerhalb einer Sprechergruppe mit gleichem muttersprachlichen Hintergrund auftretenden Varianten sich als ein *Kontinuum* beschreiben lassen, das sich von einer fast fehlerfreien zielsprachlichen Aussprache über verschiedene Zwischenstufen bis hin zu einer extrem stark akzentgefärbten Aussprache erstreckt. Der Ansatz, der in dieser Arbeit verfolgt wird, beruht auf der Idee, dieses Kontinuum in diskrete *Akzentstufen* zu unterteilen, mit denen die auf verschiedenen Kompetenzstufen auftretenden Aussprachevarianten beschrieben werden. Jede dieser Akzentstufen repräsentiert dabei *prototypische* Aussprachevarianten.

Überlegungen, die in diese Richtung weisen, wurden bereits in Kapitel 2 vorgestellt. Sie finden sich etwa in der von Nemser (1971a) entworfenen Konzeption von Lerner Sprachen als *approximative Systeme* (siehe hierzu Kap. 2.3). Dabei handelt es sich um vorübergehende, jedoch temporär stabile Sprachsysteme, die sich die Lernenden als Modell der Zielsprache aufgebaut haben. Der Sprachlernprozess läßt sich als eine Abfolge von approximativen Systemen beschreiben, die sich zunehmend dem Sprachsystem L2 annähern. Dabei wird weiterhin angenommen, daß die approximativen Systeme von Lernenden, die auf der gleichen Lernstufe stehen, Ähnlichkeiten aufweisen („*prototypische approximative Systeme*“).

Eine Konkretisierung dieser theoretischen Annahmen im Bereich der Phonetik und Phonologie findet sich bei Ternes (1976), der eine *Stufung lautlicher Interferenz* als Modell für die bei L2-Lernern verschiedener Kompetenzniveaus auftretenden Aussprachefehler vorschlägt. Ternes weist darauf hin, daß es sinnvoll sei,

[...] innerhalb einer für *alle* Lernniveaus bestimmten kontrastiven Beschreibung wenigstens eine gewisse *Abstufung* der Interferenzen vorzunehmen, etwa indem man zwischen *schweren*, *mittelschweren* und *leichten* Fehlern unterscheidet. (Ternes 1976: 11, Hvhbg. von mir)

Hieraus leitet Ternes eine chronologische Abfolge ab, nach der die Korrektur von Aussprache Fehlern im Fremdsprachenunterricht erfolgen kann. Danach seien schwere Aussprachefehler zuerst zu korrigieren, um dann schrittweise zu den leichteren Fehlern voranzugehen und letztere erst beim Erreichen höherer Lernniveaus zu korrigieren.

Ternes illustriert eine solche Interferenzstufung für die Ausgangssprache Deutsch und die Zielsprache Französisch am Beispiel der vier französischen Wörter *beau*, *bon*, *pont* und *bonne*:

	beau /bo/	bon /bõ/	pont /põ/	bonne /bõn/
1. Stufe	[b̥o:]	[b̥o:]	[p ^h o:]	[b̥õn]
2. Stufe	[bo:]	[bõŋ]	[p ^h õŋ]	[bõn]
3. Stufe	[bo:]	[bõ:]	[põ:]	[bõn]
4. Stufe	[bo]	[bõ]	[põ]	[bõn]

Tabelle 4.2: Beispiel einer Interferenzstufung (mit geringfügigen Änderungen übernommen von Ternes 1976: 12)

Die vier Stufen repräsentieren dabei verschiedene Lernerniveaus und die auf diesen Niveaus typischerweise auftretenden Aussprachefehler. Stufe 1 zeigt Formen, die auf einer sehr niedrigen Lernstufe (d.h. von Anfängern) produziert werden; mit ansteigenden Interferenzstufen nähert sich die phonetische Realisierung zunehmend der zielsprachlichen Norm an, welche schließlich in Stufe 4 vollständig erreicht ist. Eine Verwandtschaft zu Nemsers Ansatz der approximativen Systeme ist dabei deutlich erkennbar, insofern es sich bei den Akzentstufen um *vorübergehende* Realisierungsformen der L2-Lautung handelt, welche mit fortschreitendem Lernniveau zugunsten einer akkurateren L2-Aussprache zurücktreten.

Ein ähnlicher Ansatz wurde in der vorliegenden Arbeit nicht nur als deskriptives Instrument zugrunde gelegt, sondern bildet auch die unmittelbare Grundlage für die programmtechnische Implementierung eines Regelsystems zur Generierung von Akzentaussprache. Durch eine sequentielle Anwendung von Regelmodulen können verschiedene Akzentstärken – von leichten Abweichungen von der Standard-Aussprache bis hin zu einer stark fehlerhaften Aussprache – modelliert werden. Die zugrundegelegten Akzentstufen erstrecken sich dabei von einer annähernd muttersprachlichen Aussprache (als unterer Grenze des Variationspektrums) bis hin zu einer stark akzentbehafteten Aussprache (als oberer Grenze, d.h. mit maximaler Abweichung von der Norm-Aussprache).

Derzeit ist das Regelsystem so realisiert, daß für jedes Eingabewort vier prototypische Varianten generiert werden. Im Gegensatz zu der von Ternes verwendeten Konvention, nach der eine niedrige Indexzahl einen besonders stark ausgeprägten Akzent bezeichnet, wurde in dieser Arbeit die umgekehrte Konvention gewählt: Akzentstufen nehmen einen Wert im Bereich zwischen 0 und 4 an. Dabei markiert der Wert 0 die kanonische, vollkommen akzentfreie L2-Aussprache. Die Werte 1 bis 4 kennzeichnen zunehmende Abweichungen von dieser Normaussprache. Die obere Akzentstufe 4 markiert eine maximal abweichende Form, d.h. einen extrem ausgeprägten Akzent. Die folgende Abbildung illustriert dieses Modell:

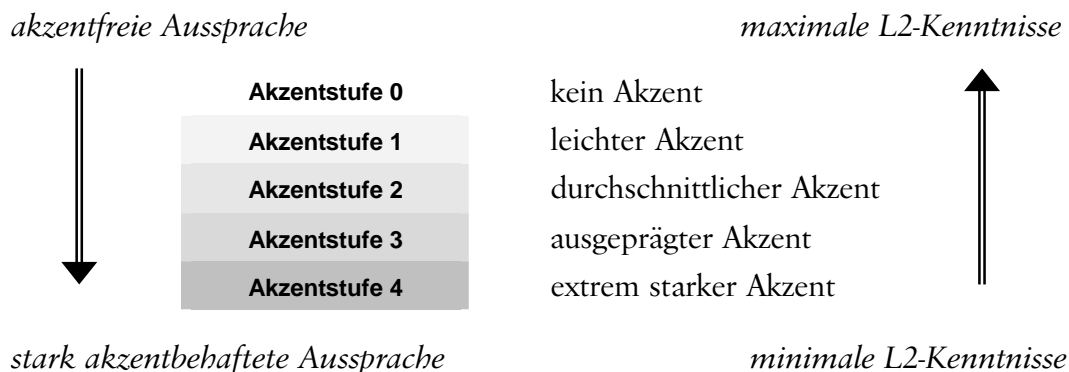


Abb. 4.3: Modell der Akzentstufen

Die hier zugrundegelegte Anzahl von 4 Stufen ist nicht festgeschrieben und kann prinzipiell reduziert oder erhöht werden. Sie hat sich aber als sinnvoll erwiesen, um einerseits die auftretenden Fehlertypen adäquat zu modellieren, dabei aber andererseits die Anzahl potentieller Varianten so zu limitieren, daß hinreichende phonetische Unterschiede zwischen den einzelnen Akzentstufen erkennbar bleiben. Tabelle 4.4 zeigt eine solche Akzentstufung am Beispiel des englischen Ortsnamens *Appleton Thorn*, gesprochen von Muttersprachlern des Deutschen:

engl. <i>Appleton Thorn</i> [æpɫən θɔ:n], L1 Deutsch		
Akzentstufe	Variante	Beschreibung des Fehlers
Stufe 0	[æpɫən θɔ:n]	keiner: kanonische L2-Aussprache
Stufe 1	[ɛpɫən θɔ:n]	phonetische Substitutionen
Stufe 2	[ɛpɫən θɔ:n]	keine Vokalreduktion
Stufe 3	[ɛpɫən sɔ:n]	phonemische Substitutionen
Stufe 4	[apɫən tɔ:n]	Transfer der GTP-Regeln aus L1

Tabelle 4.3: Akzentstufung am Beispiel des engl. Ortsnamens *Appleton Thorn* bei L1 Deutsch

Das Modell der Akzentstufen ist zunächst eine *Beschreibung* und *Typisierung* der auftretenden Aussprachevarianten. Im Vergleich zu der im vorausgehenden Abschnitt dargestellten Modellierung mittels Finite-State-Modellen beinhaltet es eine Reihe von Idealisierungen. So repräsentiert jede der Stufen prototypische, charakteristische Fehler, die bei nicht-muttersprachlichen Sprechern auftreten.

Dies bedeutet freilich nicht, daß die als Prototypen beschriebenen Formen stets in exakt dieser Weise von Sprechern produziert werden. Vielmehr ist es das Ziel des Modells, die Vielfalt der potentiell möglichen Formen auf wenige rele-

vante Fälle zu reduzieren und dabei dennoch wesentliche Fehlertypen zu erfassen. Die Stufen lassen sich phonetisch wie folgt charakterisieren:

- Auf **Stufe 1** werden geringfügige phonetische Substitutionen vorgenommen, welche auch bei Sprechern mit guten Kenntnissen der betreffenden L2 häufig zu beobachten sind, etwa der Ersatz des englischen Vokals [æ] durch deutsches [ɛ]. Dies entspricht im Kern den Fehlertypen, die von Fleges Modell der Äquivalenzklassifizierung (siehe Kap. 2.4.2) vorausgesagt werden und die auch in den Sprachdaten dieser Studie häufig zu beobachten waren.
- **Stufe 2** beinhaltet weitere phonetische Substitutionen, deren Ursache jedoch zum Teil bereits auf einer anderen Ebene zu suchen ist. So ist der Ersatz von engl. [ə] durch [ɔ] in *Appleton* weniger durch eine lautliche Ähnlichkeit zwischen [ə] und [ɔ] motiviert, als vielmehr durch Einflüsse der Schriftaussprache sowie durch eine spezifische phonologische Regel des Englischen (Vokalreduktion), die von deutschen Sprechern oftmals nicht befolgt wird (in Kap. 6.3 erfolgt eine detailliertere Darstellung dieses Fehlers). Mit Stufe 2 wird bereits ein ausgeprägter Akzent erreicht als mit Stufe 1.
- **Stufe 3** schließlich beinhaltet weitere Substitutionen, die ihrer Art nach denjenigen Fehlertypen entsprechen, welche von der Kontrastivhypothese (siehe Kap. 2.2) vorausgesagt wurden: Beinhaltet die zielsprachliche Sequenz einzelne Laute, die in der L1 des Sprechers nicht auftreten, so besteht eine mögliche Strategie darin, den jeweils „nächstliegenden“ (d.h. akustisch bzw. perzeptiv ähnlichsten) Laut des muttersprachlichen Inventars als Ersatzlaut zu produzieren. Hierdurch entsteht beim Hörer (einem Muttersprachler der fraglichen L2) zu meist bereits der Eindruck eines relativ starken Akzents; daher ist eine Modellierung dieses Fehlertyps auf einer hohen Akzentstufe sinnvoll.
- **Stufe 4** schließlich ist eine Stufe, die von nur wenigen Sprechern in dieser Weise realisiert wird. Hier wird die zielsprachliche Sequenz fast vollständig nach den Schriftaussprache-Regeln der Muttersprache des Sprechers realisiert, wodurch mitunter stark von der L2-Norm abweichende Varianten wie z.B. [aplɔtɔn tɔɹn] für *Appleton Thorn* entstehen. Da derartige Varianten aber zumindest dann auftreten können, wenn der Sprecher über keinerlei Kenntnisse der zielsprachlichen Aussprache verfügt (solche Sprecher sind Teil des in Kapitel 5.2 beschriebenen Versuchsszenarios), wurde eine solche Stufe als eine zusätzliche, fakultative Variante in das Modell aufgenommen.

Wesentlich für den Aufbau der Stufung ist weiterhin, daß *höhere Akzentstufen die auf den niedrigeren Akzentstufen modellierten Fehler einschließen*. So sind die geringfügigen phonetischen Abweichungen, die auf Stufe 1 auftreten, stets implizit in den höheren Stufen enthalten. Höhere Stufen modellieren demnach nicht *andere*, sondern *zusätzliche* Fehler. Dies ist nicht nur technisch sinnvoll (im

Sinne eines modularen Aufbaus des Regelsystems), sondern entspricht dabei durchaus der sprachlichen Realität: Ein Sprecher mit einem stark ausgeprägten Akzent wird oftmals alle Fehlertypen zugleich produzieren. Mit zunehmender Vertrautheit mit der Zielsprache werden schrittweise die Fehler der hohen Akzentstufen eliminiert, es verbleiben jedoch die Fehler der niedrigen Stufen. Dies entspricht in Kern der oben skizzierten Akzentstufung, wie sie von Ternes entworfen wurde.

Das Modell der Akzentstufen erscheint nicht nur als Instrument der Beschreibung und Typisierung des auftretenden Variationsspektrums geeignet, sondern kann unmittelbar in ein System der regelbasierten Modellierung integriert werden. Die Programmausgabe des in Kapitel 7 beschriebenen Regelsystems ist ein modifiziertes Lexikon mit 4 Aussprachevarianten pro Eingabewort. Zu diesen vier genuinen Akzentstufen tritt stets die kanonische L2-Aussprache (Stufe 0) als eine weitere, potentiell mögliche Realisierung hinzu, so daß letztlich fünf Aussprachevarianten vorliegen. Auf diese Weise kann die Wahrscheinlichkeit maximiert werden, daß die automatisch generierten Varianten die in einer Sprechergruppe beobachteten Aussprachen hinreichend genau approximieren.

Es sei abschließend angemerkt, daß das Modell der Akzentstufen sich weitgehend neutral zu einigen der in Kapitel 2 skizzierten „großen Hypothesen“ der L2-Phonetik und -phonologie verhält. Betrachtet man etwa vergleichend die „starke Kontrastivhypothese“ mit dem Konzept der Äquivalenzklassifizierung, so lassen sich aus beiden Modellen durchaus widersprüchliche Vorhersagen über die dominierenden phonetischen Prozesse ableiten, welche für das Zustandekommen akzentbehafteter Aussprache prägend sind: Während die Kontrastivhypothese tendenziell einen Ersatz jener L2-Laute prognostiziert, die in der L1 des Sprechers *nicht* auftreten, leitet sich aus dem Konzept der Äquivalenzklassifizierung die Voraussage ab, daß gerade jene Laute „ersetzt“ (d.h. nicht entsprechend der L2-Zielform realisiert) werden, welche in der L1 des Sprechers ein phonetisch ähnliches Äquivalent aufweisen. Die daraus resultierenden Prädiktionen sind nicht immer miteinander vereinbar. Mittels des Akzentstufen-Modells ist es jedoch möglich, beide Formen phonetischer Prozesse gleichberechtigt zu berücksichtigen, indem beide Typen auf unterschiedlichen Akzentstufen modelliert werden.

II.

DATENERHEBUNG UND ANALYSE

5 Erstellung einer domänenspezifischen Sprachdatenbank⁸

5.1 Zur Rolle von Sprachdatensammlungen in Linguistik und Sprachtechnologie

Innerhalb der Linguistik der vergangenen ca. fünf Jahrzehnte spielten Korpora gesprochener und geschriebener Sprache als empirische Basis sprachwissenschaftlicher Untersuchungen eine eher marginale Rolle. Durch das Aufkommen und den Erfolg verschiedener „mentalistic“ oder „rationalistischer“ Strömungen verlagerte sich der sprachwissenschaftliche Fokus auf die Beschreibung kognitiv repräsentierter Wissensstrukturen, welche die Grundlage des Sprachvermögens (die sog. *Kompetenz* bzw. *I-Language*, cf. Chomsky 1986) des individuellen Sprechers bzw. einer idealisierten Sprachgemeinschaft bildet. Eine der Hauptanforderungen, die dabei an sprachwissenschaftliches Arbeiten gestellt wurde, bestand darin, daß linguistische Strukturbeschreibungen das internalisierte sprachliche Wissen des Sprechers adäquat charakterisieren. Damit verbundene grundsätzliche Einwände gegen empirische bzw. korpusbasierte Methoden berufen sich dabei noch heute oftmals auf die von Chomsky (1957, 1959) formulierte Kritik an einer strukturalistisch-taxonomischen Linguistik, wie sie bis zu dieser Zeit verbreitet war.

Nicht zuletzt durch das Aufkommen sprachtechnologischer Anwendungen und den Erfolg datenbasierter und statistischer Methoden bei ihrer Entwicklung kann jedoch etwa seit Beginn der 1990er Jahre ein neuerliches Interesse an der Gewinnung, Analyse und dem technischen Einsatz von Sprachdaten konstatiert werden (cf. hierzu auch Church 2003). Dieser Wandel - gelegentlich auch als *return of empiricism* bezeichnet (cf. Jurafsky & Martin 2000: 14) - wirkt dabei möglicherweise künftig auch auf die Linguistik selbst zurück: So sind derzeit Bestrebungen zu beobachten, primär zu sprachtechnologischen Zwecken erhobene, oft sehr umfangreiche Datenbestände auch für die empirische linguistische

⁸ Teile dieses Kapitels sind in englischer Sprache erschienen in meinem Artikel: Schaden, S. (2002a): "A database for the analysis of cross-lingual pronunciation variants of European city names." *Proceedings Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Vol. 4, 1277-1283.

Forschung heranzuziehen (siehe z.B. van Bael, Strik & van den Heuvel 2004 als Beispiel für solche Aktivitäten). Auch wenn es sich zunächst nur um eine sekundäre Verwertung handelt, so ist hierin dennoch eine fruchtbare Forschungsperspektive einer empirischen Sprachwissenschaft zu sehen – und möglicherweise auch eine weitere Konvergenz der beteiligten Teildisziplinen zu erwarten.

Als „Sprachdatenbanken“ (*spoken language corpora*, *spoken language databases*, *SDB*) im engeren Sinne werden umfangreiche Sammlungen von strukturierten, in digitalisierter Form vorliegenden Sprachaufzeichnungen bezeichnet, die mit linguistischen Annotationen auf verschiedenen Ebenen versehen wurden (eine Einführung und Übersicht bietet Draxler 2000). Derartige Korpora spielen bei der Entwicklung sprachtechnologischer Systeme eine zentrale Rolle. Unter anderem sind hierbei folgende Einsatzbereiche zu nennen:

- Gewinnung (Training) von akustischen Modellen für die automatische Spracherkennung; Adaption von akustischen Modellen für spezielle Einsatzbereiche von Spracherkennern (z.B. spezielle Sprechstile, dialektale Färbungen);
- Test von automatischen Spracherkennern mit definierten Sprechergruppen;
- Gewinnung von Sprachbausteinen für korpusbasierte automatische Sprachsynthesysteme;
- Untersuchung von natürlichen Dialogen als Vorstufe ihrer Modellierung für automatische Sprachdialogsysteme.

Eine Gemeinsamkeit dieser Anwendungen ist, daß zumeist Sprachdaten in erheblichem Umfang benötigt werden. So werden etwa für das Training von automatischen Spracherkennungssystemen oftmals Datensammlungen von vielen hundert Sprechern verwendet, um hieraus mittels statistischer Methoden akustische Referenzmodelle für sprecherunabhängige Systeme zu generieren. Gleiches bezüglich der Quantität der Daten gilt für Korpora geschriebener Sprache (d.h. elektronisch verfügbare Texte), welche z.B. zum Training probabilistischer Grammatiken (sog. „Sprachmodelle“) für Spracherkennungssysteme eingesetzt werden. Auch hier sind Textkorpora in erheblicher Größe erforderlich, um eine umfassende lexikalische und syntaktische Repräsentation der betreffenden Sprache zu erzielen. Zumindest hinsichtlich der Quantität der Daten unterscheiden sich aktuelle datenbasierte Methoden sprachlichwissenschaftlicher Forschung somit deutlich von ihren früheren Vorläufern.

Auch für die Domäne fremdsprachlich akzentgefärbter Aussprache wurde in den vergangenen Jahren bereits eine Reihe von Sprachdatensammlungen erstellt. Zu nennen sind hier beispielsweise die folgenden Korpora:

- Die **Translanguage English Database** (Lamel et al. 1994). Sie enthält Aufzeichnungen von 188 wissenschaftlichen Kongreßvorträgen in englischer Sprache von Sprechern verschiedener Muttersprachen.
- Das **Strange Corpus I & II**⁹. Hierbei handelt es sich um Aufzeichnungen eines kurzen deutschen Textes, der von 88 Sprechern verschiedener Muttersprachen gelesen wurde (Strange Corpus I) bzw. Material verschiedener Sprechstile, das von 70 Sprechern (Strange Corpus II) artikuliert wurde.
- Das **ISLE Corpus of Non-Native Spoken English** (Menzel et al. 2000). Eine Sammlung von Aufzeichnungen deutscher und italienischer Muttersprachler (je 23 Sprecher), die 82 Sätze in der Zielsprache Englisch lesen.
- Das **NATO Native and Non-Native (N4) Speech Corpus** (Benarousse et al. 2001). Es enthält typische Kommandos des militärischen Funkverkehrs in englischer Sprache, gesprochen von deutschen, niederländischen, kanadischen und britischen Sprechern.
- Das **Hispanic-English Corpus** (Byrne et al. 1998). Es besteht aus Aufzeichnungen von 20 Stunden Telefonkonversation in englischer Sprache, gesprochen von 18 Muttersprachlern des Spanischen mit mindestens elementaren Englischkenntnissen.

Die genannten Korpora liefern eine wichtige Datengrundlage zur empirischen Untersuchung fremdsprachlicher Akzente. Es fällt dabei jedoch auf, daß die Zielsprache Englisch in den Korpora überrepräsentiert ist. Dies mag mit einer allgemeinen Dominanz des Englischen innerhalb der sprachtechnologisch orientierten Forschung zusammenhängen und ist im Sinne einer besseren Kommunizierbarkeit der Ergebnisse innerhalb einer internationalen Forschungsgemeinschaft durchaus gerechtfertigt. Die Eignung dieser Datenbanken zum Aufbau eines sprachenübergreifend einsetzbaren Regelsystems, wie es in der vorliegenden Arbeit angestrebt wird, ist damit jedoch eingeschränkt. Auf der Basis der L2 Englisch gewonnene Ergebnisse sind möglicherweise nicht ohne weiteres auf andere Zielsprachen übertragbar. Umgekehrt ist es denkbar, daß bestimmte Phänomene in der L2 Englisch gerade *nicht* auftreten. So hat sich im Verlauf der vorliegenden Arbeit die gleichzeitige Einbeziehung verschiedener Sprachen oft als vorteilhaft erwiesen, da bestimmte Charakteristika akzentgefärbter Aussprache nur bei einigen, nicht aber bei allen Zielsprachen auftreten (beispielsweise die in Kap. 7.4.1 beschriebene „Rekonstruktion stummer Phoneme“ für die Zielsprache Französisch). Bei einer Beschränkung auf nur eine Zielsprache (z.B. Englisch) wären wesentliche Merkmale akzentbehafter Aussprache unberücksichtigt geblieben.

⁹ Derzeit vertrieben durch das *Bayerische Archiv für Sprachsignale* (BAS) an der LMU München – <http://www.phonetik.uni-muenchen.de/Bas/>

Aufgrund der skizzierten Situation wurde für die vorliegende Untersuchung eine umfangreiche sprachenübergreifende Sprachdatenbank zur Analyse nicht-muttersprachlicher Aussprachevarianten erstellt. Sie wird im vorliegenden Kapitel beschrieben. Primär dient die Sprachdatensammlung im Rahmen dieser Arbeit folgenden Zwecken:

- der Untersuchung auftretender Aussprachevarianten für verschiedene Ausgangssprachen und Zielsprachen, insbesondere einer Analyse sprecherübergreifender Variation und Konstanz;
- der Bereitstellung von phonetischen Zielformen, die mittels der Regeln modelliert werden sollen;
- als Referenzkorpus für die Evaluierung der erstellten Regeln (siehe hierzu Kap. 8)

Es ist der folgenden Beschreibung vorzuschicken, daß die Datenbank in Umfang und Konzeption nicht als Trainingskorpus für Spracherkennungs- oder Sprachsynthesysteme geeignet ist, sondern primär als Forschungsinstrument zur phonetischen bzw. phonologischen Analyse von Aussprachevarianten innerhalb einer ausgewählten Domäne entworfen und realisiert wurde. Dies schließt jedoch nicht grundsätzlich aus, daß eine Reihe von Entwicklungsaufgaben – beispielsweise der Test von automatischen Spracherkennern mit akzentbehafteter Aussprache – mittels der Sprachdatensammlung durchgeführt werden können.

5.2 Beschreibung der Datenbank

5.2.1 Sprachen, Sprachenpaare und Sprachrichtungen

Die als empirische Grundlage erstellte Sprachdatensammlung beinhaltet deutlich mehr Sprachen bzw. Sprachrichtungen, als im Rahmen dieser Arbeit ausführlich behandelt werden können. Insgesamt umfaßt sie folgende sieben Sprachen: Deutsch (DEU), Englisch (EN), Französisch (FR), Italienisch (ITA), Niederländisch (NL), Spanisch (SPA) sowie eine in ihrem Umfang eingeschränkte Erweiterung für Tschechisch (CZ).

Daten, Analyse und Modellierung für die Sprachen Niederländisch, Tschechisch und Spanisch werden in dieser Arbeit nicht dargestellt. Der erweiterte Umfang der Sprachdatensammlung ist aber von großem Nutzen für mögliche Anschlußuntersuchungen, in denen z.B. die Übertragbarkeit der entwickelten Methode auf weitere Sprachen geprüft werden kann. Weiterhin kann auf der

Basis dieser Daten die Möglichkeit geprüft werden, ob und in welchem Maße bereits vorhandene Regeln für eine Sprachrichtung auf neue Sprachrichtungen übertragbar sind und sich auf diese Weise z.B. Kernregeln extrahieren lassen, die für mehrere Ausgangs- oder Zielsprachen zugleich gültig sind.

Grundsätzlich wurden die genannten Sprachen sowohl als L1 (Ausgangssprache der Sprecher) als auch als L2 (Zielsprache) in die Datensammlung einbezogen, so daß eine Vielzahl von Sprachenkombinationen/Sprachrichtungen erfaßt sind. Bei einer Anzahl von N untersuchten Sprachen ergeben sich rechnerisch $N^2 - N$ Sprachrichtungen (d.h. Kombinationen aus L1 und L2), welche jeweils eine separate Untersuchungseinheit bilden. Bei den vorliegenden 7 Sprachen entspricht dies 42 Sprachrichtungen. Jedoch wurden in der Sprachdatensammlung nicht sämtliche potentiellen Kombinationen erfaßt. Folgende Einschränkungen gelten dabei für die Sprachen Niederländisch, Spanisch und Tschechisch:

- **Niederländisch** wurde bisher nur als Zielsprache erfaßt. Hierbei sind insbesondere Aussprachevarianten von Interesse, welche bei Sprechern mit guten Kenntnissen der Fremdsprache Deutsch auftreten. Eine Hypothese ist, daß orthographische, graphemotaktische und morphologische Ähnlichkeiten zwischen Deutsch und Niederländisch dazu führen, daß Sprecher im Leseprozeß niederländisches Material nach deutschen Regeln realisieren. Hieraus lassen sich möglicherweise weiterreichende Schlüsse hinsichtlich des Phänomens der *Dritt-spracheninterferenz* (siehe Kap. 6.6) ableiten.

- **Tschechisch** wurde ebenfalls nur als Zielsprache einbezogen. Sprachaufzeichnungen mit der L2 Tschechisch liegen darüber hinaus nur für eine vergleichsweise geringe Anzahl von Sprechern vor (25 Sprecher verschiedener Muttersprachen). Die Einbeziehung der Zielsprache Tschechisch als ergänzendes Datenmaterial erfolgte vor allem mit Blick auf zwei Fragestellungen:

(1) Es ist von grundlegendem Interesse, welche Aussprachevarianten produziert werden, wenn Sprecher bisher keinen oder nur geringen Kontakt zur Zielsprache hatten. Dies gilt im Falle von Tschechisch für praktisch alle bisher erfaßten Sprecher. Daher kann die L2 Tschechisch als Beispiel für ein solches Szenario dienen.

(2) Vieles spricht dafür, daß die orthographische Repräsentation zielsprachlicher Wörter einen nicht unerheblichen Einfluß auf ihre Aussprache durch nicht-muttersprachliche Sprecher ausübt. Die tschechische Orthographie verfügt über eine Vielzahl diakritischer Markierungen (*Akut*, *Gravis*, *Háček*, *superskribierter Ring* bei <ů>) mit vielfältigen phonetischen Funktionen (Palatalisierung, Vokallängung u.a.). Dabei entsprechen die Basisgrapheme, welche mit diesen Diakritika kombiniert werden können, den aus den westeuropäischen Sprachen bekannten Graphemen (z.B. die Reihe <c, s, z, r, n, e> gegenüber <č, š, ž, ř, ň, ě>). Es ist von besonderem Interesse, welche Aussprache den diakritisch markierten

Graphemen von den Sprechern zugewiesen wird. Da die phonetische Funktion dieser Diakritika den Sprechern nahezu unbekannt ist, können aus den resultierenden Aussprachevarianten möglicherweise weiterreichende Rückschlüsse auf den Einfluß solcher orthographischer Minimaldifferenzen gezogen werden.

- **Spanisch** wurde lediglich als Ausgangssprache einbezogen, nicht jedoch als Zielsprache. Diese Erweiterung der Sprachdatensammlung erfolgte mit einem speziellen Erkenntnisinteresse: Anhand der Varianten, die von Sprechern der L1 Spanisch realisiert werden, kann in eventuellen Anschlußstudien exemplarisch überprüft werden, ob und wie weit es möglich ist, diese Varianten mit Regelsätzen zu generieren oder zu approximieren, die für andere Muttersprachen erstellt wurden (z.B. Italienisch, Französisch). Hieraus ergeben sich möglicherweise erste Hinweise auf eine sprachenübergreifende Anwendbarkeit einzelner Regelsätze und Möglichkeiten, Regeln für verschiedene L1 zusammenzufassen.

Mit diesen genannten Einschränkungen umfaßt die Sprachdatensammlung insgesamt 24 Sprachrichtungen und bildet damit ein äußerst umfangreiches Korpus innerhalb der Domäne fremdsprachlicher Akzente. Die folgende Tabelle zeigt alle in der Datensammlung erfaßten Sprachenpaare bzw. Sprachrichtungen in der Notation *Ausgangssprache (L1) → Zielsprache (L2)*.

Sprachenpaar	Sprachrichtung 1 → 2	Sprachrichtung 2 → 1
EN / FR	1. EN → FR	2. FR → EN
EN / DEU	3. EN → DEU	4. DEU → EN
EN / ITA	5. EN → ITA	6. ITA → EN
EN / NL	7. EN → NL	
DEU / FR	8. DEU → FR	9. FR → DEU
DEU / ITA	10. DEU → ITA	11. ITA → DEU
DEU / NL	12. DEU → NL	
FR / ITA	13. FR → ITA	14. ITA → FR
FR / SPA	15. SPA → FR	
FR / NL	16. FR → NL	
SPA / ITA	17. SPA → ITA	
SPA / NL	18. SPA → NL	
SPA / DEU	19. SPA → DEU	
SPA / EN	20. SPA → EN	
EN / CZ	21. EN → CZ	
FR / CZ	22. FR → CZ	
ITA / CZ	23. ITA → CZ	
SPA / CZ	24. SPA → CZ	

Tabelle 5.1: Sprachenpaare und Sprachrichtungen in der Datenbank (EN = Englisch, DEU = Deutsch, FR = Französisch, NL = Niederländisch, ITA = Italienisch, SPA = Spanisch, CZ = Tschechisch)

5.2.2 Vorbemerkung zur Selektion von Sprechern

Bei der Erstellung von Sprachdatensammlungen ist die Anwerbung und Auswahl geeigneter Sprecher (engl. *speaker recruitment*) stets eine wichtige Vorarbeit, die oftmals mit eigenen Schwierigkeiten verbunden ist (cf. hierzu auch Sanders & van den Heuvel 2001). Je nach Anwendungszusammenhang und Erkenntnisinteresse stehen unterschiedliche Anforderungen an die zu erstellende Datensammlung im Vordergrund. Wichtige Auswahlkriterien können dabei z.B. sein: der Umfang der Sprachdatensammlung (Anzahl von Sprechern), die Repräsentativität der Stichprobe für die zu untersuchende Sprechergruppe, die generelle demographische Ausgewogenheit u.v.m.. Zumeist sind mehrere dieser Anforderungen zugleich zu erfüllen, so daß die Selektion von Sprechern durchaus zu einer komplexen Aufgabe erwachsen kann.

Leider kollidiert der Wunsch nach einer „idealen Datensammlung“ jedoch nicht selten mit dem Machbaren. Allein das Auffinden und Anwerben geeigneter Sprecher ist oftmals mit erheblichen Hindernissen verbunden, da die meisten

Studien innerhalb eines begrenzten geographischen und sozialen Aktionsradius stattfinden und somit kein unbegrenzter Zugriff auf beliebige Sprecher nach frei wählbaren demographischen und individuellen Parametern gegeben ist. Je spezieller dabei die Anforderungen an die einzelnen Sprecher sind, desto unwahrscheinlicher ist es, daß eine hinreichende Anzahl von Sprechern angeworben werden kann, welche diese Kriterien exakt erfüllen.

In der Praxis wird daher – geleitet vom jeweiligen Untersuchungsziel – eine Gewichtung der Kriterien erfolgen, um zu einer forschungspraktisch adäquaten Datensammlung zu gelangen. In der vorliegenden Untersuchung wurde dem Kriterium des *Umfangs* der Datensammlung (im Sinne der Anzahl erfaßter Sprecher) Priorität vor dem Kriterium der Repräsentativität der Sprechergruppe eingeräumt. Der Hauptgrund dafür ist darin zu sehen, daß die angestrebte Modellierung akzentbehafteter Aussprache nur dann erfolgversprechend ist, wenn es gelingt, *sprecherübergreifend* auftretende charakteristische Aussprachefehler zu erfassen. Dies kann freilich nicht gelingen, wenn die Datenbasis hinsichtlich der Anzahl von Sprechern zu stark eingeschränkt ist, denn in diesem Falle bestünde stets die Gefahr, daß letztlich nur individuelle Sprechermerkmale, nicht aber sprecherübergreifend gültige Merkmale der Aussprache modelliert werden.

Zudem ist die Repräsentativität der Sprechergruppe ein Kriterium, welches im Falle akzentbehafteter Aussprache aus prinzipiellen Gründen schwer erfüllbar ist. Das Problem besteht darin, daß nicht präzise und umfassend bestimmt werden kann, welche Kriterien eine Sprechergruppe erfüllen sollte, um als repräsentativ zu gelten: Alles bisher verfügbare Wissen deutet darauf hin, daß ein spezifischer Akzent bzw. eine spezifische damit verbundene Sprechweise das Resultat einer Vielzahl ineinanderwirkender linguistischer, psychologischer, soziologischer, situativer und physiologischer Variablen ist. Alter, Herkunft, Bildung, Motivation oder Physiognomie des Sprechers sind nur einige der Faktoren, die hierbei relevant werden. Wollte man diese Variablen in all ihren verschiedenen möglichen Ausprägungen und Kombinationen angemessen berücksichtigen, so wäre bereits für vergleichsweise kleine Studien eine Anzahl von mehreren hundert Sprechern als noch zu wenig repräsentativ anzusehen.

Wird darüber hinaus – wie es in der vorliegenden Studie der Fall ist – angestrebt, verschiedene Sprachenkombinationen parallel zu untersuchen, so potenziert sich diese Anzahl noch weiter, ohne daß damit garantiert wäre, tatsächlich jeden für das Zustandekommen einer spezifischen Aussprachevariante relevanten Einflußfaktor erfaßt zu haben. So ist es beispielsweise eine noch offene Frage, ob und in welcher Weise der dialektale muttersprachliche Hintergrund eines Sprechers sich auf seine phonetische Produktion einer Fremdsprache auswirkt (cf. hierzu der Sammelband James & Kettemann 1983). Vertritt man aber die Arbeitshypothese, daß ein solcher Einfluß besteht, so bedeutete dies, daß zur Erstellung einer repräsentativen Datensammlung für *jede* untersuchte Mutterspra-

che Sprecher aus *jeder* relevanten Dialektregion heranzuziehen wären. Ganz abgesehen von der damit zusätzlich aufgeworfenen Frage, welche und wie viele Dialektregionen existieren und zu berücksichtigen wären, ist leicht zu erkennen, daß die zur Erreichung des Ziels einer repräsentativen Sprechergruppe notwendige Anzahl von Sprechern sich bald in einer Dimension bewegen würde, die forschungspraktisch kaum mehr realisierbar erscheint.

Es ist angesichts dieser Umstände geboten, eine heuristische Strategie zu wählen, welche dem angestrebten Untersuchungsziel bzw. dem aktuellen Erkenntnisinteresse bestmöglich gerecht wird, sich dabei aber an forschungspraktisch realisierbaren Maßstäben orientiert. Dieser Weg wurde bei der vorliegenden Sprachdatensammlung verfolgt.

5.2.3 Sprecher

Derzeit umfaßt die Sprachdatensammlung ca. 160 Sprecher, von denen Sprachaufzeichnungen von 40 bis 60 Minuten je Sprecher vorliegen. Die untersuchte Sprechergruppe setzt sich vorwiegend, aber nicht ausschließlich aus Studenten aus Deutschland, Frankreich, den britischen Inseln (Großbritannien und Irland) sowie Italien und Spanien zusammen, wobei die nicht aus Deutschland stammenden Sprecher nur für begrenzte Zeit in Deutschland ansässig waren (Austauschstuden-ten).

Die Anwerbung von Austauschstudenten als Versuchspersonen ist einerseits mit erheblichen Vorteilen verbunden. Zum einen konnte auf diese Weise innerhalb relativ kurzer Zeit eine große Anzahl von Personen angeworben werden, um so zu aussagekräftigen Ergebnissen hinsichtlich des Faktors der interindividuellen Variation der Aussprache zu gelangen. Zum zweiten war es dadurch relativ problemlos möglich, Sprecher verschiedener Muttersprachen in ausreichender Zahl zu erreichen und in die Datensammlung einzubeziehen. Andererseits jedoch wirkt sich eine solche Strategie in einigen Aspekten negativ auf die Sprecherdistribution innerhalb der Datensammlung aus. Vor allem sind zwei Faktoren zu nennen, in denen dadurch ein Ungleichgewicht entsteht:

(1) **Faktor Alter:** Eine Sprecherrekrutierung unter Studenten führt erwartungsgemäß zu einem relativ niedrigen Durchschnittsalter und zu einer zu geringen Streuung der Variable „Alter“. So ist der Altersdurchschnitt der Probandengruppe mit 24,6 Jahren (Standardabweichung $s = 9,2$) verhältnismäßig niedrig (siehe auch Tabelle 5.2 auf Seite 98). Vertritt man die Hypothese, daß das Alter der Sprecher als Einflußgröße bei der Aussprache wirksam wird (cf. Eklund & Lindström 2001), so ist hierin zweifellos ein Defizit zu sehen.

(2) **Faktor Sprachkenntnisse:** Zweitens ist bei einer aus Studenten zusammengesetzten Probandengruppe nicht nur ein insgesamt hoher Bildungsstand,

sondern insbesondere auch eine relativ hohe Fremdsprachenkompetenz anzunehmen. Bei Austauschstudenten, die eine deutsche Universität besuchen, sind vor allem überdurchschnittliche Kenntnisse der Fremdsprache Deutsch zu erwarten. So zeigt Tabelle 5.2 (Seite 98), daß in der Gruppe der englischen und französischen Sprecher die durchschnittlichen Kenntnisse des Deutschen mit einem Wert von ca. 3,0 angegeben werden (auf einer Skala zwischen 0 = *keine Kenntnisse* und 4 = *sehr gute Kenntnisse*). Bezüglich der Zielsprache Deutsch liegt also eine hohe Ausprägung der Variablen „L2-Kennntnis“ vor; die Verteilung der Sprecher ist hier nicht ausgewogen. Für die zusätzlich einbezogenen Zielsprachen Niederländisch und Italienisch sind die Kenntnisse hingegen bei den meisten Probanden außerordentlich gering, so daß diese Zielsprachen sich zur Untersuchung stark ausgeprägter Akzente eignen. Somit ergibt sich, wenn alle untersuchten Zielsprachen eingebezogen werden, eine ausgewogenere Verteilung der Sprecher in Bezug auf Fremdsprachenkenntnisse.

5.2.4 Sprachmaterial

Als Sprachmaterial wurden für jede der Zielsprachen zweierlei Vokabular- bzw. Texttypen herangezogen, die im folgenden skizziert sind.

(1) Ortsnamen

Als Material für den zentralen Teil der Datensammlung – die Aussprache von Toponymen – wurden für jede untersuchte Zielsprache 43 Ortsnamen zusammengestellt. Als Quelle dienten Atlanten, im Internet verfügbare Listen geographischer Namen, Routenplanungs-Programme und ähnliche Ressourcen. Unter anderem wurden folgende Kriterien bei der Selektion der Stimuli zugrunde gelegt:

Abdeckung von „kritischen Merkmalen“. Die Mehrzahl der Stimuli enthält orthographische oder phonetische Merkmale, von denen im Sinne einer Ausgangshypothese angenommen wurde, daß sie für Nicht-Muttersprachler problematisch hinsichtlich ihrer Aussprache seien. Beispiele hierfür sind <th>-Cluster, die im Englischen orthographisch die interdentalen Frikative /ð/ and /θ/ repräsentieren (*Appleton Thorn, Avonmouth*), nasalierte Vokale oder komplexe orthographische Vokalgruppen mit einfacher lautlicher Entsprechung im Französischen (z.B. *Langeais*), oder sog. ‚Umlaute‘ in deutschen Namen (*Fünfstetten, Niedermöllern*).

Hierbei ist natürlich stets zu bedenken, daß es in hohem Maße vom muttersprachlichen Hintergrund des Sprechers abhängt, ob ein Merkmal als „problematisch“ oder „kritisch“ hinsichtlich seiner akkuraten Aussprache gelten kann. So wird beispielsweise eine korrekte Realisierung des /r/-Phonems als alveolarer

Vibrant [r] in der Zielsprache Italienisch für einen Sprecher des Spanischen weniger problematisch sein als für einen Sprecher des Deutschen, da das Spanische ebenfalls über diesen Laut verfügt. Dennoch wurde für die vorliegende Studie entschieden, identische Stimuli für Probanden aller Muttersprachen zu verwenden, da so eine bessere Vergleichbarkeit der Varianten von Probanden mit unterschiedlichem muttersprachlichen Hintergrund gegeben ist.

Präferenz unbekannter Ortsnamen. In Kap. 3.3 wurde dargelegt, daß für eine Reihe von Ortsnamen lexikalisierte *Exonyme* existieren und für einige weitere Ortsnamen – insbesondere Großstadtnamen wie *Berlin, London, Marseille* – eine relativ stabile und konventionalisierte Aussprache durch Nicht-Muttersprachler vorherrscht. Der Schwerpunkt bei der Zusammenstellung der Stimuli lag daher auf verhältnismäßig unbekannt Namen, deren Aussprache durch Nicht-Muttersprachler weniger vorhersahbar ist.

Ausschluß von Ortsnamen aus regional-spezifischen Sprachzweigen und Varietäten. In Fällen, in denen Ortsnamen eines Landes verschiedenen Sprachzweigen entstammen, wurde nur die Standardvarietät berücksichtigt. Dies bedeutet, daß etwa britische Ortsnamen walisischen oder schottischen Ursprungs, welche zumindest in ihrer Schreibung keine deutliche Assimilation an die englische Sprache erfahren haben (etwa walisisch *Aberystwyth* oder schottisch *Crianlarich*), nicht als Stimuli aufgenommen wurden. Gleiches gilt für französische Namen bretonischer Herkunft (etwa *Guingamp, Plogoff*).

Die folgenden Beispiele zeigen einige der in den Versuchen verwendeten Ortsnamen:

Deutsch	<i>Gellmersdorf, Aindling, Thränitz, Fünfstetten, Blankenstein</i>
Englisch	<i>Cheltenham, Hemel Hempstead, Shrewsbury, Toddington, Plymouth</i>
Französisch	<i>Fougères, Issoire, Questembert, Riom Châtelguyon, Fréjus</i>
Niederländisch	<i>Bloemwijk, Purmerend, Enkhuizen, Goes, Arnemuiden</i>
Italienisch	<i>Pegognaga, Roncobilaccio, Faenza, Chiavari, Battipaglia</i>
Tschechisch	<i>Přísečná, Polště, Nemyšl, Ševětín, Štěpánovice, Kvítkov Dvůr</i>

Zur Entlastung der Probanden wurden den Stimulus-Listen einige wenige Einträge hinzugefügt, deren Aussprache in der Regel unproblematisch ist (z.B. *Roma, Verona, Paris, London, Berlin*). Weiterhin wurden am Ende der Listen einzelne Einträge, die bereits zu Beginn aufgeführt waren, wiederholt, um hieraus Rück-

schlüsse auf intraindividuelle Variation bzw. auf eventuell auftretende Lern- oder Gewöhnungseffekte ziehen zu können.

Der Vollständigkeit halber sei erwähnt, daß – entgegen einer bei der Kompilierung von Sprachdatensammlungen gängigen und in anderen Kontexten durchaus sinnvollen Praxis – nicht angestrebt wurde, ein phonemisch ausbalanciertes Korpus zu generieren, d.h. ein Korpus, bei dem die relative Auftretenshäufigkeit einzelner Phoneme deren Auftretenshäufigkeit innerhalb der untersuchten Sprache als ganzes reflektiert. Während dieses Kriterium etwa für Trainingsdatenbanken in der automatischen Spracherkennung von großer Bedeutung ist (cf. Gibbon, Moore & Winski 1997: 101), erscheint seine Anwendung im vorliegenden Zusammenhang nicht ohne weiteres sinnvoll, da möglicherweise gerade die *weniger häufig auftretenden Phoneme* einer Zielsprache für Nicht-Muttersprachler besondere Problemfälle darstellen und daher von größerem Interesse sind.

(2) Sätze

Zusätzlich zu den Ortsnamen wurden für jede Zielsprache (Ausnahme: Tschechisch) Listen von je 10 kurzen Sätzen zusammengestellt. Sie sind Artikeln verschiedener Online-Zeitungen¹⁰ entnommen und stehen untereinander in keiner inhaltlichen Beziehung. Die Sätze haben eine einfache syntaktische Struktur und eine maximale Länge von 60 Zeichen, wie die folgenden Beispiele zeigen:

Deutsch	<i>Ich bin allerdings anderer Meinung.</i>
Englisch	<i>This is how the market works</i>
Französisch	<i>Qu'est-ce qu'on va faire?</i>
Niederländisch	<i>We wachten op concrete plannen.</i>
Italienisch	<i>La storia d'Europa ricominciava.</i>

Im Rahmen der vorgestellten Sprachdatensammlung erfüllen die Sätze einen doppelten Zweck: Erstens eignen sie sich zur Überprüfung und eventuellen Korrektur der von den Versuchspersonen selbst angegebenen Fremdsprachenkenntnisse (siehe unten); zweitens können die Sätze als unabhängige Datensammlung Verwendung finden, etwa zur Überprüfung der Gültigkeit der anhand von Ortsnamen gewonnenen Regeln für Standard-Vokabular.

¹⁰ Zugriff über das Portal <http://www.onlinenewspapers.com>

5.2.5 Versuchsdesign und -durchführung

Die Einweisung der Probanden bildet einen wesentlichen Bestandteil einer experimentellen Untersuchung fremdsprachlicher Performanz, da stets die Gefahr besteht, daß die Versuchssituation selbst sich in unerwünschter Weise auf die sprachliche Produktion auswirkt.

Das vorrangige Problem besteht darin, daß Sprecher leicht den Eindruck gewinnen, sie würden hinsichtlich ihrer fremdsprachlichen Performanz geprüft und beurteilt (cf. hierzu auch Tomokiyo & Burger 1999). Aus diesem Grunde wurden alle Aufnahmesitzungen mit einer Einführung durch den Versuchsleiter eingeleitet, in der nicht nur die Ziele des Experiments offengelegt wurden, sondern auch betont wurde, daß es sich *nicht* um eine Testsituation handelt und eine fehlerfreie Aussprache nicht erwartet werde. Wenn es auch in einer letztlich artifiziellen Aufnahmesituation nicht möglich ist, den Eindruck eines ‚Sprachtests‘ vollständig zu entkräften, so kann eine solche Einweisung dazu beitragen, ihn abzuschwächen.

Vor den Aufnahmesitzungen (Dauer pro Sprecher ca. 40 bis 60 Min.) wurden von jedem Sprecher folgende Angaben zur Person erfragt:

- Alter, Geschlecht, Name, Adresse, Beruf oder Studienfach
- Muttersprache bzw. Muttersprachen (bilinguale Sprecher)
- Herkunftsland und -region
- Derzeitiger Wohnsitz im Heimatland
- Sprachkenntnisse nach Selbsteinschätzung
 - (a) für die im Projekt untersuchten Sprachen
 - (b) für beliebige weitere Sprachen (sofern vorhanden)

Die in den Fragebögen angegebenen Informationen sind von zentraler Bedeutung, um auftretende Merkmale der Aussprache in Beziehung zu spezifischen Sprechereigenschaften setzen zu können. Die im vorliegenden Zusammenhang vermutlich wichtigste Eigenschaft ist der sprecherspezifische Kenntnisstand der jeweiligen Zielsprache.

Trotz der damit verbundenen möglichen Einbußen an Akkuratheit wurde entschieden, den L2-Kennntnisstand auf der Basis einer Selbsteinschätzung der Sprecher zu erheben und keinerlei standardisierte Sprachtests einzusetzen. Zweifellos ist eine Selbsteinschätzung mit Vorbehalten zu betrachten. So ist beispielsweise zu erwarten, daß Sprecher ihren eigenen Kenntnisstand der Zielsprache(n) auf der Basis von aktivem Vokabularumfang, Syntax, Idiomatik und allgemeiner Geläufigkeit im Gebrauch der Fremdsprache einschätzen werden, Eigenschaften also, die nicht notwendigerweise mit akkurater L2-Aussprache korrelieren. Darüber hinaus ist eine Selbsteinschätzung stets mit der Gefahr verbunden, die eige-

nen Fähigkeiten über- oder unterzubewerten und damit eine zu positive oder negative Einschätzung zu geben.

Dennoch wurde die Methode der Selbsteinschätzung gewählt, da es in der vorliegenden Untersuchung weniger um eine exakte Quantifizierbarkeit von L2-Kenntnissen geht, sondern vielmehr um die Identifizierung von Korrelationen zwischen L2-Kenntnisstand und spezifischen Aussprachefehlern. Zudem konnten die in Teil (3) des Versuchs (siehe Kapitel 5.2.6) erhobenen Sprachdaten dazu verwendet werden, zweifelhafte Selbsteinschätzungen ggf. zu korrigieren. Generell war dabei die Tendenz festzustellen, daß die Probanden eher eine zu negative Einschätzung ihrer L2-Kompetenz abgaben, was vermutlich auf die Versuchssituation und die damit einhergehende Erwartung eines ‚Sprachtests‘ zurückzuführen ist. Die Kenntnisse der Zielsprachen wurden auf einer Skala von 0 bis 5 angegeben, welche den folgenden verbalen Deskriptoren entsprechen:

Index	Beschreibung
0	keine (<i>none</i>)
1	Grundkenntnisse / Anfänger (<i>basic knowlegde/beginner</i>)
2	durchschnittlich (<i>intermediate</i>)
3	gut (<i>good/fair</i>)
4	sehr gut (<i>excellent/near-native</i>)
5	muttersprachliche Beherrschung (<i>native</i>)

L2-Kenntnisse wurden nicht nur für die in dieser Untersuchung behandelten Zielsprachen erfragt, sondern (in Form einer offenen Frage) auch für jede weitere Fremdsprache, die der Person geläufig ist. Diese Information kann insofern von Bedeutung sein, als häufig zu beobachten ist, daß Sprecher bei der Aussprache von ihnen unbekanntem Sprachen Kenntnisse aus bekannten Sprachen heranziehen (*Drittsspracheninterferenz*, siehe Kap. 6.6).

Die folgende Tabelle gibt eine Übersicht der in der Datensammlung erfaßten Sprecher und ihrer Kenntnisse der verschiedenen Zielsprachen und faßt damit einige der bisher ausgeführten Sprechermerkmale noch einmal zusammen:

Mutter- sprache	Anzahl Sprecher	männl. / weibl.	durch- schnittl. Alter	durchschnittliche L2-Kenntnisse (0 ≤ n ≤ 5)				
				Engl.	Franz.	Deut.	Ital.	Nieder- länd.
Englisch	33	17 / 16	21,9	(5,0)	1,31	3,0	0,14	0,18
Französisch	26	8 / 18	24,3	2,6	(5,0)	2,8	0,41	0,07
Deutsch	29	18 / 11	23,7	2,97	1,25	(5,0)	0,5	0,3
Italienisch	47	18 / 29	29,3	2,15	0,64	1,74	(5,0)	0,09
Spanisch	26	11 / 15	23,7	2,62	0,96	2,31	0,65	0,04
GESAMT / DURCHSCHN.	161	72 / 89	24,6	2,58	1,04	2,46	0,42	0,13

Tabelle 5.2: Sprecher und durchschnittliche L2-Kenntnisse
(ohne Zielsprache Tschechisch)

Die Tabelle zeigt, daß in der untersuchten Sprechergruppe die Kenntnisse der Zielsprachen Englisch und Deutsch besonders ausgeprägt sind, während für die Zielsprachen Französisch und insbesondere Italienisch und Niederländisch nur geringe Kenntnisse vorhanden sind. Für die vorliegende Untersuchung ergibt sich daraus ein Ungleichgewicht zwischen den verschiedenen Zielsprachen. Die aus den Sprachdaten abgeleiteten Regeln bilden im Falle von Englisch und Deutsch als Zielsprache eine Sprechergruppe mit relativ hohen L2-Kenntnissen (und möglicherweise weniger ausgeprägtem Akzent) ab, während im Falle der verbleibenden Sprachen starke Akzentaussprache zu erwarten ist.

5.2.6 Teilaufgaben/Szenarien

Für jeden Sprecher setzte sich der Versuch bzw. die Aufnahmesitzung aus drei Teilaufgaben zusammen, die im folgenden erläutert sind. Hieraus ergeben sich drei Teilkorpora mit z.T. unterschiedlichem Material und unterschiedlichen Modi der Stimulus-Darbietung.

(a) **Ortsnamen, gelesen (orthographische Stimuli).** Den Probanden wurde die Aufgabe gestellt, 45 (43 + 2) Ortsnamen aller Zielsprachen als isolierte Wörter vorzulesen. Dabei wurden jeweils 2 Ortsnamen am Ende der Stimulus-Liste wiederholt, um ggf. prüfen zu können, ob und in welchem Ausmaß intra-individuelle Variation beim Lesen der Namen auftritt. Aufgabe (a) ist obligatorisch für alle Versuchspersonen, unabhängig von ihrer zielsprachlichen Kompetenz. Da die Stimuli als isolierte Wörter dargeboten wurden, konnte die Aufgabe

auch solchen Sprechern gestellt werden, die über nur geringe oder keinerlei Kenntnisse der Zielsprache verfügen.

Dies ist durchaus als ein wesentlicher Bestandteil der Untersuchungen zu sehen, denn gerade anhand der Aussprache von Material aus einer dem Probanden unbekanntem Sprache lassen sich oftmals die von den Sprechern angewandten Strategien erkennen (z.B. die Etablierung „spontaner“ Regeln, Anwendung von Drittsprachenkenntnissen etc.). Zudem war es den Versuchspersonen ausdrücklich gestattet, die Namen zunächst stumm zu lesen, bevor sie artikuliert wurden. Da jegliche Verzögerung beim Lesen der Namen in den Audioaufnahmen erfaßt wurde, kann diese Zusatzinformation bei der Auswertung der Daten ggf. auch dazu genutzt werden, den Schwierigkeitsgrad einzelner Namen einzuschätzen. Zu beachten ist auch, daß – im Gegensatz zu einer ähnlichen Studie von Fitt (1995), bei der die Probanden die sprachliche Herkunft der auszusprechenden Städtenamen selbst mittels eines *multiple choice*-Verfahrens zu bestimmen hatten – den Probanden im vorliegenden Versuch die linguistische Herkunft der Namen zu jeder Zeit bekannt war.

Abschließend wurden ausgewählte Sprecher aufgefordert, die Ortsnamen ihrer Muttersprache zu lesen. Dieses Material wurde ebenfalls aufgezeichnet, um später in Teil (b) des Versuchs als akustische Stimuli eingesetzt zu werden.

(b) Ortsnamen, Wiederholung akustischer Stimuli. In Teil (b) der Versuche wurden die Probanden instruiert, die von einem Muttersprachler gesprochene, akustisch dargebotene Standard-Aussprache der Ortsnamen zu wiederholen. Das dabei verwendete Sprachmaterial ist identisch mit den in Versuch (a) verwendeten 43 Ortsnamen, die jedoch in einer anderen Reihenfolge dargeboten wurden. Die akustischen Stimuli wurden über Kopfhörer dargeboten, wobei die Probanden die Möglichkeit hatten, einen Stimulus vor dem Nachsprechen mehrfach anzuhören. Für die akustischen Stimuli wurden geeignete muttersprachliche Sprecher aus vorangehenden Aufnahmen herangezogen.

Die hier beschriebene Zweiteilung der Versuche in eine Lese- und eine akustische Reproduktionsaufgabe bei identischem Sprachmaterial ermöglicht es, die Realisierungsvarianten bei orthographischer und akustischer Präsentation zu vergleichen. Hieraus lassen sich Rückschlüsse auf spezifische Einflüsse der Orthographie (Schriftaussprache, Graphem-Phonem-Beziehungen) ziehen und von rein phonetisch bedingten Interferenzeffekten abgrenzen. In späteren Kapiteln dieser Arbeit ist exemplarisch dargestellt, in welcher Weise sich die Aussprachevarianten in Teil (a) und (b) des Versuchs unterscheiden.

(c) **Sätze, gelesen (orthographische Stimuli).** Im Teil (c) der Aufzeichnungen wurde den Probanden die Aufgabe gestellt, die oben beschriebenen je 10 Sätze der Zielsprachen zu lesen. Da es nicht als sinnvoll erachtet wurde, daß Sprecher ohne jegliche Kenntnisse der Zielsprache an diesem Teilversuch teilnehmen, wurde den Probanden – im Gegensatz zu den Aufgaben (a) und (b), die für alle Sprecher obligatorisch waren – freigestellt, ob und für welche Zielsprachen sie Teil (c) durchführen.

Erwartungsgemäß führten die Probanden diesen Teil des Versuch nur für die ihnen bekannten Zielsprachen durch. Aus diesem Grund ist die Anzahl der aufgezeichneten Äußerungen je Sprecher variabel. Wie oben dargelegt, wurde bei der Stimulus-Auswahl darauf Wert gelegt, daß die Sätze hinsichtlich Länge und syntaktischer Komplexität möglichst einfach gehalten sind, um die Sprecher nicht zu überfordern und Verzögerungen, Lesefehler, Wiederholungen und dergleichen zu vermeiden. Dennoch traten derartige Phänomene erwartungsgemäß in Teil (c) der Aufzeichnungen häufiger auf als in (a) und (b).

5.3 Lexikon und phonetische Transkription

5.3.1 Kanonische Transkription

Die Sprachdatenbank enthält ein Referenzlexikon, welches die orthographische Form sowie die phonetische Referenzform (kanonische Aussprache) für alle in den Versuchen verwendeten Ortsnamen und Sätze umfaßt. Das phonetische Referenzlexikon wird auf verschiedenen Ebenen der vorliegenden Studie benötigt. Insbesondere im Rahmen der Evaluation des Regelsystems (siehe Kap. 8) kommt der kanonischen Aussprache eine wichtige Rolle zu, da hier ermittelt wird, ob und wie weit die regelbasierten Varianten eine bessere Approximation an die Sprechervarianten erzielen können als die kanonische Form. In Fällen, in denen Ortsnamen bereits in ihrer Herkunftssprache zwei alternative Aussprachevarianten aufweisen (z.B. engl. *Broughton* [bɹɑʊtən] ~ [bɹɔ:tən]), wurden beide Varianten in das Referenzlexikon aufgenommen. Für die kanonische Transkription wurde dabei das SAMPA-Alphabet der jeweiligen Sprache herangezogen (Wells 2004).

5.3.2 Transkription der Sprachdaten

Grundlage für die Auswertung der Sprachdaten auf symbolisch-phonetischer Ebene ist das Vorliegen einer phonetischen Transkription der aufgezeichneten Äußerungen. Eine wesentliche Schwierigkeit bei der Erstellung einer Transkription nicht-muttersprachlicher Aussprachevarianten liegt jedoch darin, daß die von den Sprechern produzierten Laute häufig nicht eindeutig einem Phonemsystem bzw. dessen Allophenen zuzuordnen sind. Oftmals handelt es sich um eine Synthese aus muttersprachlichen und zielsprachlichen Lauten sowie nicht eindeutig klassifizierbaren Zwischenformen.

Daher ist es unzureichend, das SAMPA-Inventar nur einer Sprache heranzuziehen. Es ist zur genaueren Differenzierung erforderlich, das Lautinventar zu erweitern und die verwendeten Symbole im Sinne sprachunabhängiger phonetischer Werte zu interpretieren. Eine zunächst naheliegende Möglichkeit wäre dabei, die phonetischen Inventare (a) der Muttersprache des Sprechers und (b) der jeweiligen Zielsprache heranzuziehen und parallel zu verwenden. Jedoch ist SAMPA so konzipiert, daß die jeweiligen Phonemsymbole stets auf der Basis ihrer sprachspezifischen phonemischen (d.h. bedeutungsunterscheidenden) Werte zu interpretieren sind. Dies führt zu Überschneidungen der sprachspezifischen Inventare und damit verbundenen phonetischen Ambiguitäten: So wird beispielsweise innerhalb von SAMPA das Symbol /r/ sowohl für den 'r'-Laut [ɹ] im Englischen (alveolarer Approximant) als auch im Italienischen [r] verwendet (Vibrant), obwohl damit in artikulatorisch-phonetischer Hinsicht deutlich unterschiedliche Realisierungen repräsentiert sind.

Derartige Differenzierungen sollten jedoch unbedingt in der Transkription repräsentiert sein. So weisen die Aussprachevarianten häufig Kennzeichen wie die folgenden auf:

- eine Übertragung muttersprachlicher Realisierungsvarianten äquivalenter Phoneme auf die Zielsprache, z.B. die Realisierung des Phonems /R/ als engl. [ɹ] anstelle des dt. Uvularlautes [ʁ] (z.B. *Brett* → [b.ɹɛt]).
- eine gegenüber der L1-Standardrealisierung markierte Form, z.B. eine Realisierung des /R/ als Vibrant („gerolltes R“) bei engl. Sprechern: *Brett* → [brɛt]
- positionsbedingte Realisierungsvarianten von Phonemen, die auf eine Übertragung muttersprachlicher phonologischer Regeln auf die Zielsprache schließen lassen, z.B. eine Kontrastierung von alveolarem und velarisiertem /l/ bei englischen Sprechern: *Holz* → [hɔʎts] vs. [hɔʎts]

Daher wurde in Fällen, in denen eine Transkription nach SAMPA Ambiguitäten verursachen würde, zur engen phonetischen Transkription der Sprechervarianten das ebenfalls von Wells (2004) konzipierte X-SAMPA-Inventar herangezogen. Es

ermöglicht die Transkription phonetischer Details, die mittels der regulären sprachenspezifischen SAMPA -Inventare nicht codiert werden können.

Eine grundsätzliche Entscheidung ist darüber zu treffen, bis zu welchem Grad phonetische Details in die Transkription einfließen sollen. Barry & Fourcin (1992) unterscheiden fünf Ebenen der phonetischen Annotation (*labelling*), die von Ebene 1 bis 4 einen abnehmenden Grad an phonetischer Detailtreue aufweisen:

	Ebene	Kurzbeschreibung
(1)	Physikalische Ebene (<i>Physical level</i>)	Annotationen sind eng am physikalischen Sprachsignal orientiert.
(2)	Akustisch-phonetische Ebene (<i>Acoustic-phonetic level</i>)	Annotationen basieren auf bekannten, etablierten phonetischen Kategorien.
(3)	Enge phonetische Ebene (<i>Narrow-phonetic level</i>)	Annotationen verwenden etablierte phonetische Transkriptionssymbole wie z.B. das IPA-Alphabet
(4)	Phonemische Ebene (<i>Phonemic level</i>)	Annotationen verwenden etablierte phonetische Transkriptionssymbole wie z.B. das IPA-Alphabet, verzeichnen dabei jedoch ausschließlich bedeutungsunterscheidende (distinktive) Phoneme der betreffenden Sprache.
(5)	Breite phonetische Ebene (<i>Broad phonetic level</i>)	Zwischenebene: Annotationen verwenden zwar ausschließlich Phonemsymbole - entsprechend Ebene (4) - , diese werden jedoch z.T. auch verwendet, um nicht-phonemische (nicht-distinktive) Erscheinungen in kontinuierlicher Sprache zu notieren.

Tabelle 5.3: Annotations-Ebenen nach Barry & Fourcin (1992)

Legt man diese Typologie zugrunde, so sind die in dieser Studie erstellten Transkriptionen der dritten Ebene („eng-phonetisch“) zuzuordnen. Diese Ebene weist eine für den vorliegenden Zweck ausreichende Detailtreue auf, insofern sie wichtige lautliche Charakteristika der untersuchten Aussprachevarianten mit hinreichender Genauigkeit repräsentiert. Dennoch sind Transkriptionen dieser Granularität noch mit vertretbarem Transkriptionsaufwand zu erstellen. Dies ist durchaus ein maßgeblicher Faktor: Hohe phonetische Detailtreue in der Transkription schlägt sich stets in einem erhöhten Transkriptionsaufwand nieder. Durch die Wahl der genannten Transkriptionsebene 3 konnten im Rahmen dieser Studie bisher ca. 13.500 Äußerungen manuell phonetisch transkribiert wer-

den, so daß eine große Anzahl von Transkriptionen mit einer für den vorliegenden Zweck ausreichenden Genauigkeit vorliegt.

5.4 Von sprachlichen Daten zu Regeln

Ein Hauptziel bei der Erstellung der Sprachdatenbank ist die Identifizierung phonetischer Regularitäten und sprecherübergreifender Konstanten, die bei der Aussprache fremdsprachlichen Materials auftreten. Angesichts des limitierten Vokabularumfangs (43 Ortsnamen + 2 Wiederholungsstimuli sowie 10 Sätze für jede Zielsprache) stellt sich dabei die Frage, ob, wie und in welchem Maße die beobachteten sprecherübergreifenden Regelmäßigkeiten auf (a) *neue Sprecher* und (b) *neues Vokabular* übertragbar sind.

Aus diesem Grunde wurde neben der engen phonetischen Transkription der Sprechervarianten auf Wortebene eine weitere Annotationsebene eingeführt, die in soziolinguistischen oder dialektologischen Studien seit langem etabliert ist (cf. Hudson 1980, Kap. 5; Chambers & Trudgill 1980: 145ff.). Hier ist die atomare Untersuchungseinheit nicht das *Wort*, sondern die *phonologische Variable*. Eine phonologische Variable (im folgenden auch: PV) ist dabei definiert als ein Element – ein Phonem, ein Phonemcluster, eine Silbe oder andere untersuchungsrelevante Einheiten –, von dem als Ausgangshypothese angenommen wird, daß es innerhalb der Untersuchungsdomäne einem gewissen Maß an Variation unterliegt – z.B. in Abhängigkeit von Sprechermerkmalen wie regionaler Herkunft, sozialer Schichtzugehörigkeit oder der aktuellen Redesituation. Mit jeder Realisierung der Variable innerhalb der erhobenen Untersuchungsdaten nimmt die PV einen spezifischen *Wert* an, der als *Variante* bezeichnet wird. In statistischer Terminologie würde man von einem *Merkmal* mit seinen verschiedenen *Ausprägungen* sprechen (cf. Schlobinski 1996: 21ff.).

Die Arbeit mit phonologischen Variablen als Basiseinheit (anstelle von z.B. Wörtern, Lexemen, Phrasen, Sätzen etc.) bietet den Vorteil, daß trotz eines begrenzten Vokabulars, an dem die Untersuchung vorgenommen wird, verhältnismäßig viele Regelmäßigkeiten identifiziert werden können, welche sich möglicherweise auf neues, bisher nicht untersuchtes Vokabular übertragen lassen. Beschränkte man sich darauf, Wörter als Basiseinheit heranzuziehen, so wäre die Untersuchung bereits an ihrem Ziel angelangt, wenn alle vorkommenden Aussprachevarianten für alle Wörter des untersuchten Vokabulars vollständig aufgelistet sind – eine Übertragung des daraus gewonnenen Wissens auf neues Vokabular wäre jedoch nicht ohne weiteres möglich.

Phonologische Variablen im hier dargelegten Sinn sind nicht zwangsläufig als isolierte Phoneme, sondern als *Phoneme in einem spezifischen Kontext* definiert.

Der Kontext (d.h. die Position des fraglichen Lautes innerhalb des Wortes) ist fester Bestandteil der Variable, so daß identische Phoneme in verschiedenen Kontexten als gesonderte Variablen behandelt werden (z.B. wort-initiales gegenüber wort-finalem [a:]). Darüber hinaus treten einzelne phonologische Variablen innerhalb des Vokabulars nicht notwendigerweise in nur einem Lexikoneintrag auf, sondern können mehrfach vorhanden sein.

Das geschilderte Prinzip wurde für die vorliegende Sprachdatensammlung und deren Auswertung adaptiert: Jedes Lexem des untersuchten Vokabulars ist Träger einer oder mehrerer phonologischer Variablen. Um zu bestimmen, welches lautliche Merkmal als eine relevante PV anzusehen ist, wurde das aufgezeichnete Sprachmaterial einer Teilgruppe von Sprechern (5-10) von einer phonetisch geschulten Person abgehört und bewertet. Dort, wo eine sprecherübergreifend auftretende, auffallende Abweichung von der kanonischen Transkription zu verzeichnen war, wurde dies zunächst als eine potentielle PV markiert (d.h. als eine PV, deren weitere Untersuchung lohnend erscheint). In einem anschließenden Schritt wurde dann die Ausprägung dieser PV für *alle* Sprecher bestimmt.

Die folgende Abbildung 5.1 illustriert das Prinzip der PVs und ihrer Varianten am Beispiel des englischen Ortsnamens *Northampton* mit seinen Variablen /ɔ:/, /θ/ und /ə/ (hier in SAMPA-Notation mit den Symbolen /O:/, /T/ und /@/ wiedergegeben).

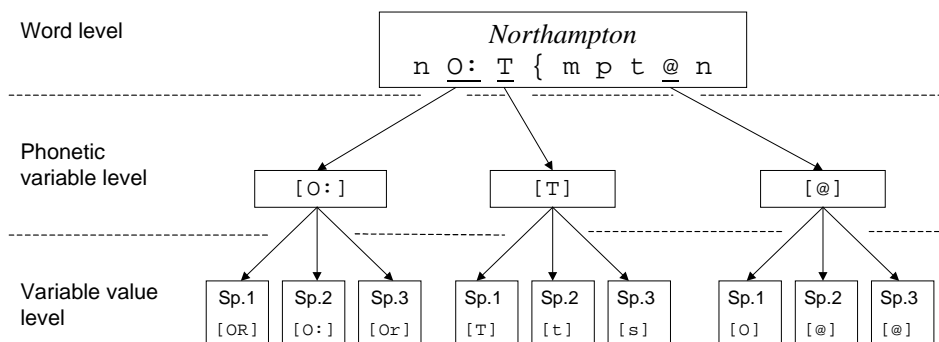


Abb. 5.1: Annotation phonetischer Variablen (aus: Schaden 2002a)

Ausgehend von einer solchen Annotation ist es möglich, aus den phonologischen Variablen und den Vorkommenshäufigkeiten ihrer einzelnen Varianten sprecherübergreifende Regelmäßigkeiten zu extrahieren und zu formalisieren. Eine solche quantitative Auswertung wird in Kap. 6 exemplarisch anhand einiger Lautsubstitutionen demonstriert.

Die relationale Datenbank, die in dieser Studie zur Speicherung sämtlicher symbolischer Annotationen verwendet wurde, unterstützt dabei eine halbautomatische Zusammenfassung der Variantenhäufigkeiten für phonologische Variablen. Aus einer solchen Zusammenfassung lassen sich in einem nächsten

Schritt bereits erste phonologische Regeln ableiten, die als Basis für die Erstellung von Regelsätzen dienen können. Die allgemeine Form dieser Regeln ist dabei

$$X \rightarrow Y \quad / \quad Z$$

wobei ein Eingabesymbol X im Kontext von Z in eine Variante Y transformiert wird. Der Wert von Y wird dabei prinzipiell definiert durch diejenige Variante von PV, welche in der untersuchten Sprechergruppe die höchste (oder zumindest eine hohe) Häufigkeit aufweist. So können etwa aus dem oben genannten Beispiel *Northampton* folgende potentielle Realisierungen der drei fraglichen phonologischen Variablen abgeleitet werden:

PV	Varianten
/ɔ:/	[ɔʌ, ɔ:, ɔɹ]
/θ/	[θ, s, t]
/ə/	[ə, ɔ]

Derselbe Zusammenhang läßt sich bereits in Form einfacher phonologischer Regeln ausdrücken:

/ɔ:/	→	[ɔʌ]
/ɔ:/	→	[ɔɹ]
/θ/	→	[s]
/θ/	→	[t]
/ə/	→	[ɔ]

Zwar finden sich hier noch keinerlei Angaben über phonetische Kontexte der Regeln sowie über weitere Anwendungsbedingungen, die möglicherweise eine Rolle spielen; dennoch bildet eine solche Auswertung von PVs und ihrer Varianten bereits eine wichtige Arbeitsgrundlage zur Erstellung komplexerer Regeln, wie sie in Kapitel 7 dieser Arbeit ausführlich beschrieben werden.

5.5 Sprachdatensammlung: Grenzen, Probleme, Ausblick

Eine der wesentlichen Beschränkungen der Datenbank in ihrem jetzigen Stadium wurde in den vorausgehenden Abschnitten bereits angesprochen. Sie betrifft die

Sprecherdistribution hinsichtlich des Durchschnittsalters und der verhältnismäßig hohen L2-Kenntnisse bei einigen Zielsprachen. Entsprechende Erweiterungen der Datenbank durch eine weitere Sprechergruppe (höheres Alter, geringere Kenntnisse der Zielsprachen Deutsch und Englisch) sind also zweifellos sinnvoll, um die Repräsentativität der Sprachdatensammlung zu verbessern. Für die Muttersprache Italienisch konnte eine solche Erweiterung in einem späten Stadium der vorliegenden Studie bereits durchgeführt werden; Angaben zu diesen Daten finden sich in Kap. 8 dieser Arbeit.

Während jedoch das Problem der Sprecherdistribution und Datenrepräsentativität durch einen weiteren Ausbau der Datensammlung zumindest im Prinzip lösbar ist, liegt eine größere forschungspraktische Herausforderung im Phänomen der *intra-individuellen Variation*. Damit ist die Erscheinung gemeint, daß ein und derselbe Sprecher in verschiedenen Situationen deutlich unterschiedliche Aussprachevarianten gleicher fremdsprachlicher Wörter oder anderer Einheiten (Sätze, Phrasen, idiomatische Wendungen u.a.) produzieren kann.

Wie verschiedene Einzelstudien nahelegen, ist es ein spezifisches Merkmal nicht-muttersprachlicher Aussprache, daß etwa die Kommunikationssituation und/oder die individuelle Sprecherdisposition einen deutlichen Einfluß auf die (u.a. auch phonetische) Performanz des Sprechers ausüben (siehe hierzu auch Kap. 2.5). Selbst bei einem einzelnen Sprecher ist daher also mit einer gewissen Variationsbreite zu rechnen, die von einer annähernd zielsprachlichen Aussprache bei stark kontrollierter, formeller Sprechweise bis hin zu einer deutlich akzentgefärbten Aussprache in informeller Spontansprache reichen kann. Im Verbund mit dem ohnehin erwartbaren Phänomen *inter-individueller Variation* potenziert sich damit das Spektrum möglicher Variation innerhalb einer L1-Sprechergruppe weiter.

Innerhalb des hier gewählten Versuchsdesigns wurde intra-individuelle Variation nicht bzw. nur in minimaler Weise berücksichtigt, indem bei der Darbietung der Stimuli in Teil (a) der Versuche (Leseaufgabe) einzelne Stimuli mit zeitlichem Abstand wiederholt wurden. Dies ermöglicht es, abweichende Realisierungen bzw. inkonsistentes Sprecherverhalten zumindest ansatzweise zu ermitteln. Es ist jedoch anzunehmen, daß in ungesteuerter (d.h. nicht im Experiment erhobener) Sprachproduktion derartige intra-individuelle Abweichungen in weit höherem Maße auftreten werden, als es hier erfaßt werden konnte.

Ein solcher Mangel an Natürlichkeit ist aber als ein generelles Defizit empirischer Sprachforschung anzusehen, bei der durch das experimentelle Umfeld eine artifizielle sprachliche Situation etabliert wird. Es handelt sich letztlich um eine Erscheinungsform des sog. „Beobachterparadoxons“, demzufolge der Forscher gerade durch seine Beobachtung das beobachtete Objekt beeinflusst und dadurch zu anderen Daten gelangt, als es ohne Beobachtung der Fall wäre.

Als Ausblick sei daher im folgenden in aller Kürze ein Szenario skizziert, mit dem dieses Problem minimiert werden könnte, ohne auf ein kontrolliertes Versuchsdesign zu verzichten. Die Rede ist dabei von sog. *Map Tasks* (Brown, Anderson, Shillcock & Yule 1984), die für eine Untersuchung innerhalb der hier besprochenen Domäne – der Aussprache von Eigennamen durch Nicht-Muttersprachler – modifiziert werden könnten. Die Grundidee dieses Versuchsdesigns besteht darin, zwei Versuchspersonen in einen aufgabenorientierten natürlichen Dialog zu verwickeln, um damit erstens die kognitive Last von der reinen sprachlichen Form auf die kommunikative Aufgabe zu lenken und zweitens Sprachverwendung in einem weitgehend natürlichen Kontext beobachten zu können, ohne dabei auf ein kontrolliertes Versuchsdesign verzichten zu müssen. Bei einer solchen typischen *Map Task* verfügen zwei Personen – der *Instruktor* und der *Instruierte* – z.B. über eine einfache gezeichnete Landkarte, auf der sich verschiedene Markierungs- und Orientierungspunkte (*landmarks*) befinden, etwa Bäume, Häuser, Autos etc.. Auf der Karte des Instructors ist eine Route eingezeichnet, welche auf der Karte des Instruierten nicht vorhanden ist. Aufgabe des Instructors ist es nun, dem Instruierten diese Route anhand der *landmarks* zu erklären. Die Karten sind dabei jedoch nicht identisch, sondern gezielt mit Abweichungen versehen, um die Probanden an definierten Punkten der Karte in eine sprachliche Diskussion zu verwickeln und damit von der reinen sprachlichen Form abzulenken.

Ein derartiges Versuchsdesign könnte für Untersuchungen zur Aussprache von Eigennamen dahingehend modifiziert werden, daß z.B. fremdsprachliche Ortsnamen als Markierpunkte dienen. Auf diesem Wege kann deren Aussprache untersucht werden, ohne daß den Probanden die Möglichkeit gegeben wird, ihre volle Aufmerksamkeit auf die sprachliche Form zu richten. Es sind jedoch zumindest informelle Vorstudien erforderlich, um die Eignung eines solchen experimentellen Vorgehens zunächst zu prüfen. So ist es denkbar, daß Probanden in einem solchen Szenario vollkommen idiosynkratische Aussprachevarianten (Abkürzungen o.ä.) einführen, um auf möglichst einfache und schnelle Weise einen Verständigungserfolg herbeizuführen. Solche Formen wären wiederum für eine detaillierte phonetische Studie nur von begrenztem Wert. In der vorliegenden Arbeit wurde daher – trotz der damit verbundenen möglichen Einschränkungen hinsichtlich der Natürlichkeit der erhobenen Sprachdaten – auf das in diesem Kapitel beschriebene Versuchsdesign zurückgegriffen.

6 Fehlertypen und Implikationen für eine regelbasierte Modellierung¹¹

6.1 Grundlage und Motivation der Fehlertypologie

Ursachen für Aussprachefehler können auf verschiedenen linguistischen Ebenen und in außerlinguistischen Bedingungen angesiedelt sein, die in vielfältiger Weise ineinandergreifen. Die Zuordnung bestimmter Fehler zu einer dieser Ebenen ist dabei nicht allein von theoretischem Interesse; sie ist auch grundlegend für die angestrebte Modellierung akzentgefärbter Aussprache, da spezifische Fehlertypen jeweils spezifische Methoden für die regelbasierte Nachbildung erfordern. Der Entwurf einer Typologie der linguistischen Ursachen für Aussprachefehler, d.h. eine Klassifizierung der auftretenden Fehler nach linguistischen Ebenen (z.B. Orthographie, Phonologie, Morphologie) bildet also eine wesentliche Grundlage für eine linguistisch sinnvolle und ökonomische Formulierung von Regeln.

Im folgenden Kapitel wird der Versuch unternommen, aus den in den Sprachdaten beobachteten Aussprachebesonderheiten und -fehlern eine grundlegende Taxonomie zu entwickeln, welche die auftretenden Aussprachevarianten nach ihren linguistischen Ursachen klassifiziert. Eine solche Taxonomie erfüllt dabei nicht allein eine deskriptive Funktion, sondern bildet auch die unmittelbare Grundlage für das in Kap. 7 beschriebene Regelsystem und dessen programmtechnische Realisierung.

Im Unterschied zu der verbreiteten Praxis, Aussprachefehler bzw. Abweichungen von der L2-Norm ausschließlich aus phonetisch-phonologischer Perspektive zu betrachten, wird im folgenden auch der Bezug zur *orthographischen Ebene* als eigenständige bzw. mit der lautlichen Ebene zusammenwirkende Fehlerquelle hergestellt werden. Obwohl bereits in einer frühen Phase der kontrastiven Linguistik auf den Status der Schriftaussprache als eigenständige Fehlerquelle hingewiesen wurde (cf. z.B. Lado 1957: 19-21), ist diese Ebene in späteren Untersuchungen weitgehend vernachlässigt worden (Ausnahmen bilden z.B. Szulc 1973; Glück 1993; Hirschfeld 1983; Altenberg & Vago 1987). Dies ist insofern als

¹¹ Teile dieses Kapitels beruhen auf meinem Artikel: Schaden, S. (2003c): "Non-native pronunciation variants of city names as a problem for speech technology applications". In: *Text, Speech and Dialogue. Proceedings of the Sixth International Conference TSD 2003*. Berlin u.a.: Springer, 229-236.

Defizit zu werten, als die geschriebene Sprachform oftmals eine wesentliche Basis der Vermittlung bzw. des Erlernens einer L2 liefert und somit schon in einem frühen Stadium als potentielle Interferenzquelle wirksam werden kann. So konnten auch in der vorliegenden Sprachdatensammlung häufig Aussprachefehler beobachtet werden, die eng mit der Schriftaussprache zusammenhängen. Das Versuchsdesign der Sprachaufzeichnungen ermöglicht dabei einen unmittelbaren Vergleich der Aussprachevarianten, die bei orthographischer und akustischer Präsentation der Stimuli produziert wurden. Dies erlaubt Rückschlüsse auf den Einfluß der Schriftaussprache auf die Aussprachevarianten und eine Abgrenzung der hier auftretenden Fehler von ausschließlich phonetischen Interferenz-Effekten.

6.2 Substitutionen, Einfügungen, Elisionen

6.2.1 Vorbemerkung: Die segmentale Betrachtungsweise

Es ist zweifellos eines der hervorstechenden phonetischen Kennzeichen akzent-behafteter Aussprache, daß einzelne Sprachlaute nicht entsprechend der kanonischen L2-Aussprache artikuliert werden, sondern durch andere Laute ersetzt werden. Hieraus resultieren charakteristische Abweichungen von der kanonischen Lautung, welche in Verbindung mit anderen Faktoren (z.B. spezifischen prosodischen Mustern) den abweichenden Gesamtklang akzent-behafteter Aussprache konstituieren.

Diese Vorstellung eines „Lautersatzes“ beinhaltet jedoch stets eine gewisse Idealisierung. So legt der Terminus nahe, daß an wohldefinierten Positionen einer sprachlichen Äußerung einzelne Lautsegmente der Gesamtsequenz gewissermaßen entnommen und durch andere Segmente - mit abweichender lautlicher Qualität - ersetzt werden. Aus physikalischer bzw. instrumentalphonetischer Perspektive stellt sich jedoch jede lautsprachliche Äußerung zunächst als ein Lautkontinuum ohne eindeutig bestimmbare Grenzen zwischen seinen Bestandteilen dar. Die Perspektive der sprachwissenschaftlich orientierten Phonologie ist dagegen eine andere: Sie geht von der Annahme aus, daß sich dieser Lautstrom in diskrete Einheiten gliedern und einer segmentalen Analyse unterziehen läßt. Je nach theoretischer oder praktischer Orientierung sind dies z.B. *Phoneme*, *distinktive Merkmale*, *Silben* oder andere Einheiten. Goldsmith (1976) bezeichnete dies als die *absolute slicing hypothesis*, welche der modernen Phonologie zugrunde liege: Ihr zufolge besteht eine lautsprachliche Äußerung aus linear an-

geordneten, voneinander trennbaren Segmenten, welche in Form ungeordneter phonologischer Merkmalsbündel vorliegen und als solche analysierbar sind.

Es wurde vielfach darauf hingewiesen, daß diese Hypothese der lautlichen Form menschlicher Sprache nicht ohne weiteres gerecht werde, da diese nicht nur durch vielfältige Koartikulations- und Assimilationsphänomene, sondern auch durch suprasegmentale Merkmale wie etwa ihre prosodische und rhythmische Strukturen gekennzeichnet sei.

Die Frage, ob und wie weit eine phonologisch-segmentale Gliederung menschlicher Lautsprache grundsätzlich als Basishypothese sinnvoll ist, kann hier nicht in gebührendem Umfang diskutiert werden. Jedoch sprechen verschiedene Gründe durchaus auch *für* eine solche Perspektive. Evidenz dafür liefern z.B. Phänomene wie sprachliche Fehlleistungen („Versprecher“), die offensichtlich auf einer segmentalen Vertauschung von Lauten beruhen (z.B. „Schlurzkuß“, „leitergeweitet“, „Corritical pollectness“), sowie sprachhistorische und dialektale Einzelphänomene wie die Metathese (Umstellung) von Lautsegmenten (*Brunnen* ~ *Born*; *Balg* ~ *Blag*; *Bernstein* ~ *brennen*). Auf einer allgemeineren Ebene kann darüber hinaus auch argumentiert werden, daß die gesamte Entwicklung von Alphabetschriften letztlich eine vorausgehende „phonologische“ Denkweise im weitesten Sinne voraussetzt, da ihr eine wie auch immer geartete Segmentierung lautlicher Einheiten historisch vorausgehen muß (cf. Ternes 1999: 26/27).

Zwar unterliegt eine Betrachtung fremdsprachlicher Akzente unter einem ausschließlich segmentalen Gesichtspunkt prinzipiell den gleichen Beschränkungen und Vorbehalten, die einer segmentalen Betrachtung von Lautsprache an sich entgegengebracht werden können. So sind akzentbehafte Äußerungen in hohem Maße auch durch spezifische prosodische Muster geprägt, die kennzeichnend sowohl für das Vorliegen eines Akzents an sich wie auch für die muttersprachliche Herkunft des Sprechers sind (cf. Jilka 2000).

Dennoch erscheint die Arbeitshypothese der Segmentierbarkeit als Ausgangspunkt einer Untersuchung durchaus geeignet: Zum einen eröffnet dies die Möglichkeit, auf erprobte Beschreibungsverfahren phonologischer Analysen zurückzugreifen. Zweitens fügt sich eine segmentale Beschreibung von Akzenten relativ nahtlos ein in gängige Methoden der Sprachtechnologie und kann daher leicht in bestehende Systeme integriert werden. Zudem schließt diese Perspektive eine Erweiterung um suprasegmentale Gesichtspunkte zu einem späteren Zeitpunkt keineswegs aus.

6.2.2 Typen des Lautersatzes

Wenn im folgenden allgemein von einem „Lautersatz“ die Rede ist, so schließt dieser Begriff eine Reihe von Interferenztypen ein, die in anderen in der Litera-

tur vorgeschlagenen Fehler-Typologien z.T. in viele weitere Unterkategorien gegliedert sind. So entwickelt beispielsweise Ternes (1976: 23-55) eine Typologie der lautlichen Interferenz, bei der insgesamt sieben Typen des Lautersatzes unterschieden werden. Eine solche Feingliederung soll hier nicht vorgenommen werden. Vielmehr wird zunächst eine vereinfachte Typologie zugrunde gelegt, die sämtliche Interferenztypen umfaßt. Unterschieden werden dabei lediglich (a) Substitution, (b) Einfügung und (c) Tilgung von Lauten. Die elementaren Operationen lassen sich dabei wie folgt formalisieren:

- | | | | | |
|-----|---------------------|---|---|----|
| (a) | <i>Substitution</i> | X | → | Y |
| (b) | <i>Einfügung</i> | X | → | XY |
| (c) | <i>Tilgung</i> | X | → | ∅ |

Über die Beschaffenheit der ersetzten Laute und der Ersatzlaute sind damit noch keinerlei Aussagen getroffen. Beispielsweise ist die Frage nach dem phonemischen oder phonetischen Status der Laute X und Y – d.h. die Frage, ob ein Phonemsprung oder nur eine allophonische Variation aus der Lautersetzung resultiert – hiermit nicht unmittelbar berührt. Statt dessen sollen Lautsubstitutionen zunächst nach folgenden Kategorien unterschieden werden:

- (1) **Substitution durch phonetische Äquivalente.** Ein L2-Laut wird vom Sprecher durch einen L1-Laut ersetzt, der dem ersetzten Laut phonetisch sehr ähnlich, aber nicht identisch mit ihm ist. Hierbei handelt es sich oftmals um das Resultat einer *Äquivalenzklassifizierung*, wie sie in Kapitel 2.4 beschrieben wurde: Ein L1-Laut wird auf der Basis ausgeprägter perzeptiver oder artikulatorischer Ähnlichkeit als Ersatz für den zielsprachlichen L2-Laut herangezogen.
- (2) **Substitution durch Interferenzformen.** Falls zum betreffenden L2-Ziellaut kein identischer oder ausgeprägt ähnlicher L1-Laut als Äquivalent existiert, können oftmals Lautersetzungen beobachtet werden, wie sie im Rahmen der „Kontrastivhypothese“ (siehe Kapitel 2.2) als Ergebnis von negativem Transfer und Interferenz diskutiert wurden. Hier treten Ersetzungen wie etwa deutsch [s] für englisch [θ] auf (*thick* → [sɪk]). Die resultierenden Abweichungen von der korrekten L2-Form sind dabei in der Regel stärker ausgeprägt als im Falle einer Äquivalenzklassifizierung.
- (3) **Substitution durch phonetische Approximationen.** Oftmals treten phonetische Approximationen und Mischformen auf, die nicht eindeutig dem L1- oder dem L2-Inventar zugeordnet werden können. Dies ist beispielsweise der Fall, wenn der Sprecher den Ziellaut zwar perzeptiv von ähnlichen L1-Lauten abgrenzen kann, ihn jedoch artikulatorisch nicht realisieren kann.

(4) **Orthographisch motivierte Substitutionen.** Bei diesem Sonderfall (der später in diesem Kapitel ausführlicher dargestellt wird) ist die Wahl des Ersatzlautes durch die orthographische Repräsentation des Ziellautes beeinflusst. Die auftretenden Ersatzlaute lassen sich somit nicht rein phonetisch erklären. Der Sprecher wendet hierbei in der Regel Schriftaussprache-Regeln seiner L1 auf zielsprachliches Material an. Dieser Fall tritt hauptsächlich bei gelesener Sprache auf.

Im folgenden werden einige charakteristische Beispiele für Lautersetzungen, -elisionen und -einfügungen gegeben, die in den vorliegenden Sprachdaten häufig auftraten. Dabei ist es jedoch nicht das Ziel, eine detaillierte Analyse jedes beobachteten Lautersatzes zu geben. Vielmehr stehen die Beispiele stellvertretend für spezifische Fehlertypen, deren Modellierung Bestandteil des später entwickelten Regelsystems ist. Dieses Vorgehen wurde gewählt, weil eine solche Typologie der Fehler als theoretische Grundlage dieses Regelsystems von weitaus größerer Bedeutung ist als die jeweiligen phonetischen Inhalte, in denen sich die Fehlertypen manifestieren.

6.2.3 Alternationen der Vokalqualität

Eine in allen untersuchten Sprachrichtungen häufig auftretende Form der Lautsubstitution ist eine gegenüber der L2-Aussprache veränderte Vokalqualität. So ersetzen Muttersprachler des Deutschen englisches [æ] oftmals durch deutsches [ɛ], wodurch Varianten wie *Aston* → [ɛstən] oder *Aberdeen* → [ɛbədi:n] auftreten. Als phonetisches Einzelphänomen bedarf dieses Beispiel an dieser Stelle keiner umfassenderen Diskussion. Es handelt sich um ein charakteristisches Beispiel einer Äquivalenzklassifizierung, bei der ein phonetisch sehr ähnlicher, aber nicht identischer Ersatzlaut für einen L2-Laut eintritt. Dennoch lassen sich an diesem einfachen Beispiel einige wichtige generelle Prinzipien illustrieren.

Erstens zeigt sich hierin, daß allein die artikulatorische Ähnlichkeit kein zuverlässiger Indikator für den gewählten Ersatzlaut ist. Häufig weisen zwei oder mehrere L1-Laute den gleichen Grad artikulatorischer Ähnlichkeit zum Ziellaut auf, so daß prinzipiell beide Laute gleichermaßen als Ersatzlaute in Frage kommen. So liegt der vordere Vokal [æ] (d.h. der Ziellaut) auf der artikulatorischen Dimension *offen-geschlossen* zwischen den vorderen Vokalen [ɛ] und [a]. Somit käme der Laut [a] mit gleicher Berechtigung als Ersatzlaut für englisch [æ] in Frage. Bei deutschen Sprechern sind jedoch Varianten wie *Aston* → [astən] oder *Aberdeen* → [abədi:n] kaum zu beobachten; der gewählte Ersatzlaut ist hier fast ausnahmslos [ɛ]. Bei französischen Sprechern hingegen verändert sich dieses Verhältnis deutlich: Hier wird [æ] weitaus häufiger als [a] wiedergegeben. Das Beispiel zeigt somit, daß nicht allein die isoliert betrachtete lautliche Ähnlichkeit

als Kriterium für die Wahl des Ersatzlautes herangezogen werden kann, sondern die L1 des Sprechers bzw. ihr Phonemsystem hierbei eine wichtige Rolle spielt. Für eine regelbasierte Modellierung bedeutet dies, daß die von den Regeln zu generierenden Ersatzlaute nicht auf der Basis allgemeiner phonetischer Prinzipien (z.B. Lautähnlichkeit) bestimmt werden können, sondern stets als spezifische Regeln für eine bestimmte Sprachrichtung formuliert werden müssen.

Eine Substitution von Lauten muß nicht notwendigerweise die Form einer Ersetzung eines einzelnen Segmentes durch ein anderes annehmen. Es kann auch der Fall beobachtet werden, daß eine Folge mehrerer Segmente durch ein einzelnes Segment bzw. umgekehrt ein einzelnes Segment durch eine mehrgliedrige Sequenz ersetzt wird.

So realisieren Sprecher des Englischen die deutsche Lautsequenz [ɔʁ] (z.B. in *Dorf*) oftmals durch den akustisch ähnlichen englischen Vokal [ɔ:]. Begünstigt wird diese Substitution dabei durch orthographische Einflüsse. In beiden Sprachen sind die fraglichen Laute [ɔʁ] (dt.) und [ɔ:] (engl.) durch die Graphemsequenz <or> repräsentiert (vgl. *Dorn* vs. *thorn*). Wird das L2-Material gelesen, so ist es schon aufgrund dieser orthographischen Analogie wahrscheinlich, daß ein Lautersatz [ɔʁ] → [ɔ:] stattfindet¹². Das Beispiel zeigt, daß die orthographische Repräsentation relevant für die Wahl des Ersatzlautes sein kann, wenn aus einer rein phonetischen Perspektive mehrere Alternativen in Frage kommen.

Eine Besonderheit des genannten Beispiels einer Ersetzung [ɔʁ] → [ɔ:] liegt darin, daß die Relation von ersetzttem Laut und Ersatzlaut zwischen Ausgangssprache und Zielsprache *umkehrbar* ist. Denn in gleicher Weise neigen deutsche Sprecher dazu, englisches [ɔ:] durch deutsches [ɔʁ] zu ersetzen, so daß z.B. englisch *thorn* als [θɔʁn] wiedergegeben wird. Eine solche Umkehrbarkeit ist aber als Ausnahmefall zu betrachten; keinesfalls ist es möglich, durch eine bloße Umkehrung von Ersetzungsregeln der Sprachrichtung L1 → L2 adäquate Regeln für die Richtung L2 → L1 zu gewinnen.

Das bei der Gewinnung der Sprachdaten verwendete Untersuchungsdesign ermöglicht es zu bestimmen, ob und in welchem Maße die orthographische Repräsentation eines zielsprachlichen Lautes auf seine Realisierung durch Nicht-Muttersprachler Einfluß nimmt. Während die oben skizzierte Substitution [ɔʁ] → [ɔ:] durch die Orthographie zumindest begünstigt wird, ist bei anderen Lautersetzungen kein solcher Einfluß feststellbar. So ist beispielsweise bei französischen Sprechern mit der Zielsprache Deutsch eine ausgeprägte Tendenz zu

¹² Das genannte Beispiel geht dabei vom Standard des britischen Englisch aus. Bei rhotazierten Varietäten des Englischen (z.B. *Standard American English*) muß die Regel [ɔʁ] → [ɔ:] umformuliert werden in [ɔʁ] → [ɔr].

beobachten, silbische Konsonanten wie [ŋ] oder [l] in wortfinaler Stellung durch Schwa-Formen wie [əŋ] oder Vollvokale wie [œŋ], [ɛŋ] zu ersetzen (z.B. in *Vöhringen* → [fœʋŋgəŋ] oder *Apen* → [apœŋ]).

Der Vergleich zwischen den Realisierungen bei gelesenen und nachgesprochenen Varianten zeigt dabei, daß hier ein von der Orthographie unabhängiger Fehler vorliegt, der bei akustischer Darbietung ebenso auftritt wie bei graphemischer Darbietung der betreffenden Stimuli. In Kapitel 7.1 wird sich zeigen, daß eine solche Klassifizierung der auftretenden Fehler nicht allein von taxonomischem Interesse ist, sondern unmittelbar auf die gewählte Methode der regelbasierten Modellierung solcher Fehler zurückwirkt.

6.2.4 Alternation allophonischer Realisierungen

Oftmals sind die beobachteten Lautsubstitutionen das Resultat allophonischer Realisierungsvarianten, die in der L1 des Sprechers die Standard-Realisierung eines spezifischen Phonems bilden. Ein Beispiel hierfür ist das Phonem /R/ mit seinen vielfältigen möglichen Realisierungen: Das Spektrum reicht hier vom uvularen Frikativ [ʁ] über die Vibranten [r] und [rr] bis zum alveolaren Approximanten [ɹ]; darüber hinaus treten auch vokalische Realisierungen des Phonems als [ø] und [v] auf.

Häufig – jedoch keineswegs immer – wird die in der L1 des Sprechers vorherrschende Standard-Realisierung des /R/ auf die Zielsprache übertragen. So findet man bei Sprechern des Englischen in verschiedenen Zielsprachen häufig die Variante [ɹ], bei Sprechern des Deutschen einen Transfer des [ʁ] auf die Zielsprache, und bei Sprechern des Italienischen und Spanischen eine Realisierung als einfacher [r] oder doppelter (bei L1 Italienisch) Vibrant [rr]. Allein durch das breite Spektrum möglicher /R/-Realisierungen entsteht hierbei eine Vielzahl potentieller Aussprachevarianten. Im Rahmen einer regelbasierten Modellierung kann dieses Variationsspektrum zunächst erfaßt werden, indem Regeln formuliert werden, in denen die zielsprachlich „korrekte“ Realisierung des /R/-Phonems durch die Standardrealisierung in der jeweiligen L1 des Sprechers ersetzt wird. Für das deutsche Phonem /R/ mit seiner Standardrealisierung [ʁ] würde diese beispielsweise wie folgt aussehen:

Deutsch	L1 Englisch	L1 Italienisch
[ʁ]	[ʁ] → [ɹ]	[ʁ] → [r] oder [ʁ] → [rr]

Jedoch ist die sprachenübergreifende Realisierung des /R/ zusätzlicher Variation unterworfen, weil häufig bereits innersprachliche (lokale, dialektale) und auch distributionelle (stellungsbedingte) Unterschiede in der Realisierung auftreten.

Zum Teil wirken diese sich auch auf die Aussprache fremdsprachlichen Materials aus. So existieren auf den britischen Inseln verschiedene regionale Dialekte (u.a. in Wales, Irland, Schottland) in denen nicht der alveolare Approximant [ɹ], sondern der Vibrant [r] als Standard-Realisierung des /R/ gilt (cf. Gimson 1990: 207ff.). Eine ebenfalls stark ausgeprägte dialektal-regionale Variabilität des /R/ besteht im Deutschen, das neben der heute als Standard geltenden Realisierung als uvulares [ʁ] auch die regionale Variante [r] sowie verschiedene andere Subvarianten aufweist (cf. hierzu ausführlich Wiese 2003).

Die in dieser Arbeit erhobenen Sprachdaten zeigen deutlich, daß diese Form innersprachlicher Variabilität als Einflußgröße bei der phonetischen Beschaffenheit akzentbehaffeter Aussprachevarianten einer Fremdsprache wirksam werden kann. So traten bei Sprechern mit irischer und walisischer Herkunft bei der Aussprache von deutschen und französischem Material häufig Realisierungen des Phonems /R/ als [r] auf. Dies stellt nicht nur eine auffällige Abweichung von der kanonischen L2-Realisierung dar, sondern weicht darüber hinaus auch von der erwarteten Aussprache mit englischem [ɹ] deutlich ab. Eine einfache sprachspezifische Regel wie die oben genannte Ersetzung [ʁ] → [ɹ] vermag derartige Abweichungen nicht zu erfassen.

Darüber hinaus ist die Situation denkbar, daß es durch das spezifische phonetische Verhältnis von dialektalem Hintergrund des Sprechers und Zielsprache zu einer *Neutralisierung* des Akzents kommen kann. Dies wäre z.B. dann der Fall, wenn die Dialektbasis des Sprechers eine Realisierung einzelner Phoneme aufweist, die mit der zielsprachlichen Realisierung identisch oder ihr ähnlich ist. So würden beispielsweise die genannten englischen Sprecher mit irischem oder walisischem Dialekthintergrund mit ihrer Realisierung des /R/-Phonems als [r] in den Zielsprachen Italienisch oder Spanisch eine annähernd korrekte zielsprachliche Realisierung erzielen.

Beispiele wie diese zeigen, daß eine Formulierung von Lautersatzungs-Regeln allein auf der Basis einer angenommenen Normaussprache die sprachliche Realität nicht immer hinreichend erfassen kann. Wie bereits in Kapitel 2 angesprochen wurde, bestehen vielfältige Einflüsse sprecherspezifischer Variablen auf die zu erwartende Aussprache fremdsprachlichen Materials. Das hier genannte einfache Beispiel des dialektalen Einflusses auf die Realisierung der einzelnen phonetischen Variable /R/ vermag zu illustrieren, mit welcher erheblichen Variationsbreite grundsätzlich zu rechnen ist, wenn die Realisierung *mehrerer* phonetischer Variablen durch *mehrere* Sprecher untersucht wird.

6.2.5 Elisionen

Oftmals tritt an die Stelle eines L2-Lautes *kein* Ersatzlaut – es erfolgt also eine vollständige Tilgung (Elision) des Ziellautes. Beispielsweise besteht – einem geläufigen Stereotyp zufolge – ein charakteristisches Merkmal eines ‚französischen Akzents‘ in der Elision des Lautes [h] in prävokalischer Position (*Haus* → [aus]). Aus einer kontrastiv-phonetischen Perspektive, wie sie in Kapitel 2.2 dargelegt wurde, ließe sich diese Beobachtung aus einer Interferenz zwischen den Lautsystemen des Französischen und z.B. des Deutschen ableiten: Da im Französischen [h] als Allophon der phonetischen Oberflächenform nicht existiert¹³, ist eine Elision dieses Lautes vorhersagbar.

Die in dieser Studie vorliegenden Sprachdaten zeigen jedoch, wie wenig verlässlich derartige Prädiktionen sind, wenn sie allein aus der Gegenüberstellung der Lautsysteme L1 und L2 abgeleitet sind. Kontrastiv-linguistischen Prognosen zufolge wäre in den Daten ein häufiges Auftreten von Formen wie *Hamburg* [ambʊʁk] oder *Heßberg* [ɛsbɛʁk] erwartbar. Tatsächlich liegt jedoch in der untersuchten Sprechergruppe der Anteil elidierter [h]-Laute bei lediglich 7,6%¹⁴. Eine interessante Beobachtung ergibt sich zudem bei einem Vergleich der /h/-Realisierungen bei (1) orthographischen Stimuli (d.h. bei gelesenen Material) und (2) akustischen Stimuli (d.h. bei nachgesprochenem Material): Gegenüber der Lesesituation sinkt die ohnehin geringe Zahl der [h]-Elisionen (7,6%) bei der Wiederholung gesprochener Stimuli weiter auf unter 4% ab.

Dieses Ergebnis begründet Zweifel an der Hypothese, daß es sich bei der [h]-Elision um ein rein phonetisches Interferenzprodukt handelt, welches sich allein aus dem Fehlen eines Lautes [h] im Französischen herleiten läßt. Vielmehr deutet es darauf hin, daß ein Einfluß der Orthographie wirksam wird: Möglicherweise ist ein Teil der [h]-Elisionen, die in verschiedenen L2 bei französischen Sprechern zu beobachten sind, vorwiegend auf eine Übertragung der französischen Schriftaussprache-Regeln auf die L2 zurückzuführen, denen zufolge orthographisches <h> nicht als eigener Laut erscheint. Zur Prüfung dieser Hypothese sind weitere Untersuchungen erforderlich; jedoch deuten die vorliegenden Daten

¹³ Umstritten ist dagegen der *phonemische* Status von /h/ im Französischen: Argumente für eine phonologische Existenz von /h/ mit einer Oberflächenrealisierung als ‚Nullphonem‘ berufen sich dabei in der Regel auf bestimmte *Liaison*-Phänomene im Französischen, welche mittels eines zugrundeliegenden, aber letztlich ‚stumm‘ realisierten Phonems /h/ erklärt werden können (cf. Schane 1968; Klein 1973).

¹⁴ Die vorliegende Analyse beschränkt sich auf [h] in wort-initialer Position, da eine [h]-Elision in anderen Kontexten (z.B. bei morphem-initialer Stellung in *Becht[h]eim*) als ein natürlicher Assimilationsprozeß betrachtet werden kann, der auch in nicht-akzentgefärbter Aussprache auftritt.

darauf hin, daß die Orthographie hier einen deutlichen Einfluß auf die produzierten Aussprachevarianten ausübt.

Eine Tilgung von phonetischen Segmenten unter Berücksichtigung ihrer orthographischen Repräsentation ist als eigenständiger Regeltyp in Kapitel 7 beschrieben. Fehler wie die genannte [h]-Elision können mittels dieses Regeltyps modelliert werden.

6.2.6 Einfügungen

Sowohl Elisionen als auch Einfügungen von Lauten sind oftmals das Resultat von abweichenden Silbenstrukturen in L1 und L2. Einzelsprachen weisen erhebliche Unterschiede hinsichtlich Aufbau und Komplexität ihrer möglichen Silben und deren Verkettung auf. Entsprechende Aussprachefehler manifestieren sich in einem Transfer der muttersprachlichen Silbenstruktur auf die Zielsprache. Eine häufige Folge eines solchen Transfers sind *Einfügungen* von Lauten, die letztlich für den Sprecher die Funktion erfüllen, die Silbenstruktur von L2-Material seiner L1 anzugleichen.

Wells (1999) führt in diesem Zusammenhang einige charakteristische Beispiele von Sprechern der L1 Japanisch bei der Produktion der Zielsprache Englisch an. Hier besteht die ausgeprägte Tendenz, einfache Konsonanten oder komplexe Konsonantengruppen im Anlaut oder Auslaut englischer Wörter durch Einfügung zusätzlicher Vokale der japanischen Silbenstruktur anzugleichen, etwa in *play*, das als [ptuleɪ] realisiert wird, ebenso wie in *school* [stuku:l] oder *cheese* [tʃi:zu]. Diese Adaption – und hierin liegt eine Besonderheit dieser phonetischen Interferenz-Erscheinungen – geht dabei über den individuellen Sprechakt hinaus: Denn englische Lehnwörter werden bei ihrer dauerhaften Aufnahme in das Japanische oftmals hinsichtlich ihres Silbenbaus umstrukturiert – so erscheint englisch *screen* als japanisch *sukuriin*.

Derartig auffällige Resyllabifizierungen konnten in den hier vorliegenden Sprachdaten nicht beobachtet werden. Dennoch kommt es auch bei den untersuchten europäischen Sprachen zu Interferenzen der Silbenstruktur und in der Folge zu Einfügungen zusätzlicher Laute. Die auffälligsten Beispiele hierfür liefern Sprecher der L1 Italienisch mit den Zielsprachen Englisch oder Deutsch. Während das Deutsche relativ komplexe Konsonantengruppen im Silbenauslaut erlaubt (z.B. in *Strumpf*, *wünschst*, *Arzt*), weist das Italienische eine einfachere Struktur der Silbenkoda auf. Entsprechend kann bei einigen Sprechern des Italienischen die Tendenz beobachtet werden, auslautende Silben des Deutschen in Richtung einer vereinfachten Silbenbildung zu modifizieren. Dies tritt am deutlichsten bei Obstruenten im Wortauslaut in Erscheinung, wie sie in Namen wie *Heßberg*, *Thränitz*, *Otzing*. Die Sprecher neigen hier dazu, der Silbenkoda einen

vokalischen Laut hinzuzufügen, so daß häufig Varianten wie *Thränitz* [trenitsə] oder *Otzing* [otsɪŋə] auftreten. Zumeist wird dieser Laut kurz und stark reduziert realisiert (als zentralisierter Vokal [ə]). Es kann eine deutliche sprecherübergreifende Tendenz zu dieser Form der Resyllabifizierung beobachtet werden.

Die Konsequenz dieses Sachverhalts für eine regelbasierte Modellierung ist u.a. darin zu sehen, daß die syllabische Struktur von Wörtern als Anwendungsbedingung in einzelne Regeln eingehen sollte. Aussprachefehler wie die genannte Einfügung von [ə] in wortfinaler Position können zwar prinzipiell als Lautersatz modelliert werden (im Sinne einer Einfügung eines Segmentes an einer „leeren“ Position). Jedoch ist es hierzu erforderlich, Elemente wie z.B. Silbengrenzen oder Wortgrenzen in den Regelkontext einzubringen, die keine phonetischen „Segmente“ im üblichen Wortsinne sind.

Das in Kapitel 7 vorgestellte Regelsystem beinhaltet eine solche Möglichkeit erst in Ansätzen, da eine Voraussetzung für die Modellierung dieses Fehlertyps darin besteht, zunächst eine automatische Syllabifizierung der phonetischen (und möglicherweise auch orthographischen) Eingabewörter vorzunehmen. Diese muß jedoch für *alle* in dieser Studie untersuchten Sprachen implementiert werden, um gewinnbringend eingesetzt werden zu können, womit der Rahmen der vorliegenden Arbeit bei weitem überschritten wäre. Dennoch liegt hierin eine wichtige potentielle Weiterentwicklung des vorgestellten Ansatzes.

6.2.7 Kontextfaktoren bei Lautsubstitutionen

In der bisherigen Darstellung wurden Lautersetzungen, -einfügungen und -elisionen so behandelt, als könnten sie in Form von Zuordnungsregeln formuliert werden, die unabhängig vom lautlichen Kontext des betroffenen Lautes sind. Eine Regel der Art „Ersetze den englischen Laut [æ] durch [ɛ]“ würde demnach ohne Einschränkungen, d.h. für *jedes Vorkommen* des Lautes [æ] gelten. Hierin liegt jedoch eine gezielte Vereinfachung, die der sprachlichen Realität nicht gerecht wird. Tatsächlich sind Lautersetzungen (ebenso Tilgungen und Einfügungen) oftmals an bestimmte lautliche Umgebungen gebunden, in denen die zu ersetzenden Laute auftreten. Beispiele für solche Kontextbindungen wurden bereits im vorausgehenden Abschnitt implizit eingeführt: Wenn z.B. die Rede davon war, daß bestimmte Lautersetzungen im Wortauslaut oder im Silbenanlaut auftreten, so ist damit bereits eine lautliche Umgebung spezifiziert, in der die Ersetzung auftritt.

Eine Vielzahl von Lautersetzungen, die in den vorliegenden Sprachdaten zu beobachten waren, ist an solche Kontexte gebunden. Dies bedeutet, daß Zuordnungen zwischen zielsprachlichen Lauten und ihren Ersatzlauten nur unter Angabe der phonetischen Umgebung des betreffenden Lautes formuliert werden

können. So wird deutsches [a:] von englischen Sprechern in einigen Kontexten durch den Diphthong [eɪ], in anderen Kontexten dagegen durch [ɑ:] substituiert. Möglich ist zudem der Fall, daß bestimmte Ersetzungen in einigen Umgebungen blockiert sind, d.h. prinzipiell nicht auftreten. Oftmals besteht bei derartigen durch den Kontext determinierten Fehlern ein enger Zusammenhang mit der Silbenstruktur der Zielsprache, welche von den Sprechern aufgrund von Beschränkungen der muttersprachlichen Silbenstruktur nur schwierig zu realisieren ist (siehe hierzu das oben genannte Beispiel der Einfügung eines Schwa-Lautes [ə] im Wortauslaut durch Sprecher des Italienischen).

Zahlreiche weitere Beispiele für derartige kontextabhängige Alternationen der Ersatzlaute lassen sich anführen. Entscheidend für den vorliegenden Zusammenhang ist jedoch zunächst der Sachverhalt, daß eine Modellierung dieser Ersetzungen mittels Regeln die Möglichkeit beinhalten muß, *kontextabhängige* Regeln zu formulieren. Eine solche Regel könnte beispielweise lauten: „Ersetze [a:] durch [eɪ], wenn ein Konsonant und mindestens ein Vokal folgen; ersetze dagegen [a:] durch [ɑ:] in allen anderen Kontexten“. In Kapitel 7 wird ausführlich dargestellt, in welcher Weise dieser Regeltyp realisiert wurde und welche Anwendungsmöglichkeiten dafür bestehen; dort finden sich auch weitere Beispiele kontextabhängiger Lautersetzungen.

6.3 Die Rolle phonologischer Regeln aus L1 oder L2

6.3.1 Sprachspezifische phonologische Regeln

Phonologische Beschreibungen, darunter vor allem solche, die in der Tradition der generativen Phonologie (Chomsky & Halle 1968) stehen, gründen sich zu meist auf die Annahme zweier Repräsentationsebenen der Lautstruktur einer Sprache: (a) einer abstrakten zugrundeliegenden Repräsentation (*underlying representation*, UR) einerseits und (b) einer phonetischen Oberflächenform andererseits. Die UR enthält dabei phonologische Basisformen, welche phonetisch noch unterspezifiziert sind. Sie sind nur mit solchen lautlichen Merkmalen gekennzeichnet, welche in der betreffenden Sprache als distinktiv (bedeutungsunterscheidend) wirksam werden. Die Oberflächenform (*surface representation*) – d.h. die tatsächliche lautliche Realisierung einer Äußerung – läßt sich mittels sprachspezifischer Derivationsregeln aus der UR ableiten.

So werden beispielsweise im Deutschen eine Reihe von Plosiven in bestimmten Kontexten aspiriert. Sie erscheinen in der phonetischen Oberflächenform – d.h. in ihrer konkreten lautlichen Realisierung – als aspirierte Formen [t^h], [k^h]

oder [p^h], etwa in *Tasse* [t^hasə], *Kamm* [k^ham] oder *Pater* [p^ha:tə]. Jedoch wird dieses Merkmal im Deutschen nicht zur phonologischen Opposition genutzt, d.h. es existiert kein Fall, in dem ein Wechsel [t^h] ~ [t] zu einer Bedeutungsänderung führt. Kenstowicz (1994: 57-59) trifft daher eine grundlegende Unterscheidung von *distinktiven* und *redundanten* Eigenschaften der Lautstruktur einer Sprache: Während z.B. der Sachverhalt, daß das deutsche Wort *Tasse* mit einem alveolaren, stimmlosen Plosiv [t] beginnt (und sich darin z.B. von *Masse* oder *Kasse* unterscheidet) nicht vorhersagbar ist – es sich also um ein *arbiträres* Merkmal handelt –, ist die Eigenschaft, daß dieser Plosiv in bestimmten Kontexten aspiriert wird, aus der Kenntnis der Lautstruktur des Deutschen *vorhersagbar*. Für den Wechsel [t] ~ [t^h] kann eine Regel formuliert werden, der zufolge der Laut [t] immer dann aspiriert wird, wenn er in wortinitialer Position vor einem Vokal auftritt:

$$[t] \quad \rightarrow \quad [t^h] \quad / \quad \#\# \quad _ \quad [+vokal]$$

In Kenstowicz' Terminologie handelt es sich bei der Aspiration um ein *redundantes* Merkmal der Lautstruktur des Deutschen, da es (a) nicht-distinktiv und (b) regelhaft prädictierbar ist. Mittels der genannten Regel kann für alle Vorkommen des Lautes [t] in einer entsprechenden Position die phonetische Oberflächenform [t^h] vorausgesagt werden. Es ist daher nicht erforderlich, dieses Merkmal in der UR des Wortes *Tasse* sowie aller anderen Wörter, welche [t] in dieser Position aufweisen, zu verzeichnen. Aus Gründen der Beschreibungsökonomie wird statt dessen eine UR /tasə/ angenommen, welche die o.g. phonologische Regel durchläuft und so in die Oberflächenform [t^hasə] überführt wird.

Da nun weiterhin zu beobachten ist, daß andere Laute des Deutschen sich analog zu [t] verhalten, d.h. ebenfalls von der Aspirationsregel betroffen sind, kann man in einem nächsten Schritt die genannte Regel in allgemeinerer Weise formulieren, so daß alle betroffenen Laute mit nur einer Regel erfaßt werden:

$$\left[\begin{array}{l} + \text{ plosiv} \\ - \text{ stimmhaft} \end{array} \right] \quad \rightarrow \quad \left[\begin{array}{l} + \text{ plosiv} \\ - \text{ stimmhaft} \\ + \text{ aspiriert} \end{array} \right] \quad / \quad \#\# \quad _ \quad [+vokal]$$

Mittels dieser Regel werden nun *alle* stimmlosen Plosive des Deutschen, sofern sie in der spezifizierten Position auftreten, mit dem Merkmal der Aspiriertheit versehen, so daß auch auf [k] oder [p] beginnende Wörter erfaßt sind und mit [k^h] oder [p^h] in der phonetischen Oberflächenform erscheinen.

Phonologische Regeln dieser Art sind stets sprachspezifisch. Zwar ist der Fall denkbar, daß zwei Sprachen dieselbe phonologische Regel aufweisen (die genannte Aspirationsregel ist beispielsweise im Deutschen und Englischen ähnlich),

generell sind sie jedoch nicht auf andere Sprachen übertragbar. So gilt beispielsweise die oben genannte Aspiriertheitsregel für das Französische nicht; hier erscheinen stimmlose Plosive in der betreffenden Position stets nicht-aspiriert (etwa in *pâte, comme, tard*). In einer Reihe phonologischer Arbeiten (cf. Kenstowicz & Kisseberth 1979; Singh & Ford 1986; Rubach 1984) wurde nun auf die Rolle sprachspezifischer phonologischer Regeln beim Zustandekommen eines fremdsprachlichen Akzents hingewiesen: Sprecher neigten oftmals dazu, redundante Eigenschaften (= systematisch auftretende phonologische Regeln) ihrer L1 unbewußt auf L2 zu übertragen. Ein solcher Transfer führe zu einigen charakteristischen lautlichen Merkmalen fremdsprachlicher Akzente.

Oftmals ist die Entscheidung darüber, ob es sich bei einem beobachteten Merkmal der Lautstruktur um eine systematisch auftretende phonologische Regel der betreffenden Sprache handeln soll oder nicht, stark vom Theorierahmen abhängig, in dem die Analyse stattfindet. So wurde beispielsweise innerhalb verschiedener Teilströmungen der generativen Phonologie die Unterscheidung zwischen (a) „genuinen“ *phonologischen Regeln*, also solchen Regeln, die allein auf der phonologischen Ebene operieren und (b) *morphonologischen Regeln*, also Regeln der Morphemalternation und deren phonologischer Effekte betont und auf die Relevanz dieser Unterscheidung für die L2-Phonologie hingewiesen. Dabei wurde herausgestellt, daß nur Kategorie (a) einem Transfer auf L2 unterworfen sein kann und sich in spezifischen L2-Aussprache Fehlern niederschlägt, während Kategorie (b) hiervon nicht betroffen sei (cf. Singh & Ford 1986; Rubach 1984).

Im Rahmen dieser Arbeit soll diese Erörterung nicht fortgeführt werden. Es soll aber zumindest informell angemerkt werden, daß eine Sichtung der Sprachdaten unter diesem Aspekt tatsächlich keinerlei Evidenz für einen Transfer morphonologischer Regeln auf die Zielsprache lieferte, wohl aber für einen Transfer „genuiner“ phonologischer Regeln. Dies ist jedoch insofern nicht überraschend, als morphonologische Regeln stets eines spezifischen morphologischen Kontextes bedürfen, in dem sie zur Anwendung kommen können. Ein solcher Kontext wird aber in L2-Material prinzipiell nicht auftreten, weil L1 über ein vollständig anderes Morphem-Inventar als L2 verfügt.

In den nachfolgenden Abschnitten werden einige Beispiele für Aussprachevarianten vorgestellt, die im oben beschriebenen Sinne als Transfer systematisch auftretender phonologischer Prozesse von L1 auf L2 interpretiert werden können. Dieser Fehlertyp weist dabei eine Reihe von Besonderheiten auf: Im Unterschied zu den bisher behandelten Fehlertypen betrifft ein Transfer phonologischer Regeln zumeist nicht einzelne Lautsegmente, sondern ganze *Lautklassen* (z.B. Obstruenten, Plosive, Nasale etc.), die in bestimmten Kontexten gleichartigen Alternationen unterworfen sind. Für eine regelbasierte Modellierung bedeu-

tet dies, daß entsprechende Derivationsregeln nicht als Ersetzungen einzelner Segmente, sondern als generalisierende Regeln für Phonemklassen formuliert werden können. Eine solche Regel kann z.B. besagen, daß alle Obstruenten in einer bestimmten Position des Wortes ihre Stimmhaftigkeit verlieren oder daß alle vorderen Vokale in einem spezifischen Kontext als Langvokale realisiert werden. Dies ermöglicht oftmals eine ökonomischere Regelformulierung, da nur eine Regel für die gesamte Lautklasse erstellt werden muß.

Eine weitere Besonderheit eines Transfers phonologischer Regeln ist darin zu sehen, daß entsprechende Regeln zum Teil unabhängig von der Zielsprache formuliert werden können. Während Lautsubstitutionen, wie sie bisher skizziert wurden, als spezifische Regeln für ein Sprachenpaar bzw. eine Sprachrichtung formuliert sind, werden systematische phonologische Regeln von Sprechern meist auf mehrere Zielsprachen zugleich projiziert. So wird z.B. die oben dargestellte Regel der Aspiration von Plosiven von deutschen Sprechern nicht nur auf das Französische, sondern in gleicher Weise auf das Englische, das Italienische sowie weitere Sprachen übertragen, sofern sie einen entsprechenden Kontext zur Regelanwendung bieten. Hinsichtlich einer regelbasierten Modellierung der betreffenden Aussprachefehler resultiert daraus die grundsätzliche Möglichkeit, die Regeln unabhängig von der jeweiligen Zielsprache zu formulieren bzw. die Regeln *zielsprachenübergreifend* anzuwenden.

Auffällig bei dem zur Diskussion stehenden Fehlertyp ist weiterhin, daß die durch einen Transfer phonologischer Regeln verursachten Aussprachefehler oftmals auch bei Sprechern mit sehr guten L2-Kenntnissen auftreten. Offensichtlich handelt es sich um eine Fehlerklasse, die relativ resistent gegenüber Lernprozessen ist. Dies kann aus theoretischer Sicht als ein stützendes Argument für die These herangezogen werden, daß es sich bei den redundanten phonologischen Prozessen einer Sprache um unbewußt angewandte Regeln handelt, die einen Teil einer internalisierten Kompetenz des Sprechers bilden – und gerade aus diesem Grund nur äußerst schwierig zu überwinden sind (cf. Kenstowicz & Kisseberth 1979: 31). Relevant für den vorliegenden Zusammenhang ist vor allem, daß Fehler dieses Typs auf *jeder* L2-Kompetenzstufe auftreten. Sie sind daher in einem Modell der Akzentstufen (siehe Kapitel 4.4) bereits auf einer sehr niedrigen Stufe, d.h. bei einem nur leicht ausgeprägten Akzent zu berücksichtigen. Generell ist festzustellen, daß ein Transfer phonologischer Regeln typischerweise keine gravierenden L2-Aussprachefehler, sondern nur relativ moderate Abweichungen nach sich zieht. Dessen ungeachtet bilden die so verursachten Aussprachefehler ein charakteristisches Merkmal fremdsprachlicher Akzente und sind daher als gesonderte Klasse zu berücksichtigen.

Es erscheint grundsätzlich sinnvoll, zwei Formen von Interferenz durch phonologische Regeln zu unterscheiden: (a) Einen *Transfer* phonologischer Regeln

aus L1 auf L2 und (b) die *Nicht-Anwendung* phonologischer Regeln von L2 durch L1-Sprecher.

6.3.2 Transfer phonologischer Regeln aus L1 auf L2

Hier wird eine phonologische Regel der Muttersprache des Sprechers auf die Zielsprache übertragen und führt dort zu spezifischen Abweichungen von der kanonischen Aussprache, weil die betreffende Regel in L2 keine Gültigkeit hat. Voraussetzung für einen solchen Transfer ist jedoch stets, daß das zu artikulierende zielsprachliche Material den phonologischen oder orthographischen Anwendungskontext für die betreffende Regel enthält. So ist z.B. die deutsche „Auslautverhärtung“ in einer Reihe von Kontexten auf englischen Wortschatz anwendbar, etwa in Englisch *dog* [dɒg], nach deutschen Regeln als [dɔk] realisiert. Läßt die Zielsprache dagegen auslautende stimmhafte Obstruenten ohnehin nicht zu, so ist ein potentieller Anwendungskontext für die Regel der Auslautverhärtung ist nicht gegeben.

Als Beispiel für einen Transfer phonologischer Regeln aus L1 auf L2 soll hier der *Transfer der Vokalnasalisierung* bei Sprechern des Französischen eingehender dargestellt werden: Das Französische weist eine phonologische Regel auf, der zufolge Vokale in bestimmten Kontexten als nasalierte Vokale V[+nasal] realisiert werden. Diese tritt z.B. in Kraft in *mont* [mɔ̃] oder *prendre* [pʁɑ̃dʁ], aber auch in Eigennamen wie *Cavaillon* [kavajɔ̃] oder *Dunkerque* [dœ̃kɛʁk].

In verschiedenen phonologischen Arbeiten wurde diskutiert, ob es sich bei diesen nasalisierten Vokalen um eigenständige Phoneme des Französischen oder um Allophone der entsprechenden Oralvokale handelt, welche aus einer zugrundeliegenden Basisform V[+oral] systematisch abgeleitet werden können (cf. hierzu zusammenfassend Love 1981: 99ff.). Für eine solche Allophonie spricht die Beobachtung, daß die Alternation zwischen Oral- und Nasalvokalen etwa in Paaren wie *prendre* ~ *prenez* durch spezifische phonetische Kontexte motiviert ist, die eine Ableitung aus einer gemeinsamen Basisform /pʁɛn/ mittels einer phonologischen Regel bedingen. Eine solche Repräsentation ermöglicht eine Vereinfachung bzw. ökonomischere Formulierung des Lexikons, denn in diesem Fall muß nur eine einzige phonologische Basisform in das Lexikon aufgenommen werden, aus denen allophonische Varianten mit einem Nasalvokal V[+nasal] mittels der allgemeinen Regel ableitbar sind.

Schane (1968: 48) formuliert diese Ableitung von nasalisierten Vokalen als eine zweistufige phonologische Regel, in der zunächst in Regel (R1) ein Vokal nasalisiert wird, wenn ihm ein Nasalkonsonant C[+nasal] plus ein weiterer Konsonant oder eine Wortgrenze folgt, und anschließend in (R2) der Nasalkonsonant

C[+nasal], der als rechter Kontext für Regel R1 wirksam wurde, getilgt wird. Formal läßt sich die Regelsequenz wie folgt schreiben:

(R1) Nasalisierung von Vokalen:

$$V [-nasal] \rightarrow V [+nasal] \quad / \quad _ \quad C [+nasal] \alpha \quad \left\{ \begin{array}{c} C \\ \#\# \end{array} \right\}$$

(R2) Tilgung des nasalen Konsonanten:

$$C [+nasal] \alpha \rightarrow \emptyset \quad / \quad V [+nasal] _$$

Im vorliegenden Zusammenhang ist nun von Interesse, ob bei Sprechern der L1 Französisch ein Transfer dieser Regel auf eine Zielsprache L2 zu beobachten ist. Betrachtet man beispielsweise die Zielsprache Deutsch, so zeigt sich, daß eine Reihe französischer Sprecher die oben formulierte Regel tatsächlich auch auf deutsche Wörtern anwenden, sofern hier der Kontext für die Regelanwendung erfüllt ist. In der Liste der verwendeten Stimuli ist dies etwa in *Blankenstein* oder *Frankfurt* gegeben, welche beide den geforderten Kontext VOKAL + NASAL-KONSONANT enthalten. Französische Sprecher produzieren hier oftmals Formen wie [bläkŋstam] und [fɛäkfvɛt], die entsprechend der o.g. Regel einen nasalisierten Vokal aufweisen. Um einen solchen Aussprachefehler zu modellieren, ist es möglich, die oben genannte Regel des Französischen postlexikalisch auf die kanonische deutsche Aussprache anzuwenden. Hieraus ergibt sich folgende Ableitungssequenz:

Kanonisch		Regel (R1)		Regel (R2)
[bläkŋstam]	→	[bläkŋstam]	→	[bläkŋstam]
[fɛäkfvɛt]	→	[fɛäkfvɛt]	→	[fɛäkfvɛt]

Auf den ersten Blick scheint die Datenlage hier mit der phonologischen Theorie im Einklang zu stehen. Betrachtet man nun jedoch ergänzend die quantitative Auswertung dieser phonologischen Variablen in den vorliegenden Sprachdaten, so ergibt sich ein geringfügig anderes Bild bezüglich einer rein phonetisch-phonologischen Natur dieses Aussprachefehlers. Tabelle 6.1 zeigt die Distribution von oralen vs. nasalen Vokalen in den entsprechenden Kontexten:

Variable: V [-nasal] (C [+nasal]) Kontext: __ # , __ C Beispiel: <i>Fr[an]kfurt, Bl[an]kenstein</i>				
Variante	<i>gelesen</i>		<i>wiederholt</i>	
	abs	rel %	abs	rel %
V [-nasal]	56	77,8	36	90
V [+nasal]	16	22,2	4	10
GESAMT	72	100	40	100

**Tabelle 6.1: Distribution von nasalierten Vokalen
L1 Französisch → L2 Deutsch**

Erläuterung der Darstellung in der Tabellen 6.1 und 6.2

Wie in Kapitel 5.4 dargestellt, wurde für eine Reihe phonetischer Variablen eine quantitative Analyse durchgeführt, bei der die Häufigkeit einzelner phonetischer Realisierungsvarianten der betreffenden Variablen ermittelt wurde. In Tabellen 6.1 und 6.2 ist diese Auftretenshäufigkeit in absoluten (*abs*) und relativen (*rel%*) Werten angegeben. Die Werte sind jeweils auf eine Gruppe von 20 Sprechern bezogen. Unterschieden wird nach den Häufigkeiten bei (a) gelesenen Material (Spalte *gelesen*) und (b) der Wiederholung der akustisch dargebotenen korrekten zielsprachlichen Aussprache durch einen Muttersprachler von L2 (Spalte *wiederholt*). Signifikante Differenzen zwischen den beiden Darbietungsmodi *gelesen* und *wiederholt* deuten dabei meist auf besondere Einflüsse der Schriftaussprache als Fehlerursache hin.

Vergleicht man hier die in der Leseaufgabe und der Wiederholungsaufgabe realisierten Varianten, so fällt auf, daß die Nasalierungsregel weit seltener angewandt wird, wenn *kein* orthographischer Stimulus vorliegt. Während in der Leseaufgabe der Anteil der Nasalvokale an allen Realisierungen bei relativ hohen 22,2% liegt¹⁵, sinkt er in der Wiederholungsaufgabe auf nur 10% ab.

Dies kann als Indiz für einen Einfluß orthographischer Merkmale auf die phonetischen Realisierungen der Sprecher gewertet werden. Die Annahme einer allein auf der Lautebene wirksamen *autonomen* phonologischen Regel ist demnach in Frage zu stellen. Die Orthographie wird hier zumindest als verstärkender

¹⁵ Die Einschätzung, daß eine Fehlerhäufigkeit von 22,2% als „relativ hoch“ zu bewerten ist, erfolgt hier vor allem vor dem Hintergrund der hohen L2-Kompetenz, d.h. der guten bis sehr guten Deutschkenntnisse innerhalb der Sprechergruppe.

Einfluß bei einem zunächst nur potentiellen Transfer der Regel auf L2 wirksam. Obwohl also mit der oben beschriebenen zweistufigen Regel der fragliche Aussprachefehler prinzipiell adäquat modelliert werden kann, sollte vor der tatsächlichen Regelanwendung der orthographische Kontext geprüft werden. Derartige orthographische Einflüsse werden weiter unten (Kap. 6.4) im Detail diskutiert.

6.3.3 Nichtanwendung phonologischer Regeln aus L2

Neben einem Transfer phonologischer Regeln aus L1 auf L2 ist häufig zu beobachten, daß Sprecher von L1 systematisch auftretende phonologische Regeln der Zielsprache nicht anwenden und dadurch spezifische Aussprachefehler produzieren. Eine Nichtanwendung zielsprachlicher phonologischer Regeln ist u.a. auch mit Blick auf eine regelbasierte Modellierung als gesonderter Fehlertyp zu berücksichtigen. Denn während der oben genannte Fehlertyp eines Transfers von L1-Regeln auf die Zielsprache prinzipiell dadurch zu modellieren ist, daß eine Regel aus L1 auf eine Phonemsequenz von L2 projiziert wird, ist es bei Typ (2) unter Umständen erforderlich, eine in der kanonischen L2-Form bereits zur Anwendung gekommene Regel nachträglich wieder zu *eliminieren* (das zu diesem Zweck verwendete Verfahren ist in Kap. 7.4 skizziert).

Wiederum soll ein Beispiel der L1 Französisch mit der Zielsprache Deutsch zur Illustration dieses Fehlertyps dienen: In der Standardaussprache des Deutschen erhält die orthographische Sequenz <ng> die Lautung [ŋ], beispielsweise in *schwingen*, *Rang* oder *Angel*. In der phonetischen Oberflächenform erscheint also das orthographische <g> nicht als gesonderter Laut, sondern verschmilzt mit dem vorausgehenden <n> zu einem einzelnen velaren Nasallaut [ŋ]. Das <n> wiederum wird in seiner Artikulationsstelle mit dem nachfolgenden <g> assimiliert; es wird also velar artikuliert. Diese Regel gilt im Deutschen mit nur wenigen Ausnahmen, die in einer Reihe fremdstämmiger Wörter (z.B. *Hangar*, *Ungar*, *Tango*) auftreten¹⁶.

Es wurde nun verschiedentlich angenommen, daß die zugrundeliegende phonologische Repräsentation des Lautes [ŋ] ein bisegmentales /ng/ sei, welches mittels einer phonologischen Regel in die Oberflächenform [ŋ] überführt wird (cf. beispielsweise Bierwisch 1972). Es handle sich hierbei um eine zweistufige Regel, bei der zunächst die Assimilation /n/ → [ŋ] erfolge und anschließend in der zweiten Stufe das /g/ getilgt werde. Sie läßt sich wie folgt darstellen:

¹⁶ In einigen Varietäten des Deutschen (z.B. Österreichisch) gilt die Regel jedoch nicht in vollem Umfang; hier erscheint /ng/ als [ŋg] (z.B. in *Angel* [aŋgəl]).

(R1) Assimilation von /n/

$$\begin{bmatrix} + \text{alveolar} \\ + \text{nasal} \end{bmatrix} \rightarrow \begin{bmatrix} + \text{velar} \\ + \text{nasal} \end{bmatrix} \quad / \quad \text{---} \begin{bmatrix} + \text{obstruent} \\ + \text{velar} \end{bmatrix}$$

(R2) Tilgung von /g/

$$\begin{bmatrix} + \text{obstruent} \\ + \text{velar} \\ + \text{stimmhaft} \end{bmatrix} \rightarrow \emptyset \quad / \quad \begin{bmatrix} + \text{nasal} \end{bmatrix} \text{---}$$

Zu beachten ist hierbei, daß Regel (R1) sowohl für die Verbindung /ng/ als auch für /nk/ gilt, während (R2) nur auf *stimmhafte* Obstruenten zugreift und somit für /nk/ nicht gilt. Andernfalls würde die Regel Formen wie *Anker* *[aŋkɛ] generieren. R2 muß daher für /nk/ blockiert werden:

	UR		(R1)		(R2)	Ergebnis
(1)	/angəl/	→	JA	→	JA	[aŋəl]
(2)	/ankɛɾ/	→	JA	→	JA	* [aŋɛɾ]
(3)	/ankɛɾ/	→	JA	→	NEIN	[aŋkɛɾ]

Bei französischen Sprechern mit der Zielsprache Deutsch zeigt sich eine deutliche Tendenz zu einer Realisierung der Sequenz <ng> als bisegmentaler Laut [ŋg] oder [ŋk]. Dies gilt sowohl in wortfinaler Position (z.B. bei *Otzing* → [ɔtsɪŋk]) wie auch in prävokalischer Stellung (*Vöhringen* → [fø:ɾŋgən]). Die folgende Tabelle zeigt die Auftretenshäufigkeiten der Varianten für diese Variable (Datengrundlage: 16 Sprecher L1 Französisch):

Variable: [ŋ]				
Kontext: __# , __ ##				
Beispiele: Otzi[ng], Vöhri[ng]en				
Variante	gelesen		wiederholt	
	abs	rel %	abs	rel %
[ŋ]	28	38,9	26	65,0
[ŋk]	44	61,1	14	35,0
[ŋg]				
GESAMT	72	100	40	100

Tabelle 6.2: Realisierungen von [ŋ] bei Sprechern L1 Französisch, L2 Deutsch

Es zeigt sich, daß die Varianten [ŋg] und [ŋk] (in der Tabelle zu einer Variante zusammengefaßt) bei gelesenen Material mit einer hohen Frequenz von 61% auftreten. Bei einer Wiederholung akustischer Stimuli liegt der Anteil von [ŋg]/[ŋk] noch immer bei relativ hohen 35%. Im Sinne der oben dargelegten zweistufigen phonologischen Regel des Deutschen läßt sich diese Beobachtung dahingehend interpretieren, daß Sprecher des Französischen die Regel nicht bzw. nicht in vollem Umfang anwenden:

	UR		(R1)		(R2)	Ergebnis
Deutsch	/angəl/	→	JA	→	JA	[aŋəl]
L1 Französ.	/angəl/	→	JA	→	NEIN	[aŋgəl]

Zwar wird Regel R1 angewendet, so daß /n/ als velares [ŋ] realisiert wird, die zweite Teilregel R2 wird dagegen übersprungen, so daß das Segment /g/ erhalten bleibt. Diese Interpretation als zweistufige Regel generiert genau die phonetischen Oberflächenformen, die in den Sprachdaten beobachtet werden konnten. Die höhere Auftretenshäufigkeit der Variante [ŋg] bei gelesenen Material kann dabei im Sinne von Bierwisch (1972) dahingehend interpretiert werden, daß die phonologische Basisrepräsentation /ng/ in der Orthographie erhalten geblieben ist und daher von den Sprechern unmittelbar in eine phonetische Oberflächenform überführt wird.

6.4 Einflüsse der Schriftaussprache

6.4.1 Transfer der Graphem-Phonem-Beziehungen

Die bisher skizzierten Fehlertypen können prinzipiell mittels Regeln formalisiert werden, deren Anwendungsbedingungen bzw. -restriktionen ausschließlich auf der phonetisch-phonologischen Ebene lokalisiert sind. Zur Modellierung solcher Fehler sind demzufolge keinerlei Informationen anderer linguistischer Ebenen als der Lautebene erforderlich. Lediglich in einigen Fällen wurden Morphem- oder Wortgrenzen als Anwendungskontexte von Regeln wirksam. Erste Einschränkungen hinsichtlich einer Autonomie lautlicher Interferenzen mußten jedoch bereits bei denjenigen Aussprachefehlern geltend gemacht werden, deren Auftretenshäufigkeit in Abhängigkeit vom Medium der Darbietung (orthographisch vs. akustisch) variiert: Tritt ein Aussprachefehler vorwiegend bei orthographischer Darbietung des Stimulus – d.h. beim Lesen des L2-Materials – auf, so kann hierin ein Hinweis darauf gesehen werden, daß dieser Fehler zumindest teilweise aus der orthographischen Repräsentation und ihrer spezifischen Beziehung zur Lautebene resultiert.

Wird das fremdsprachliche Material gelesen, so treten Aussprachefehler – insbesondere bei Sprechern mit geringen L2-Kenntnissen – vor allem durch die Übertragung der muttersprachlichen Schriftaussprache auf die Zielsprache auf. Hirschfeld (1983: 172) verwendet für derartige Einflüsse der graphemischen Repräsentation auf die Aussprache einer L2 den Terminus „Schriftbildinterferenz“. Gemeint ist damit jedoch nicht, wie man auf dieser Begrifflichkeit leicht schließen könnte, eine primär visuelle Interferenz zwischen den beteiligten Graphem-Inventaren, sondern vielmehr eine in ihrem Resultat *phonetische* Interferenz, deren Ursache jedoch in unterschiedlichen Graphem-Phonem-Relationen liegt, d.h. in Unterschieden der Schriftaussprache-Regeln von L1 und L2. Daher soll hier zur Charakterisierung dieses Fehlertypus der Terminus *Transfer der Graphem-Phonem-Beziehungen* (Abk: *Transfer der GTP-Beziehungen*) verwendet werden. Gegenüber den bisher diskutierten Lautsubstitutionen, welche häufig nur relativ moderate Abweichungen von der kanonischen Aussprache generieren, resultiert ein Transfer der GTP-Beziehungen häufig in stark akzentgefärbten Varianten, die auffällige Phonemsprünge gegenüber der kanonischen Aussprache aufweisen.

Zur Etablierung einer grundlegenden Fehlertypologie erscheint es sinnvoll, zunächst zwei elementare Situationen zu unterscheiden, durch die das Verhältnis der geschriebenen Formen von L1 und L2 bestimmt sein kann. Hieraus kann eine erste Eingrenzung möglicher Fehler abgeleitet werden:

(1) **Graphemotaktische Analogie von L1 und L2.** Das zielsprachliche Wort weist orthographische Sequenzen auf, die auch *in der L1 des Sprechers gültige Sequenzen sind*, d.h. mit der Graphemotaktik von L1 im Einklang stehen. Dies eröffnet dem Sprecher die Möglichkeit, entsprechende Schriftaussprache-Regeln seiner Muttersprache in unmodifizierter Weise auf L2-Material anzuwenden. Dabei ergeben sich z.B. Varianten wie die folgenden:

L2	Ortsname	Kanonisch	L1	Variante
Französisch	<i>Mulhouse</i>	[myluz]	Englisch	[muθhaʊz]
Französisch	<i>Tarbes</i>	[taʁb]	Englisch	[t ^h ɑ:bz]
Englisch	<i>Stafford</i>	[stæfəd]	Französisch	[stafɔʁ]
Englisch	<i>Plymouth</i>	[pliməθ]	Französisch	[plimut]
Niederländ.	<i>Schiphol</i>	[sxɪphɔl]	Deutsch	[ʃifo:l]
Niederländ.	<i>Straelen</i>	[stra:lən]	Deutsch	[ʃtʁɛ:lən]
Portugies.	<i>Estoril</i>	[ɪʃturil]	Deutsch	[estɔʁil]
Schwedisch	<i>Linköping</i>	[linçø:piŋ]	Deutsch	[lɪŋkø:piŋ]

(2) **Graphemotaktischer Kontrast zwischen L1 und L2.** Weist das zielsprachliche Wort dagegen orthographische Sequenzen auf, welche in der L1 des Sprechers *keine gültigen Sequenzen* sind, so ist ein Transfer, wie er oben skizziert wurde, nicht ohne weiteres möglich, da die L1 des Sprechers keine adäquaten Regeln der Phonemisierung dieser Sequenzen anbietet. So wären z.B. bei der L1 Deutsch die folgenden fremdsprachlichen Graphemsequenzen keinem vollständigen Transfer der deutschen Schriftaussprache-Regeln unterworfen, da es sich nicht um mögliche orthographische Sequenzen des Deutschen handelt:

engl.	<i>brought</i>	<ought>
ungar.	<i>Zsusa</i>	<zs>
franz.	<i>Bayeux</i>	<ayeu>
tschech.	<i>Hrdlička</i>	<hrdl>

Die deutschen Schriftaussprache-Regeln könnten in diesen Fällen lediglich in Form einer sequentiellen Aneinanderfügung von Phonemen angewendet werden, die jedoch in höchst artifiziell wirkenden Lautfolgen resultieren würde (beispielsweise engl. *brought* als [b ʁ o ũ k h t]). Wie weiter unten noch ausgeführt wird, tritt eine solche Strategie der Phonemisierung von L2-Material nur selten auf.

Festzuhalten ist zunächst der Umstand, daß die oben skizzierten Konstellationen (a) *Analogie* und (b) *Kontrast* zumeist nicht in reiner Form auftreten. Weitaus häufiger ist die Situation vorzufinden, daß zielsprachliche Wörter, Phrasen oder Sätze sowohl Analogien als auch Kontraste aufweisen. Einige Teilsequenzen des zielsprachlichen Materials enthalten dabei orthographische Sequenzen, die auch in der L1 des Sprechers gültig sind, andere hingegen sind L2-spezifisch und können nicht mittels L1-Regeln phonemisiert werden.

Aus diesem Sachverhalt folgt ein wichtiges Charakteristikum akzentbehafteter Aussprachevarianten, die durch einen Transfer von Graphem-Phonem-Beziehungen gekennzeichnet sind: Findet ein solcher Transfer statt, so bedeutet dies nicht notwendigerweise, daß die *gesamte* L2-Sequenz davon betroffen ist. Vielmehr können nur Teilsequenzen betroffen sein, während andere Teilsequenzen annähernd oder vollständig nach den Regeln der Zielsprache realisiert werden. Hieraus resultieren vielfältige Mischformen, die nicht allein auf der Basis der Schriftaussprache-Regeln von L1 oder L2 erklärbar sind. Es ist aus diesem Grunde sinnvoll, im Zusammenhang mit einem Transfer von Graphem-Phonem-Beziehungen auf die Zielsprache eine grundlegende Unterscheidung zwischen *lokalem* und *globalem* Transfer zu treffen:

(a) Globaler Transfer muttersprachlicher GTP-Beziehungen

Ein *globaler* Transfer tritt auf, wenn fremdsprachliches Material vollständig nach den Schriftaussprache-Regeln der L1 des Sprechers realisiert wird. Eine solche Strategie führt mitunter zu recht artifiziell wirkenden Aussprachevarianten, etwa deutsch [ma:ʁzai:lə] für französisch *Marseille* oder deutsch [hɑ:tləpɔ:l] für englisch *Hartlepool*.

(b) Lokaler Transfer muttersprachlicher GTP-Beziehungen

Weitaus häufiger tritt der Fall auf, daß zielsprachliche Graphemsequenzen nur teilweise nach den muttersprachlichen GTP-Regeln realisiert werden, während die verbleibenden Graphemfolgen annähernd nach den zielsprachlichen GTP-Regeln ausgesprochen werden. Bei einem solchen *lokalen* Transfer kommt es zu Mischformen, in denen sowohl L1- als auch L2-Ausspracheregeln wirksam werden. So enthält z.B. dt. *Ravensberg* in der Realisierung [ʁa:vənsbɜ:g] durch englische Muttersprachler Elemente der englischen als auch der deutschen Graphem-Phonem-Beziehungen.

Die vorliegenden Sprachdaten liefern starke Evidenz dafür, daß ein globaler Transfer der muttersprachlichen Graphem-Phonem-Beziehungen auf fremdsprachlichen Wortschatz verhältnismäßig selten auftritt. Hierfür sind vor allem zwei Gründe anzuführen:

Erstens verfügen auch Sprecher mit sehr geringen oder keinen Kenntnissen der Zielsprache häufig über elementares Wissen über die Aussprache spezifischer zielsprachlicher Graphemsequenzen – oder zumindest über das Wissen, daß diese Sequenzen *nicht* den muttersprachlichen Schriftaussprache-Regeln folgen. So ist beispielsweise die Aussprache des deutschen <sch> als [ʃ] oftmals auch Sprechern ohne formale Kenntnisse des Deutschen geläufig; nur wenige Sprecher würden hier den Versuch unternehmen, die Aussprache von <sch> aus einer Phonemisierung der Einzelsegmente <s> <c> <h> herzuleiten und damit zu Varianten wie [skh], [stsh] oder [ssh] gelangen. In diese Richtung weisen auch Befunde von Fitt (1998) und Trancoso et al. (1999). Zweitens generiert ein uneingeschränkter Transfer muttersprachlicher Graphem-Phonem-Beziehungen in vielen Fällen unnatürliche Phonemreihungen, da bestimmte L2-Graphemfolgen nicht der Graphemotaktik von L1 entsprechen – z.B. engl. *Shrewsbury*, das nach unmodifizierten deutschen Regeln etwa als [shɛʁfsbu:ri] zu realisieren wäre. Tatsächlich aber kommt es in diesen Fällen weitaus häufiger zur Anwendung von Teilkenntnissen, Drittsprachenkenntnissen oder zu spontanen Regelbildungen durch die Sprecher.

Diese Beobachtungen sowie die Differenzierung zwischen *globalem* und *lokalem* Transfer der Graphem-Phonem-Beziehungen hat unmittelbare Konsequenzen für die im Rahmen dieser Studie gewählte Methode der regelbasierten Modellierung. Wäre ein globaler Transfer der Regelfall, so ließe sich ein großer Teil der Aussprachevarianten dadurch modellieren, daß die Schriftaussprache-Regeln der L1 des Sprechers auf orthographischen Eingabetext der Zielsprache L2 angewendet werden. Gerade die Beobachtung aber, daß ein globaler Transfer nur höchst selten auftritt, war im Rahmen dieser Arbeit ausschlaggebend, von dieser Methode Abstand zu nehmen (siehe hierzu auch Kapitel 7.3.4).

6.4.2 Interferenz der Graphem-Inventare und Diakritika

Zwar verwenden alle in dieser Studie untersuchten Sprachen ein lateinisches Schriftsystem und damit ein gemeinsames alphabetisches Basisinventar, jedoch verfügen sie darüber hinaus über eine Reihe von sprachspezifischen Sonderzeichen und Diakritika¹⁷. Diese erfüllen phonetische Funktionen, d.h. sie repräsentieren spezielle Laute und Lautunterscheidungen auf der orthographischen Ebene. Sprachspezifische Sonderzeichen dieses Typs werden mitunter als eine eigenständige Quelle von Aussprache Fehlern von Nicht-Muttersprachlern wirksam. Dies ist beispielsweise der Fall, wenn einzelne Grapheme

¹⁷ Es wurden hier nur einige wenige Beispiele zur Illustration dieses Fehlertyps herausgegriffen. Eine detaillierte Übersicht der gebräuchlichen Diakritika in Sprachen mit lateinischem Alphabet gibt Wells (2001b).

- (a) in der L1 des Sprechers keine genaue Entsprechung haben, d.h. nicht zum L1-Inventar gehören, oder
- (b) im L1-Inventar zwar auftreten, jedoch dort einen anderen Lautwert repräsentieren als in der Zielsprache.

Fall (b) läßt sich prinzipiell als eine Interferenz der Graphem-Phonem-Beziehungen klassifizieren, wie sie weiter oben bereits skizziert wurde. Auch seine Behandlung im Rahmen einer regelbasierten Modellierung unterscheidet sich nicht grundsätzlich davon (siehe Kap. 7). Fall (a) ist dagegen als Ursache eines eigenständigen Fehlertyps zu betrachten, der daraus resultiert, daß ein L2-spezifisches Graphemzeichen im L1-Inventar des Sprechers nicht existiert. Beispiele hierfür sind etwa sog. „Umlaute“ <ä, ö, ü> im Deutschen (vordere Vokale, die orthographisch durch ein Trema gekennzeichnet sind) oder diakritische Markierungen durch *accent* <é, è, ê> oder *cedille* <ç> im Französischen. Liegen derartige sprachspezifische Sonderzeichen vor, so kann eine *Interferenz der Grapheminventare* als Quelle für Aussprachefehler wirksam werden.

In diesen Fällen ist häufig zu beobachten, daß Sprecher bei Lesen von L2-Material das *jeweils ähnlichste Graphem* ihrer L1 heranziehen und die Aussprache des L2-Graphems hieraus herleiten. So tritt beispielsweise bei englischen Sprechern mit der Zielsprache Deutsch häufig der Fall auf, daß die deutschen Grapheme <ä, ö, ü> beim Lesen deutschen Materials durch ihre unmarkierten Basisgrapheme <a, o, u> ersetzt und entsprechend phonemisiert werden. Hieraus resultieren Aussprachevarianten wie [dʊsəldɔ:f] für *Düsseldorf* oder [krɔ:nɪŋ] für *Kröning*. Nach dem gleichen Prinzip einer graphemischen Reinterpretation entstehen Fehler wie die Realisierung des französischen <ç> als Plosiv [k] oder des spanischen <ñ> als [n] (als Ersatzlaut für [ɲ]). Derartige Fehler können grundsätzlich immer dann beobachtet werden, wenn die Zielsprache über sprachspezifische Grapheme verfügt, die in der L1 des Sprechers nicht auftreten.

Um diese Annahme zu prüfen, wurden im Rahmen der in dieser Studie erstellten Sprachaufzeichnungen informelle Untersuchungen mit der Zielsprache Tschechisch durchgeführt. Tschechisch erschien zu diesem Zweck als besonders geeignet, da es über eine Reihe von Graphemen verfügt, bei denen die Basisgrapheme <c, s, z, r, n, e> durch die diakritische Markierung mit einem *Háček* als <č, š, ž, ř, ň, ě> einen deutlich veränderten Lautwert repräsentieren. Da die phonetische Funktion dieses Diakritikons den meisten Sprechern der vorliegenden Sprachdatensammlung unbekannt ist, ist hier eine starke Tendenz zur graphemischen Reinterpretation zu erwarten. Tatsächlich zeigten die Versuche deutlich, daß sich die Sprecher bei der lautlichen Realisierung von <č, š, ž, ř, ň, ě> fast ausnahmslos von den orthographischen Basisgraphemen und ihren phonemischen Entsprechungen ihrer L1 leiten lassen. Die folgende Tabelle zeigt typische Aussprachevarianten tschechischer Ortsnamen, wie sie bei

Aussprachevarianten tschechischer Ortsnamen, wie sie bei Sprechern des Deutschen auftreten:

Name	kanonische Aussprache	Graphemische Reinterpretation	Resultierende Variante
Příbram	[pr̩ɪbram]	Pribram	[pɾi:br̩am]
Kroměříž	[kr̩ɔm̩ʲɛr̩ʒɪʃ]	Kromeriz	[kr̩ɔ:m̩ɛr̩ɪts]
Strážný	[stra:ʒni:]	Strazny	[stratsni]
Ševětín	[ʃɛv̩ʲɛt̩ɪn]	Sevetin	[ze:v̩ɛt̩ɪn]

Tabelle 6.3: Graphemische Reinterpretation diakritisch markierter Grapheme am Beispiel der Zielsprache Tschechisch¹⁸

Ähnliche Ergebnisse sind z.B. bei Zielsprachen wie Dänisch, Schwedisch oder Norwegisch zu erwarten, bei denen die diakritische Markierung von <a> und <o> mit superkribiertem Ring <å> (Norwegisch, Schwedisch) bzw. einer Durchstreichung <ø> (Dänisch, Norwegisch) einen deutlichen Phonemsprung markiert. Einige informelle Versuche haben auch hier gezeigt, daß z.B. deutsche Sprecher ohne Kenntnisse des Schwedischen das Graphem <å>, welches im Schwedischen das Phonem /o/ repräsentiert, fast ausnahmslos als [a:], [a] oder [ɑ] realisieren.

Die aus dem zuletzt skizzierten Fehlertyp resultierenden potentiellen Lautsubstitutionen lassen sich für jede Sprachrichtung systematisch in Form von *graphemischen Interferenztabellen* darstellen. In einer solchen Tabelle sind zunächst für jede Zielsprache alle sprachspezifischen Sonderzeichen aufgeführt. Dabei ergeben sich für jede einzelne L1 potentielle *Ersatzgrapheme*. Zumeist handelt es sich dabei um die entsprechenden Basisgrapheme ohne die dazugehörigen Diakritika. Der vom Sprecher eingesetzte Ersatzlaut ist seinerseits bestimmt von diesem Ersatzgraphem: In der Regel tritt derjenige Laut als Ersatz ein, welcher in der Muttersprache des Sprechers durch das jeweilige Ersatzgraphem repräsentiert wird. Das folgende Beispiel zeigt eine solche Tabelle für die Zielsprache Französisch in Kombination mit den L1 Deutsch, Englisch und Italienisch:

¹⁸ Ich danke Alena Böhmová und Nino Peterek (Prag) für ihre freundliche Unterstützung bei der Transkription tschechischer Namen.

Graphem	Lautung	Ersatz- graphem	Mögliche Ersatzlaute (abhängig von L1 des Sprechers)		
			L1 Deutsch	L1 Englisch	L1 Italienisch
<é>	[e]	<e>	[e: , ε , ε: , ə]	[e , eɪ , ɪ , ə]	[e , ε]
<è>	[ɛ]	<e>	[e: , ε , ε: , ə]	[e , eɪ , ɪ , ə]	[e , ε]
<ê>	[ɛ]	<e>	[e: , ε , ε: , ə]	[e , eɪ , ɪ , ə]	[e , ε]
<ë>	[ɛ]	<e>	[e: , ε , ε: , ə]	[e , eɪ , ɪ , ə]	[e , ε]
<â>	[a]	<a>	[a: , a]	[ɑ , æ, eɪ, ə]	[a: , a]
<ô>	[o]	<o>	[o: , ɔ]	[əʊ, ɔ:, ɒ, ə]	[o:, ɔ]
<œ>	[œ]	<oe> (ö)	[o:e , o:ə , ø:]	[əʊ, əʊɪ]	[oε, oε]
<ç>	[s]	<c>	[k]	[k]	[tʃ, k]

Tabelle 6.4: Orthographische Sonderzeichen und Diakritika (am Beispiel der Zielsprache Französisch) und ihre Ersetzung durch L1-Grapheme sowie die daraus resultierenden Ersatzlaute

Wie aus der Tabelle abzulesen ist, kann der Ersatzlaut bzw. können die Ersatzlaute in Abhängigkeit von der L1 des Sprechers stark variieren. Der Grund hierfür liegt in den unterschiedlichen Schriftaussprache-Regeln, die zur Anwendung kommen, um dem Ersatzgraphem eine Lautung zuzuweisen.

Derartige Tabellen erlauben zwar eindeutige Prognose des tatsächlich auftretenden Ersatzlautes. So ist beispielsweise zu berücksichtigen, daß das jeweilige Ersatzgraphem meist mehrere mögliche lautliche Entsprechungen aufweist, deren Selektion u.a. vom orthographischen Kontext abhängt. Dennoch ermöglichen Tabellen graphemischer Interferenzen nicht nur eine systematische Darstellung potentieller Fehler, sondern können zugleich als Basis der Erstellung von Regeln zur Modellierung solcher Fehler dienen. In Kap. 7.4 wird beispielhaft gezeigt, wie sich solche Zuordnungen in phonologische Regeln überführen lassen.

Zwei weitere Sonderfälle dieses Fehlertyps seien der Vollständigkeit halber erwähnt: In den oben genannten Beispielen ist die Interferenz dadurch charakterisiert, daß sprachspezifische, diakritisch markierte Grapheme durch ähnliche, unmarkierte Basisgrapheme der L1 des Sprechers ersetzt werden. Grundlage der Ersetzung ist dabei die visuelle Ähnlichkeit der Basisgrapheme. Darüber hinaus sind auch folgende Fälle denkbar:

(a) **Abweichende phonetische Funktion der Diakritika.** In einigen Sprachenpaaren finden graphisch identische Diakritika Verwendung, repräsentieren je-

doch dabei unterschiedliche lautliche Eigenschaften. So treten beispielsweise der Gravis <`> und der Akut <'> sowohl im Französischen als auch im Italienischen auf, haben aber in beiden Sprachen abweichende phonetische Funktionen. Während im Französischen ausschließlich eine Änderung der Vokalqualität (Grad der Öffnung) angezeigt wird, verwendet das Italienische dasselbe Diakritikon zur Markierung der Wortbetonung *und* der Vokalqualität (*caffè, perché, città*). Im Falle dieses Sprachenpaares liegt hier also eine weitere Interferenzquelle vor. In anderen Sprachen wiederum markieren die Diakritika nicht die *Vokalqualität*, sondern deren *Quantität* (Länge), etwa im Ungarischen oder im Tschechischen, wodurch eine weitere Fehlerquelle eröffnet wird: So könnte ein Sprecher des Französischen dazu neigen, tschechisch <é> (phonetisch [ɛ:]) auf der Basis der Funktion des Akut im Französischen mit der Lautung [e] zu realisieren, während umgekehrt ein Sprecher des Tschechischen französisches <é> möglicherweise mit einem Langvokal realisieren wird.

(b) **Graphische Ähnlichkeit von Diakritika.** Als eine weitere potentielle Interferenzquelle ist der Fall zu nennen, daß Diakritika zweier Sprachen eine hohe graphische Ähnlichkeit aufweisen und daher beim Lesen fehlerhaft interpretiert werden. Dies ist beispielsweise denkbar beim Sprachenpaar Deutsch/Ungarisch, bei dem eine Verwechslung von Trema <¨> und Doppelakut <˝> mit entsprechenden Auswirkungen auf die Lautung möglich ist (deutsch <ö> gegenüber ungarisch <ő>). Während der Doppelakut im Ungarischen eine Vokallängung markiert, kennzeichnet das Trema im Deutschen einen Wechsel von hinteren zu vorderen Vokalen („Umlaute“). Im vorliegenden Fall <ö> vs. <ő> ist jedoch keinerlei ausgeprägte Interferenz zu erwarten, da ungarisch <ő> ohnehin einen vorderen Vokal [ø:] repräsentiert, so daß die potentielle Interferenz sich allein auf die Vokallänge beschränken wird. Für einen solchen Interferenztypus würde der von Hirschfeld vorgeschlagene Terminus der „Schriftbildinterferenz“ tatsächlich angemessen erscheinen.

6.4.3 Einflüsse der Schriftaussprache: Vorläufiges Fazit

Die bis hierher vorgeschlagene Typologie ist weitgehend aus den in dieser Studie behandelten Sprachen abgeleitet und daher notwendigerweise ergänzungsbedürftig. Zweifellos kann eine Vielzahl weiterer Beispiele für eine Interferenz der Grapheminventare angeführt werden. Die praktische Schwierigkeit hierbei besteht jedoch darin, daß derartige Interferenzen für jedes Sprachenpaar gesondert erfaßt werden müssen, da die potentiellen Fehlerquellen aus der je spezifischen Kombination der Grapheminventare L1 und L2 resultieren. Verallgemeinerungen über mehrere Sprachenpaare sind somit kaum möglich.

Wichtiger erscheint es im Rahmen dieser Arbeit jedoch, solche Fehlerquellen zunächst zu identifizieren und auf dieser Grundlage Möglichkeiten ihrer (regelbasierten) Modellierung zu erarbeiten. Entscheidend ist dabei die Einsicht, daß dieser Fehlertypus mittels Regeln, die ausschließlich auf der Lautebene operieren und die orthographische Repräsentation unberücksichtigt lassen, nicht erklärt oder modelliert werden kann. Es ist daher ein eigener Regeltyp erforderlich. In Kap 7.4 wird gezeigt, wie eine Regel aufgebaut sein kann, um Fehler zu modellieren, deren Ursache z.B. in einer Interferenz der Grapheminventare liegt.

Es besteht jedoch Bedarf an weiteren Studien, in denen Schrifteinflüsse auf Aussprachefehler von Nicht-Muttersprachlern eingehender untersucht werden. So bleibt in der vorliegenden Studie ein großes Untersuchungsfeld völlig unberücksichtigt: Im Falle grundlegend verschiedener Schriftsysteme (beispielsweise beim Aufeinandertreffen logographischer Schriften oder Silbenschriften mit alphabetischen Schriften) verläuft der Einfluß der Schrift auf die Aussprache einer L2 grundlegend anders, als es in den vorausgehenden Abschnitten dargestellt wurde. Eine detaillierte Untersuchung solcher Fehler und ihre Systematisierung kann und soll im Rahmen dieser Arbeit nicht geleistet werden; sie stellt jedoch zweifellos ein lohnendes und noch wenig bearbeitetes Untersuchungsgebiet dar.

6.5 Einflüsse der Morphologie

Die bisher dargestellten Fehlertypen sind allein unter Bezugnahme auf (a) die phonetische-phonologische und (b) die orthographische Ebene und ihre spezifischen Interaktionen erklärbar. Eine Reihe von Aussprachefehlern resultiert jedoch aus Fehlern, deren Ursache auf anderen linguistischen Ebenen zu suchen ist. Insbesondere der morphologische Aufbau von Eigennamen bzw. Ortsnamen, dessen Besonderheiten an anderer Stelle bereits angesprochen wurde (siehe Kap. 3.3), führt hier zu spezifischen Schwierigkeiten für nicht-muttersprachliche Sprecher. Ergänzend werden daher im folgenden einige Fehlertypen dargestellt, die auf einer Fehlinterpretation der morphologischen Struktur zielsprachlicher Wörter beruhen. Obgleich das Resultat solcher Fehlinterpretationen letztlich eine gegenüber der kanonischen L2-Form veränderte Aussprache ist, ist ihre Ursache nicht unmittelbar phonetischer Natur.

6.5.1 „Dual Route“-Strategien bei der Aussprache von Eigennamen

Wie bereits in Kapitel 3.3 ausgeführt, unterscheidet sich die morphologische Struktur von Eigennamen in verschiedener Hinsicht von der des appellativen Wortschatzes. Ein charakteristisches und in vielen Sprachen auftretendes Merk-

mal des Aufbaus von Ortsnamen besteht darin, daß eine Kombination von zwei Morphemtypen vorzufinden ist:

- (1) **Lexikalische Morpheme**, d.h. Wortbildungselemente, welche in der betreffenden Sprache noch produktiv sind, wie z.B. *-haus*, *-bach*, *Ober-*, *-tal* im Deutschen, sowie
- (2) **Onomastische Morpheme**, d.h. Konstituenten, die ausschließlich in Eigennamen auftreten, im Deutschen z.B. *-ing(en)*, *-ow*, *-roda*, *-stedt*.

Dieser morphologische Aufbau wirkt auf indirekte Weise zurück auf die von Nicht-Muttersprachlern angewendeten Strategien der Phonemisierung solcher Namen und steht dabei in enger Beziehung zu den L2-Kenntnissen der individuellen Sprecher: Sprecher mit (zumindest elementaren) Kenntnissen der Zielsprache können prinzipiell eine doppelte Strategie verfolgen, die in der sprachpsychologischen Literatur als *Dual Route-Modell* bekannt ist (cf. Coltheart 1978; Paap, Noel & Johansen 1992). Diesen Modellen des Lesens bzw. der Schriftaussprache liegt die Annahme zugrunde, daß dem Sprecher grundsätzlich zwei parallele Strategien zur Verfügung stehen, um orthographisches Material zu phonemisieren. Abb. 6.1 zeigt schematisch ein solches Dual Route-Modell:

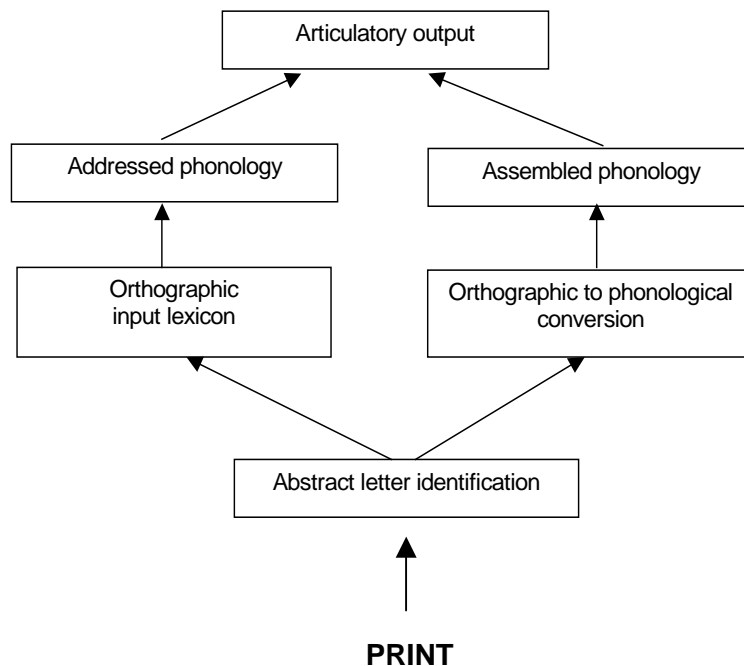


Abb 6.1: Dual Route-Modell der Phonemisierung im Leseprozess
(Abbildung nach Paap, Noel & Johansen 1992: 296)

Als Eingabe eines solchen Modells dient stets die orthographische Repräsentation („PRINT“), die vom Leser identifiziert und analysiert werden muß (hier als

abstract letter identification bezeichnet). Zur Phonemisierung dieser Eingabe stehen nun zwei parallele Strategien zur Verfügung:

(a) Der erste (jedoch nicht notwendigerweise primäre) Weg führt dabei über lexikalische oder morphologische Einheiten, die als Ganzes, einschließlich ihrer Aussprache, gelernt und in einem „mentalen Lexikon“ gespeichert sind (hier als *addressed phonology* bezeichnet). Bei solchen Einheiten kann die Phonemisierung vom Sprecher *ohne* den Einsatz von Regeln oder ähnlichen sequentiellen Strategien segmentaler Umsetzung abgerufen werden.

(b) Parallel dazu steht eine zweite Strategie zur Verfügung, bei der die orthographische Repräsentation (= Grapheme) mittels Zuordnungen oder Regeln in Phoneme überführt wird (*assembled phonology*). Kern dieser Strategie sind also Buchstaben-Laut-Zuordnungen, die in ihrer Gesamtheit die „Schriftaussprache-Regeln“ einer Sprache bilden.

Aus einem solchen Modell ergeben sich Konsequenzen auch für die Aussprache fremdsprachlichen Materials. Grundsätzlich haben auch nicht-muttersprachliche Sprecher Zugriff auf ein mehr oder weniger umfangreiches Inventar erlernter morphologischer oder lexikalischer Einheiten der Zielsprache; der Umfang dieses Inventars variiert in Abhängigkeit vom Grad ihrer L2-Kenntnis. Für dieses Inventar steht den Sprechern prinzipiell auch Strategie (a) zur Verfügung; sie können die Aussprache der betreffenden Lexeme oder Morpheme dieser Einheiten als Ganzes abrufen. Bei zielsprachlichem Material, bei dem diese Strategie nicht anwendbar ist – z.B. bei unbekanntem Wörtern oder Wortbildungselementen – besteht die Möglichkeit, Strategie (b) ergänzend hinzuzuziehen, d.h. Graphem-Phonem-Regeln anzuwenden. Wird jedoch Strategie (b) angewandt, so werden zugleich alle potentiellen Fehlerquellen aktiviert, die in diesem Kapitel im Zusammenhang mit einem Transfer der L1-Schriftaussprache genannt wurden: Prinzipiell können Sprecher hierbei Regeln ihrer L1, korrekte zielsprachliche Regeln, Regeln aus einer dritten Sprache oder auch Mischformen anwenden. Hieraus ergibt sich ein breites Spektrum potentieller Aussprachevarianten.

Eine hieraus ableitbare Hypothese besteht nun darin, daß die Anwendungshäufigkeit dieser beiden Strategien in Abhängigkeit von den jeweiligen L2-Kenntnissen des Sprechers variiert. Je ausgeprägter diese Kenntnisse sind, desto größer ist das dem Sprecher verfügbare Inventar lexikalischer oder morphologischer Einheiten, welche auf dem Wege einer *addressed phonology* phonemisiert werden können; ihre Aussprache wurde als Ganzes memoriert und kann somit auch als Ganzes abgerufen werden. Eine Konsequenz hieraus wäre, daß die Aussprache dieser Einheiten relativ stabil und weniger fehleranfällig ist. Dies gilt jedoch nur für Sprecher mit relativ guten L2-Kenntnissen. Sprecher mit geringen oder keinen L2-Kenntnissen hingegen profitieren von diesen parallelen Strategien nicht, da sie praktisch über kein memoriertes lexikalisch-morphologisches Inven-

tar verfügen. Hier überwiegt zwangsläufig Strategie (b), d.h. eine Phonemisierung nach sequentiell angewendeten Graphem-Phonem-Zuordnungsregeln.

Eine solche Hypothese stünde im Einklang mit der in den Sprachdaten dieser Arbeit zu beobachteten Tendenz, daß bei Sprechern mit guten L2-Kenntnissen die Akkuratheit der Aussprache von L2-Ortsnamen auch von deren morphologischem Aufbau abhängt. Bei Sprechern mit geringen oder keinen L2-Kenntnissen scheint ein solcher Einfluß dagegen nicht zu bestehen. Einige Beispiele sollen dies verdeutlichen: Es wurde gesagt, daß sich Ortsnamen oftmals aus lexikalisierten *und* onomastischen Wortbildungselementen (Morphemen) zusammensetzen. Anhand der Sprachdaten läßt sich zeigen, daß eine Reihe von lexikalischen Morphemen, beispielsweise

in englischen Namen: -*stone*, -*water*, -*castle*

in deutschen Namen: -*dorf*, -*berg*, -*hof*, *Nieder-*, *Bischofs-*

von den Probanden mit relativer Sicherheit und nur geringen Abweichungen von der kanonischen Aussprache realisiert werden. Demgegenüber weisen onomastische Morpheme – die auch im selben Namen auftreten können – weitaus ausgeprägtere Aussprachefehler auf. Die folgenden Beispiele illustrieren dies anhand einiger deutscher Ortsnamen. Die als Elemente der Namen auftretenden lexikalischen Morpheme erscheinen hier umrahmt:

<i>Rüdesheim</i>	Rüdes - heim
<i>Stadtkyll</i>	Stadt - kyll
<i>Oberpörlitz</i>	Ober - pörlitz
<i>Fünfstetten</i>	Fünf - stetten
<i>Kleingölit</i>	Klein - gölit

Charakteristisch für solcherart aufgebaute Namen ist, daß Aussprachefehler vorwiegend innerhalb der onomastischen Elemente auftreten, während der Aussprache lexikalischer Morpheme deutlich stabiler und weniger fehlerträchtig ist. Möglicherweise werden hier also bereits innerhalb einer Worteinheit zwei unterschiedliche Phonemierungsstrategien angewendet, mit dem Ergebnis, daß der Grad der Annäherung an die kanonische zielsprachliche Lautung zwischen onomastischen und lexikalisierten Morphemen variiert und somit unterschiedliche Approximationsgrade innerhalb eines Wortes auftreten können.

Eine aus dieser Beobachtung resultierende Konsequenz für die Modellierung von Aussprachefehlern liegt darin, daß die Anwendung bestimmter phonetischer Substitutionsregeln davon abhängig zu machen ist, in welchem Morphemtyp der

betreffende Laut auftritt. In der derzeitigen Implementierung der regelbasierten Modellierung ist diese Möglichkeit zwar prinzipiell vorgesehen, jedoch noch nicht technisch implementiert. Der Grund hierfür liegt darin, daß es zur automatischen Identifizierung lexikalischer Morpheme in der Eingabesequenz notwendig wäre, zunächst einen Abgleich mit einem Morphem- oder Ganzwortlexikon durchzuführen. Der Aufbau hierzu geeigneter Lexika für mehrere Zielsprachen übersteigt jedoch den Rahmen dieser Arbeit bei weitem, da jedes dieser Lexika bereits einen erheblichen Umfang aufweisen müßte, um nur die wichtigsten lexikalischen Morpheme der betreffenden Sprache abzudecken. Dennoch ist hierin eine sinnvolle potentielle Weiterentwicklung des in Kap. 7 dargestellten Regelsystems zu sehen.

6.5.2 Fehlerhafte Identifizierung von lexikalischen Morphemen

Bisher war die Rede davon, daß die Kenntnis morphologischer oder lexikalischer Einheiten der Zielsprache die akkurate Aussprache tendenziell erleichtert. In bestimmten Fällen jedoch wirkt die Fähigkeit eines Sprechers, innerhalb eines fremdsprachlichen Eigennamens bekannte lexikalische Morpheme zu identifizieren und deren Aussprache aus memorierten Einheiten herzuleiten, einer akkuraten zielsprachliche Aussprache entgegen, ist also als kontraproduktiv anzusehen.

Dies ist etwa dann der Fall, wenn eine Konstituente des fraglichen Namens eine hohe graphematische Ähnlichkeit mit einer bekannten morphologischen Einheit aufweist, so daß eine fehlerhafte Identifizierung von Morphemen beim Lesen auftritt. Das in dieser Studie verwendete Sprachmaterial enthält eine Reihe derartiger Stimuli, so z.B. der deutsche Ortsname *Hollenbek*, in dem von Sprechern verschiedener Muttersprachen das Morphem */hölle/* identifiziert wurde. Hieraus resultiert die phonetische Variante [hœlɪnbek]. Die graphemische Analogie bzw. die Ähnlichkeit mit lexikalischen Morphemen wird hier als eine eigenständige Fehlerursache wirksam. Bei der ihnen gestellten Aufgabe, fremdsprachliches onomastisches Material zu lesen, zu phonemisieren und zu artikulieren, versuchen Sprecher demnach aktiv, Einheiten zu identifizieren, die ihnen als strukturierendes Merkmal dienen können. Unter Umständen ziehen Sprecher hierbei jede ihnen bekannte linguistische Einheit heran. So traten beispielsweise beim italienischen Stimulus *Spotorno* mehrfach die phonetischen Varianten wie [spɔɐ̯t...] oder [spɔɐ̯t...] auf, welche offenbar auf dem fehlerhaft identifizierten Morphem */sport/* beruhen. Je geringer die beim Sprecher vorhandenen Kenntnisse der jeweiligen L2 sind, desto ausgeprägter scheint die Tendenz, auch solche – zunächst abwegig erscheinenden – Strukturmerkmale heranzuziehen, die wiederum als eigenständige Fehlerquelle bei der Aussprache wirksam werden können.

Ein weiterer Kontext, in dem eine Identifizierung bekannter morphologischer Einheiten zu Aussprachefehlern führt, liegt vor, wenn lexikalisierte Morpheme innerhalb von Namen auftreten, dabei aber eine vom Standard-Wortschatz abweichende Aussprache aufweisen. Dies ist beispielsweise der Fall bei englischen Ortsnamen, in denen lexikalische Konstituenten wie *-stone*, *-land* oder *-mouth* auftreten. Diese unterliegen oftmals einer Vokalreduktion und werden dementsprechend nicht mit einem Vollvokal produziert. Durch die Interferenz mit den entsprechenden lexikalischen Morphemen ergeben sich Varianten wie die folgenden:

Name	kanonische Lautung	lexikalisches Morphem	Variante
<i>Folkestone</i>	[fəʊkstən]	<i>-stone-</i> [stəʊn]	[fəʊkstəʊn]
<i>Bournemouth</i>	[bɔːnməθ]	<i>-mouth-</i> [maʊθ]	[bɔːnmaʊθ]
<i>Crooklands</i>	[krʊklændz]	<i>-land-</i> [lænd]	[krʊklændz]

Es handelt sich hierbei zwar um relativ geringfügige lautliche Abweichungen von der kanonischen Form, jedoch treten diese mit hoher sprecherübergreifender Regelmäßigkeit auch bei Sprechern mit sehr guten L2-Kenntnissen auf.

Ein weiterer Typ morphologisch bedingter Fehler tritt schließlich auf, wenn Morphemgrenzen (innerhalb von Ortsnamen) nicht erkannt werden und in Folge dessen fehlerhafte Ausspracheregeln zur Anwendung kommen. So tritt z.B. sowohl bei englischen als auch französischen Muttersprachlern häufig die Variante [ɛyːdəʃaɪm] für deutsch *Rüdesheim* auf. Diese Variante ist offensichtlich dadurch motiviert, daß die Sprecher die morphologische Struktur *Rüdes # heim* (# bezeichnet die Morphemgrenze) nicht zuverlässig identifizieren. Die Folge ist, daß sie für die in diesem Namen auftretende orthographische Sequenz <sh> eine Graphem-nach-Phonem-Regel <sh> → [ʃ] anwenden. Erst die morphologische Fehlinterpretation *Rüde [sh] eim* verursacht daher die fehlerhafte Variante [ɛyːdəʃaɪm]; es findet also eine Interaktion linguistischer Strukturebenen statt, welche unmittelbar auf die phonetische Beschaffenheit akzentbehalteter Aussprachevarianten einwirkt.

6.6 Kumulative Fehler und Grenzen der Typologie

Beispiele wie das zuletzt genannte zeigen deutlich, in welcher Weise einzelne linguistische Ebenen bei der Entstehung von Aussprachevarianten und -fehlern zusammenwirken können. Spezifische Fehler resultieren erst aus der *Interaktion*

der Ebenen. Fehler und Fehlertypen, die in den vorausgehenden Abschnitten analytisch getrennt wurden, treten in der Praxis in vielfältigen Kombinationen auf und bilden erst in dieser spezifischen Kombination die phonetischen Oberflächenformen, wie sie in den vorliegenden Sprachdaten zu beobachten sind.

Häufig bedingen einzelne Fehler einander, so daß es zu *kumulativen Fehlern* kommt, bei denen Fehler A erst die notwendige Bedingung für einen Folgefehler B bildet. Ein einfaches Beispiel illustriert diese Art von kumulativen Fehlern: Betrachtet man etwa die phonetische Variante [dʌs|dɔ:f], wie sie häufig bei englischen Sprechern für den deutschen Ortsnamen *Düsseldorf* auftritt, so ist auf den ersten Blick nicht offenkundig, welchem der in den vorausgehenden Abschnitten genannten Fehlertypen diese Variante zuzuordnen ist. Das Problem bildet hierbei die Substitution von deutsch [ʏ] durch englisch [ʌ]. Ein Lautersatz entsprechend einer Regel

$$[ʏ] \rightarrow [ʌ]$$

erscheint aus einer ausschließlich phonetischen Perspektive nicht motiviert, da beide Vokale zu verschieden sind, um als phonetische Äquivalente in Frage zu kommen. Ein Transfer einer Schriftaussprache-Regel des Englischen auf das Deutsche ist als alleinige Ursache ebenfalls auszuschließen, da das Englische nicht über das Graphem <ü> verfügt und somit keine entsprechende Regel mit diesem Graphem assoziiert ist. Betrachtet man die Aussprachevariante [dʌs|dɔ:f] hingegen im Sinne des zuvor Gesagten als einen kumulativen Fehler, so ist es durchaus möglich, sie als Resultat der in diesem Kapitel genannten Fehlertypen zu analysieren. Eine solche Analyse würde beispielsweise postulieren, daß hier ein sequentielles Auftreten zweier Fehlertypen vorliegt, bei dem der erste Fehler Voraussetzung und notwendige Bedingung des zweiten wirksam wird:

(1) **Graphemische Reinterpretation <ü> → <u>**

$$\begin{array}{l} <ü> \quad \quad \quad \rightarrow \quad <u> \\ <Düsseldorf> \rightarrow \quad <Dusseldorf> \end{array}$$

(2) **Anwendung einer muttersprachlichen Schriftaussprache-Regel**

$$\begin{array}{l} <u> \quad \quad \quad \rightarrow \quad [ʌ] \\ <Dusseldorf> \rightarrow \quad [dʌs|dɔ:f] \end{array}$$

Oftmals ist jedoch die Genese derartiger Fehler weniger transparent als diesem Beispiel. Sie bilden daher stets einen besonderen Problembereich bei einer regelbasierten Modellierung. Erschwerend kommt hinzu, daß für eine spezifische Aussprachevariante oftmals verschiedene konkurrierende Analysen bzw. Interpre-

tationen nebeneinander bestehen können, die identische phonetische Ergebnisse hervorbringen.

Eine vollständig befriedigende Lösung dieses Problem kann zumindest vorläufig nicht angeboten werden. Jedoch können mit Hilfe des Modells der Akzentstufen, wie es in Kap. 4.4 dargestellt wurde, einige kumulative Fehler durchaus realistisch und zutreffend modelliert werden. Das Modell ist so angelegt, daß höhere Akzentstufen (d.h. die Modellierung eines starken Akzents) stets die auf niedrigen Stufen (schwacher Akzent) auftretenden Fehler einschließen. Durch diesen Aufbau besteht die Möglichkeit, Fehler der hohen Akzentstufen als Folgefehler niedriger Akzentstufen zu modellieren. Eine Reihe der hier nur oberflächlich skizzierten kumulativen Fehler kann durch dieses Verfahren durchaus erfaßt werden.

Die in diesem Kapitel vorgeschlagene Fehlertypologie erhebt keinen Anspruch auf Vollständigkeit. Sie kann zweifellos um weitere Fehlertypen ergänzt werden, die zwar im engeren Sinne linguistisch motiviert sind, aber in den vorausgehenden Abschnitten nicht im Detail diskutiert wurden. Hierzu gehören beispielsweise folgende Typen, welche in den Sprachdaten vielfach beobachtet werden konnten:

- **Phonetische Übergeneralisierungen.** Verfügt eine L2 über einen Laut, der im Inventar der L1 des Sprechers nicht auftritt, so ist mitunter zu beobachten, daß L1-Sprecher den betreffenden L2-Laut in übergeneralisierender Weise produzieren, d.h. auch in solchen Kontexten anwenden, in denen der Laut nach den Regeln von L2 *nicht* auftreten sollte. Ein häufig zu beobachtendes Beispiel hierfür ist der Ersatz des englischen [e] durch [æ] bei Muttersprachlern des Deutschen (*head* → [hæd]). Offenbar wird der im Deutschen nicht auftretende Laut [æ] von den Sprechern als besonders charakteristisch für die Lautgestalt des Englischen empfunden und daher auch in inadäquaten Kontexten eingesetzt. Zum Teil resultieren hieraus auch Phonemverschiebungen in der Zielsprache (z.B. im Falle des genannten Beispiels *head* → [hæd] eine erhöhte Verwechselbarkeit von *had* [hæd], *head* [hed] und *hat* [hæt]).

- **Dritt Sprachen-Interferenz.** Hiermit ist der Fall gemeint, daß ein Sprecher bei der Aussprache von L2-Material Kenntnisse einer weiteren Fremdsprache (L3) heranzieht und diese auf L2 anwendet. Eine Interferenz tritt hier nicht zwischen L1 und L2, sondern vielmehr zwischen L3 und L2 auf. In den vorliegenden Sprachdaten finden sich viele Aussprachevarianten, die im Sinne einer Dritt Sprachen-Interferenz interpretiert werden können. Zwei Bedingungen scheinen dabei ihr Auftreten zu begünstigen: (i) Der Sprecher verfügt über nur geringe Kenntnisse der betreffenden L2; (ii) L2 und L3 werden vom Sprecher als „ähnliche“ Sprachen betrachtet. So konnte Dritt Sprachen-Interferenz häufig bei der

Zielsprache Niederländisch beobachtet werden. Italienische und Französische Sprecher mit guten Kenntnissen des Deutschen, aber keinen Niederländisch-Kenntnissen wenden hier oftmals die ihnen bekannten Regeln des Deutschen auf niederländisches Material an. Hieraus resultieren Aussprachevarianten, die nicht durch einen Einfluß der L1 des Sprechers erklärt werden können, sondern nur durch den Einfluß der intervenierenden Drittsprache (z.B. niederl. *Schiphol* [ʃχiphɔl] nach deutschen Regeln als [ʃifo:l] realisiert).

Fehlertypen wie die zuletzt genannten können grundsätzlich als Lautersetzungen beschrieben und somit in das hier entwickelte Verfahren einer regelbasierten Modellierung integriert werden. Im Gegensatz zu den zuvor in diesem Kapitel dargestellten Fehlern ist ihr Auftreten jedoch in deutlich höherem Maße von individuellen Voraussetzungen und Kenntnissen einzelner Sprecher beeinflusst. So kann beispielsweise ein Auftreten von Drittspracheninterferenz kaum zuverlässig prognostiziert werden, ohne die spezifischen Fremdsprachenkenntnisse des jeweiligen Sprechers zu berücksichtigen. Nicht zuletzt aus diesem Grunde werden Fehlerquellen wie diese im Rahmen dieser Arbeit – deren Schwerpunkt auf sprecherübergreifend auftretenden Fehlern liegt – zunächst nicht erfaßt. Eine Untersuchung der Interaktion zwischen diesen Fehlertypen und sprecherspezifischen Variablen kann jedoch prinzipiell auf den hier erarbeiteten Grundlagen erfolgen.

Darüber hinaus ist die vorgestellte Fehlertypologie in ihrem Erklärungsanspruch aber auch dadurch grundlegend eingeschränkt, daß vielfältige außerlinguistische Bedingungen spezifischer Aussprachevarianten, von denen einige in Kap. 2.6 genannt wurden (z.B. Motivation und psychologische Disposition des Sprechers, Kommunikationssituation und -partner u.v.m), nicht berücksichtigt wurden. Es ist jedoch nicht das Ziel dieser Arbeit, das gesamte situative und individuell-psychologische Bedingungsgefüge, welches bei der Produktion spezifischer Aussprachevarianten wirksam wird, so weit zu spezifizieren, daß eine unmittelbare Umsetzung in eine technische Anwendung möglich wird. Es muß beim derzeitigen Stand des Wissens ohnedies als fraglich gelten, ob eine solche Operationalisierung ein realistisches Forschungsziel darstellt. Sinnvoller ist es möglicherweise, eine technische Modellierung von Sprecherverhalten im Rahmen sprachtechnologischer Forschung stets als eine Approximation zu betrachten, deren Güte vor allem danach zu beurteilen ist, wie adäquat sie die sprachliche Realität im Hinblick auf den angestrebten Einsatzbereich der sprachtechnologischen Anwendung wiedergibt.

Es soll daher – im Sinne einer vereinfachten Arbeitshypothese – zunächst die Annahme zugrundegelegt werden, daß durch eine linguistisch motivierte Fehlertypologie sowie das Modell der Akzentstufen, wie sie bis hierher vorgestellt wurden, auch solche Varianten mit hinreichender Genauigkeit erfaßt werden können, die ihrerseits aus „außerlinguistischen“ Bedingungen resultieren.

III.

PROGRAMMTECHNISCHE UMSETZUNG

UND EVALUIERUNG

7 Regelbasierte Modellierung von Aussprachefehlern: Methoden, Regeltypen, Beispiele¹⁹

Im vorausgehenden Kapitel wurden Fehlertypen vorgestellt und diskutiert, die in der vorliegenden Sprachdatensammlung mit hoher Regelmäßigkeit auftraten. Diese Taxonomie bildet die unmittelbare Grundlage der technischen Implementierung einer regelbasierten Modellierung von Aussprachefehlern. Das folgende Kapitel beschreibt die programmtechnische Umsetzung des Regelsystems. Dabei stehen jedoch weniger technische Details der Implementierung im Mittelpunkt, sondern vielmehr das generelle Verfahren zur Umsetzung der bisher dargestellten Ergebnisse in ein praktisch einsetzbares Regelsystem.

Regelsätze wurden bisher für insgesamt neun Sprachrichtungen erstellt. Die Anzahl der Regeln pro Sprachrichtung ist dabei variabel. Die folgende Tabelle zeigt die bisher erfaßten Sprachrichtungen sowie die Anzahl der Einzelregeln für die jeweilige Sprachrichtung.

Sprachrichtung L1 → L2	Anzahl Einzelregeln
1. Englisch → Deutsch	98
2. Deutsch → Englisch	93
3. Französisch → Deutsch	102
4. Deutsch → Französisch	85
5. Englisch → Französisch	41
6. Französisch → Englisch	66
7. Italienisch → Deutsch	82
8. Deutsch → Italienisch	42
9. Italienisch → Englisch	15
GESAMT	624

Tab 7.1: Vorliegende Regelsätze, Anzahl der Einzelregeln je Regelsatz

Mit Ausnahme des Regelsatzes 9 für die Sprachrichtung L1 Italienisch → L2 Englisch, der sich noch in einem Versuchsstadium befindet, sind alle Regelsätze

¹⁹ Teile dieses Kapitels beruhen auf dem Artikel: Schaden, S. (2003b): "Generating non-native pronunciation lexicons by phonological rules". In: *Proceedings 15th International Conference of Phonetic Sciences (ICPhS 2003)*, Barcelona, 2545-2548.

so weit ausgearbeitet, daß wichtige Aussprache fehler bzw. -varianten, die innerhalb der betreffenden Sprechergruppe bzw. in der jeweiligen Sprachrichtung auftreten, mittels dieser Regelsätze modelliert werden können.

Grundsätzlich erscheint es jedoch nur bedingt sinnvoll, im vorliegenden Zusammenhang von einer „Vollständigkeit“ von Regelsätzen zu sprechen. Es existieren stets Aussprache fehler, die aufgrund der notwendigerweise begrenzten Datengrundlage dieser Studie nicht in den Regeln berücksichtigt sind. Weiterhin ist zu betonen, daß auch die im vorausgehenden Kapitel Fehlertypologie keinen Anspruch auf Vollständigkeit erheben kann. Es handelt sich somit *immer* um eine offene und potentiell zu erweiternde Menge von Einzelregeln. Darüber hinaus kann teilweise durch eine ökonomischere Regelformulierung – z.B. durch eine Zusammenfassung mehrerer Einzelregeln zu einer einzigen, allgemeineren Regel – eine Reduzierung des Umfangs von Regelsätzen erreicht werden, ohne daß ihre Leistung damit reduziert würde. Auch aus diesem Grunde ist es nicht möglich, aus der Anzahl der Regeln unmittelbar auf den Grad ihrer Ausarbeitung oder ihre Leistungsfähigkeit zu schließen.

7.1 Allgemeiner Aufbau des Regelsystems

Im Laufe der Durchführung der vorliegenden Arbeit wurden verschiedene Verfahren zur lexikalischen Modellierung akzentbehafteter Aussprache erprobt. Dabei wurden zwei Methoden fokussiert und technisch implementiert. Diese Modelle sollen in den folgenden Abschnitten mit ihren spezifischen Vor- und Nachteilen skizziert werden. Das erste und ältere Modell beruht auf einer Modellierung akzentbehafteter Aussprachevarianten auf der Grundlage der orthographischen Form des fremdsprachlichen Materials. Das zweite, neuere Modell, welches schließlich als besser geeignetes und überlegenes Modell für die aktuelle Implementierung übernommen wurde, basiert dagegen auf einer *postlexikalischen* Modifikation bereits existierender phonetischer Transkriptionen.

7.1.1 Modell I: Erweiterter Graphem-nach-Phonem-Umsetzer

Das erste Modell einer regelbasierten Modellierung von akzentbehafteter Aussprache beruht im wesentlichen auf der Hypothese, daß es möglich ist, ein Programm zur automatischen Graphem-nach-Phonem-Umsetzung – d.h. zur Umsetzung von Text in eine entsprechende phonetische Repräsentation in einer Lautschrift – so zu modifizieren, daß eine Phonemisierung nach den Aussprache-regeln der L1 des jeweiligen Sprechers automatisch erzeugt wird.

Ausgangspunkt war dabei vor allem die Überlegung, daß akzentbehaftete Aussprachevarianten häufig den Schriftaussprache-Regeln der L1 des Sprechers folgen. Somit sollte ein Programm zur Umsetzung von Text in eine phonetische Repräsentation, welches für die L1 des Sprechers erstellt wurde, annähernd eine „akzentbehaftete“ phonetische Transkription generieren, wenn man es – entgegen seinem ursprünglichen Bestimmungszweck – auf eine andere Sprache L2 anwendet. Würde z.B. der französische Name *Clermont-Ferrand* [kleʁmɔ̃ fɛʁɑ̃] vollständig nach deutschen Ausspracheregeln phonemisiert, so erhielte man eine phonetische Variante [kleʁmɔ̃nt fɛʁɑ̃nt], die einer möglichen Aussprachevariante eines deutschen Sprechers ohne jegliche Kenntnisse des Französischen entspricht.

Um nun darüber hinaus spezifische L2-Teilkenntnisse in diesem Modell zu integrieren, wurden zusätzliche Einzelregeln entworfen, die dem Graphem-nach-Phonem-Umsetzer hinzugefügt wurden. Um beispielsweise den Sachverhalt zu erfassen, daß vielen Sprechern des Deutschen die korrekte Lautung der französischen Graphemsequenz <eau> (z.B. in *Bordeaux*) als [o] bekannt ist, kann dem Graphem-nach-Phonem-Umsetzer für das Deutsche eine Regel <eau> → [o] hinzugefügt werden. Da die Graphem-Sequenz <eau> im Deutschen in einem vergleichbaren Kontext nicht auftritt²⁰, entsteht hierbei keinerlei Konkurrenz zu den originären GTP-Regeln für das Deutsche, so daß beide Regeltypen ohne Konflikte in einem einzigen Regelsatz kombiniert werden können. Das aus dieser Konzeption resultierende Basis-Modell kann graphisch wie folgt dargestellt werden:

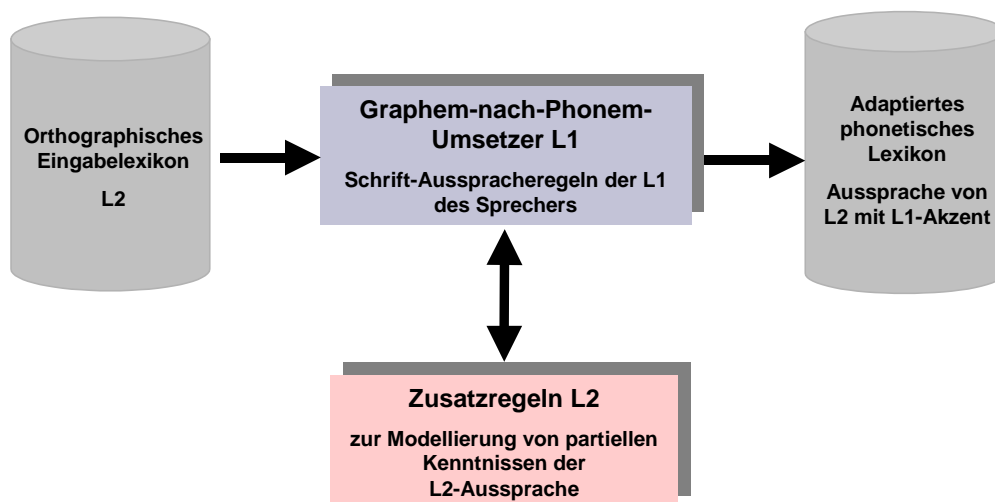


Abb. 7.1: Konzeption eines erweiterten Graphem-nach-Phonem-Umsetzers

²⁰ Genauer: Die Sequenz <eau> ist, wenn sie im Deutschen auftritt, stets durch eine Morphemgrenze unterbrochen – z.B. in *be-auftragen*. Alle anderen Vorkommen von <eau> innerhalb des deutschen Lexikons sind Entlehnungen aus anderen Sprachen (z.B. *Niveau, Plateau, Beauty*).

Die Grundidee dieses Systems besteht demnach darin, orthographische Sequenzen der Zielsprache nach den Schriftaussprache-Regeln der L1 des jeweiligen Sprechers zu phonemisieren, dabei jedoch Teilkenntnisse der L2-Aussprache mittels eines zusätzlichen Regelmoduls bzw. Teilregelsatzes einfließen zu lassen. Ein solches System wurde für die Sprachen Deutsch, Englisch, Französisch und Italienisch in verschiedenen Kombinationen erstellt und ist im Detail in Schaden (2002b) beschrieben.

Im praktischen Einsatz hat sich jedoch gezeigt, daß mit dem genannten Konzept eines erweiterten Graphem-nach-Phonem-Umsetzers verschiedene Probleme verbunden sind. Ein solches System muß prinzipiell imstande sein, zwei unter Umständen konträre Anforderungen zugleich zu erfüllen:

- (1) eine korrekte Graphem-nach-Phonem-Umsetzung nach den Regeln der Zielsprache sowie
- (2) eine gezielte Modellierung von Fehlern bzw. Aussprachevarianten, wie sie in akzentgefärbter Aussprache auftreten.

Gerade aufgrund dieser doppelten Anforderung hat sich diese Konzeption als nicht optimal erwiesen. Sie führte bald zu verhältnismäßig intransparenten Regelsystemen, die eine Modifikation und Wartung der Regeln stark erschwerten. Durch wechselseitige Abhängigkeiten der Regeln ergeben sich oftmals konfligierende Regelsequenzen, die insbesondere durch die doppelte Anforderung an das Regelsystem – sowohl korrekte Phonemisierung nach den Regeln der Zielsprache als auch adäquate Fehlermodellierung – bedingt sind. So ist beispielsweise eine vollständig fehlerfreie Graphem-nach-Phonem-Umsetzung eine wichtige Voraussetzung für die Anwendbarkeit der darauf aufbauenden Akzentregeln, welche eine korrekte L2-Form als Eingabe erwarten. Fehler in der Graphem-nach-Phonem-Komponente des Regelsystems produzieren Folgefehler bei der Anwendung der Akzentregeln – der Aufbau und Test der Akzentregeln wird dadurch deutlich erschwert.

Als ein weiteres Problem tritt hinzu, daß – wie bereits an früherer Stelle erwähnt – eine Phonemisierung orthographischer Sequenzen von L2 mittels unmodifizierter Regeln von L1 vielfach sehr unnatürliche Phonemfolgen generiert. So würde man beispielsweise für Namen wie *Geoffrey*, *Lawrence*, *Jacques*, *Françoise*, wenn man sie mittels eines deutschen Graphem-Phonem-Umsetzers phonemisieren läßt, Formen wie [ge:ofʁai], [lavʁentsə], [jakvəs] / [jatskvəs] oder [fʁaŋkoyzə] erhalten. Hierbei handelt es nicht um Varianten, die ein deutscher Muttersprachler typischerweise artikulieren würde. Die Ursache für die Generierung solcher artifizieller Formen ist u.a. darin zu sehen, daß die Eingabewörter oftmals Grapheme und Graphemsequenzen enthalten, die in den L1-Regelsätzen nicht vorgesehen sind. Weitere Probleme entstehen durch sprachspezifische Sonderzeichen (etwa dt. <ü, ö, ä> oder frz. <é, è, ê> etc.), welche in den entspre-

chenden Graphem-Phonem-Umsetzern nicht vorgesehen sind und daher zunächst nicht phonemisiert werden können. Eine entsprechende Modifikation der Programme bzw. Regelsätze ist daher erforderlich.

Es sprechen jedoch nicht allein praktisch-technische Erwägungen gegen den hier dargestellten Ansatz: Im Rahmen der Fehlertypologie in Kapitel 6 wurde dargelegt, daß ein vollständiger Transfer der muttersprachlichen Graphem-Phonem-Beziehungen auf L2 („globaler Transfer“) als Ausnahmefall zu betrachten ist. Weitaus häufiger sind die beobachteten Aussprachevarianten durch einen nur partiellen Transfer der Graphem-Phonem-Beziehungen gekennzeichnet („lokaler Transfer“). Die daraus resultierenden Aussprachevarianten lassen sich eher als Mischformen charakterisieren, die zum Teil den L2-Regeln, zum anderen Teil aber den L1-Regeln entsprechen. Eine Phonemisierung vollständiger L2-Graphemsequenzen nach unmodifizierten Regeln von L1 entspricht jedoch genau dem – weitaus selteneren – Fall eines *globalen* Transfers der Graphem-Phonem-Beziehungen, d.h. der Aussprache eines Sprechers, der die Schriftaussprache-Regeln seiner Muttersprache in naiver Weise auf zielsprachliche Wörter anwendet. Die charakteristischen phonetischen „Mischformen“ sowie die unterschiedlich ausgeprägten L2-Teilkenntnisse, von denen oben die Rede war, werden damit gerade nicht erfaßt.

Eigene Versuche mit einem solchen Ansatz haben deutlich gezeigt, daß vielfältige Modifikationen eines Graphem-Phonem-Umsetzers notwendig sind, um eine ansatzweise realistische Modellierung solcher Formen zu gewährleisten (cf. Schaden 2002b). Wenn auch der Nutzen dieser Methode für bestimmte Domänen der automatischen Spracherkennung zweifellos gegeben ist (cf. Cremelie & ten Bosch 2001; Schaden 2001), wurde im Rahmen dieser Arbeit ein anderer Weg beschritten.

7.1.2 Modell II: Postlexikalische Modifikation

Aus den zuletzt genannten Gründen wurde die Konzeption eines erweiterten Graphem-nach-Phonem-Umsetzers aufgegeben und durch ein Regelsystem ersetzt, welches eine bereits existierende kanonische Transkription mittels postlexikalischer Regeln *nachträglich modifiziert* und dabei mit spezifischen phonetischen Merkmalen akzentbehafteter Aussprache versieht.

Der wesentliche Unterschied zur ursprünglichen Konzeption ist darin zu sehen, daß die Aussprachevarianten bzw. -fehler hier nicht aus einer orthographischen Eingabe, sondern aus einer *bereits vorhandenen phonetischen Referenztranskription* generiert werden, welche erst nachträglich gezielt mit Aussprache Fehlern versehen wird. Wesentliche Elemente des ursprünglichen Konzepts (Fehlertypologie, Akzentstufung) konnten dabei übernommen werden;

insbesondere konnten die bereits erarbeiteten Regeln durch eine weitgehende Beibehaltung des Regelformats problemlos in das neue Konzept integriert werden.

Fehler, die durch einen Transfer der muttersprachlichen Schriftaussprache auf die Zielsprache bedingt sind, wurden im ursprünglichen System dadurch modelliert, daß Graphem-nach-Phonem-Regeln von L1 auf L2 angewendet werden. Diese Funktion wird im neuen Konzept übernommen von *graphemabhängigen Lautersetzungsregeln*, welche auf Informationen der orthographischen Ebene zurückgreifen können (dieser Regeltyp wird weiter unten in diesem Kapitel dokumentiert). Die folgende Abbildung zeigt den Gesamtaufbau eines auf dem Prinzip der postlexikalischen Modifikation beruhenden Systems:

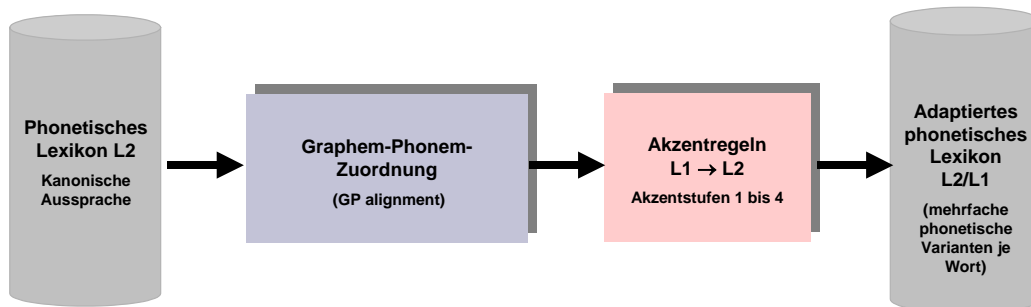


Abb. 7.2: Akzentmodellierung durch postlexikalische Regeln

Ein wesentlicher praktischer Vorteil dieses Konzepts liegt in der Entkopplung des Teilproblems „Graphem-Phonem-Umsetzung“ vom Problem der Modellierung akzentbehafteter Aussprache, welche den eigentlichen Fokus der vorliegenden Arbeit bildet. Das Resultat ist u.a. eine deutlich erhöhte Regeltransparenz sowie die Unabhängigkeit von möglichen Fehlern des verwendeten Graphem-nach-Phonem-Umsetzers, welche oftmals Folgefehler bei der Modellierung akzentbehafteter Aussprachevarianten verursachen.

Zudem ermöglicht es dieses Modell, die Regeln zur Generierung akzentbehafteter Aussprache auf verschiedene bereits existierende phonetische Lexika anzuwenden (z.B. die zu Forschungszwecken frei verfügbaren Lexika CMUDICT, UNISYN, PHONOLEX für die englische Sprache), um z.B. die Regeln an bisher unberücksichtigtem Wortschatz testen zu können. Dennoch ist bei Bedarf auch eine Kopplung des Regelsystems mit Graphem-nach-Phonem-Umsetzern problemlos möglich, indem eine Phonemisierung des Standard-Lexikons, welches als Systemeingabe dient, zunächst mittels eines solchen Programms automatisch generiert wird und erst in einem nachfolgenden, davon unabhängigen Schritt die Akzentregeln durchläuft.

In den folgenden Abschnitten sind die Komponenten eines Modells der postlexikalischen Generierung von Aussprachevarianten, wie es bis hierher nur skizziert wurde, im Detail dargestellt.

7.2 Komponenten des Regelsystems

7.2.1 Regeln und Regelinterpreter-Programm

Zur programmtechnischen Umsetzung des dargestellten Konzepts zur Generierung von Aussprachevarianten wurde ein Regelinterpreter-Programm in der Programmiersprache *Perl* (Wall 2001) entwickelt. Die Syntax bzw. der formale Aufbau der Regeln wurde in Teilen aus der Textvorverarbeitungs-Stufe (Graphem-nach-Phonem-Umsetzung) des Bochumer Sprachsynthese-Systems SyRUB übernommen (Böhm 1993) und mit einer Reihe von anforderungsspezifischen Erweiterungen versehen. Dennoch unterscheidet sich das entwickelte Modell in seiner Konzeption und Umsetzung deutlich von Programmen zur automatischen Graphem-nach-Phonem-Umsetzung; wesentliche Unterschiede werden im Verlauf dieses Kapitels näher erörtert. Unter anderem ist ein Modul für die segmentale Zuordnung von Graphemen und Phonemen innerhalb einer Eingabesequenz (*Graphem-Phonem-Alignment*) ein wesentlicher Bestandteil des Programms.

Als Eingabe erhält das Programm stets ein orthographisch und phonetisch codiertes Lexikon (zum generellen Aufbau solcher Lexika siehe Kapitel 4.2) und generiert hieraus mittels sprachenspezifischer Regelsätze alternative Aussprachelexika mit prototypischen nicht-muttersprachlichen phonetischen Varianten²¹. Die hierzu verwendeten phonetischen Regelsätze sind vom Quelltext des Programms getrennt abgelegt, so daß sie auch ohne Eingriff in den Programmcode geändert, erweitert und gewartet werden können. Wie oben dargestellt, wurde das Regelsystem in Form eines *postlexikalisch* operierenden Moduls realisiert. Dies bedeutet u.a., daß es auf bereits existente phonetische Lexika unterschiedlicher Herkunft anwendbar ist, welche außerhalb der vorliegenden Studie erstellt wurden. Um eine möglichst hohe Kompatibilität mit solchen vorhandenen oder zukünftig verfügbaren Aussprachelexika zu gewährleisten, wurde bei der Konzeption des Regelsystems darauf Wert gelegt, die Anforderungen an das Eingabelexikon hin-

²¹ Das allgemeine Verfahren einer *lexikalischen Adaption*, auf dem das Programm beruht, wurde bereits in Kapitel 4.2 dieser Arbeit skizziert, so daß dessen Grundlagen hier nicht mehr erörtert werden.

sichtlich seiner Struktur und der in ihm enthaltenen linguistischen Repräsentationsebenen so niedrig wie möglich zu halten.

Diese Konzeption erfolgte nicht zuletzt mit Blick auf die praktische Einsetzbarkeit des Regelsystems: Phonetische Lexika für Spracherkennungs- oder Sprachsynthesysteme verfügen oftmals nur über ein Minimum linguistischer Strukturinformationen. Zum Teil handelt es sich lediglich um einfache Wortlisten, in denen jedem Eintrag eine Aussprache (bzw. seltener: mehrere Aussprachevarianten) in einer computerlesbaren Lautschrift (z.B. SAMPA, ARPAbet, TIMIT u.a.) zugeordnet ist. Der folgende Ausschnitt illustriert dies am Beispiel des *Carnegie Mellon Pronouncing Dictionary* (CMUDICT), einem Aussprachelexikon für das Englische (derzeit insgesamt ca. 116.000 Einträge):

ABSOLUTE	AE1 B S AH0 L UW2 T
ABSOLUTELY	AE2 B S AH0 L UW1 T L IY0
ABSOLUTENESS	AE1 B S AH0 L UW2 T N AH0 S
ABSOLUTION	AE2 B S AH0 L UW1 SH AH0 N
ABSOLUTISM	AE1 B S AH0 L UW2 T IH2 Z AH0 M
ABSOLVE	AH0 B Z AA1 L V
ABSOLVE (2)	AE0 B Z AA1 L V
ABSOLVED	AH0 B Z AA1 L V D
ABSOLVED (2)	AE0 B Z AA1 L V D
ABSOLVES	AH0 B Z AA1 L V Z
ABSOLVES (2)	AE0 B Z AA1 L V Z
ABSOLVING	AH0 B Z AA1 L V IH0 NG

Tabelle 7.2: Ausschnitt aus dem *Carnegie Mellon Pronouncing Dictionary*, (Release 0.4) für amerikanisches Englisch

Wie das Beispiel zeigt, ist hier für jeden Lexikoneintrag lediglich die orthographische Form sowie eine dazugehörige phonetische Transkription angegeben. Für einige Einträge ist eine alternative Aussprache vorgesehen, welche mit (2) gekennzeichnet ist. Die einigen Phonemsymbolen nachfolgenden Ziffern (AA1, UW2 ...) liefern zudem Angaben zur Wortbetonung (Primär- und Sekundärbetonung). Darüber hinaus jedoch sind keinerlei strukturelle Angaben - etwa über Morphemgrenzen, Silbenstruktur, häufige allophonische Varianten, Korrespondenzen zwischen Graphemen und Phonemen etc. - zu finden. Eine wichtige Anforderung an das vorgestellte Regelsystem ist daher, daß bereits ein solches Minimum an Eingabeinformation hinreichend zur Anwendung der Regeln sein sollte.

7.2.2 Phonem-Graphem-Zuordnungen

Als minimale Anforderung an das phonetische Eingabelexikon des Regelsystems wird vorausgesetzt, daß neben der phonetischen Form eines Wortes stets auch seine orthographische Form verfügbar ist. Diese Minimalvoraussetzung ist bei praktisch allen existierenden maschinenlesbaren Aussprachelexika gegeben und setzt daher keine speziellen Vorverarbeitungs-Schritte des Eingabelexikons voraus.

Dabei wird die orthographische Form des zu verarbeitenden Wortes als eine wichtige Informationsquelle in die Regeln einbezogen, um Aussprachefehler zu modellieren, die durch die orthographische Repräsentation der Zielsprache verursacht werden (zu diesem Fehlertyp siehe Kapitel 6.4). Wie bereits weiter oben angedeutet, ist der Regeltyp der *graphemabhängigen Lautersetzungen* ein wesentlicher Bestandteil des vorgestellten Systems. Um solche Regeln formulieren zu können, ist es erforderlich, die orthographische Form des Wortes als eine zusätzliche Repräsentationsebene in die Regeln einzuführen und insbesondere eine eindeutige Zuordnung (engl. *alignment*) zwischen orthographischer und phonetischer Ebene automatisch zu definieren. Für jedes Phonemsegment eines Wortes muß daher zunächst ermittelt werden, durch welches Graphemsegment es repräsentiert wird. Im weiteren Verlauf dieses Kapitels wird deutlich werden, welchen Nutzen eine solche Zuordnung für die Modellierung akzentbehäfteter Aussprachevarianten aufweist.

Ausgangspunkt der Gestaltung des Regelsystems war zunächst der Umstand, daß maschinenlesbare Aussprachelexika oft in Form einfacher Listen vorliegen, bei denen die orthographische und phonetische Form innerhalb einer Datenzeile sequentiell hintereinander aufgeführt sind, wie etwa im folgenden Beispiel:

(1)	<i>Bexbach</i>	[bɛksbax]
(2)	<i>München</i>	[mʏnçŋ]
(3)	<i>Passau</i>	[p ^h asau]
(4)

Dabei verlaufen aber die graphemischen und phonetischen Repräsentationen eines Wortes nicht synchron, d.h. sie sind nicht aufeinander bezogen. Als Beispiel hierfür kann der Eintrag (1) *Bexbach* [bɛksbax] des oben gezeigten fiktiven Lexikonausschnitts herangezogen werden: Sowohl die orthographische Repräsentation dieses Ortsnamens als auch seine phonetische Umschrift [bɛksbax] setzen sich aus sieben Einzelsymbolen zusammen. Führt man nun zwei elementare Repräsentationsebenen **G** (Graphem) und **P** (Phonem) ein und bildet den Lexikoneintrag sequentiell darauf ab, indem man jedem Eingabesymbol (Graphem und Phonem) genau einen von sieben Sequenz-Schritten zuweist, so ergibt sich folgende fehlerhafte Zuordnung:

Sequenz-Schritt	1	2	3	4	5	6	7
G-EBENE	b	e	x	b	a	c	h
P-EBENE	b	ε	k	s	b	a	x

Offensichtlich entspricht diese Zuordnung nicht den tatsächlich vorliegenden Graphem-Phonem-Relationen: Im Deutschen kann ein Phonem grundsätzlich durch mehrere Grapheme repräsentiert sein; in gleicher Weise kann aber auch ein einzelnes Graphem mehrere phonetische Segmente repräsentieren. Eine einfache lineare Zuordnung der Ebenen G und P liefert daher meist nicht das gewünschte Ergebnis. Im Beispiel *Bexbach* ergeben sich mehrere Fehler zugleich: Das Graphem <x> repräsentiert phonetisch die Sequenz [ks], wird jedoch fälschlicherweise allein dem Phonem [k] zugeordnet. Ebenso wenig wird erfaßt, daß das Phonem [x] durch die zwei Graphem-Symbole <ch> repräsentiert wird. Darüber hinaus entstehen durch fehlerhafte Zuordnungen im linken Teil der Sequenz Folgefehler in den nachfolgenden Symbolen, so daß es schließlich zu vollkommen inadäquaten Zuordnungen wie ⇔ [s] oder <c> ⇔ [a] kommt.

Es ist daher erforderlich, für jedes Eingabewort zunächst eine korrekte Zuordnung der Ebenen G und P vorzunehmen. So wird für das Beispiel *Bexbach* die Information benötigt, daß im Deutschen folgende Repräsentationsbeziehungen zwischen Graphemen und Phonemen bestehen können:

<x> ⇔ [ks]
 <ch> ⇔ [x]

Erst mittels dieser Information ist es möglich, aus dem gegebenen Lexikon-eintrag eine korrekte Zuordnung der Ebenen G und P zu erzeugen, die sich wie folgt darstellen läßt:

G-EBENE	b	e	x	b	a	ch
P-EBENE	b	ε	ks	b	a	x

Im Rahmen dieser Arbeit wurde daher ein Programm-Modul erstellt, welches die Zuordnung von Graphem- und Phonem-Ebene automatisch vornimmt. Es beruht auf sprachspezifischen Zuordnungstabellen, in denen für jede untersuchte Sprache alle mehrgliedrigen Graphemsegmente (Digrapheme, Trigrapheme) sowie alle mehrgliedrigen Phonemsegmente (Diphthonge, Affrikaten) aufgeführt und

aufeinander bezogen sind. Mittels dieser Zuordnungen können graphemische und phonetische Eingabesequenzen automatisch so in Übereinstimmung gebracht werden, daß die Relation zwischen G-Ebene und P-Ebene in der oben skizzierten Form zutreffend repräsentiert ist. Das Modul ist den eigentlichen „Akzentregeln“ (phonetisch-segmentale Ersetzungen) vorgeschaltet; es bildet allerdings eine obligatorische Vorverarbeitungs-Stufe, damit die akzentgenerierenden Regeln in der gewünschten Weise arbeiten können.

Da die Zuordnungsrelationen zwischen den Ebenen G und P sprachenspezifisch sind, ist es notwendig, das Modul für jede Zielsprache anzupassen bzw. sprachspezifische Abbildungsregeln zu formulieren. Bisher wurden derartige Zuordnungs-Regeln für die Sprachen **Englisch**, **Französisch**, **Deutsch** und **Italienisch** formuliert und implementiert, wobei die Zuordnungen von unterschiedlicher Komplexität sind. So ist beispielsweise für das Französische (z.T. auch für das Englische) der Fall zu berücksichtigen, daß einige Grapheme *keinen* Laut repräsentieren: Wörter oder Namen wie *mère*, *gant*, *souris* oder *Hérault* weisen in ihrer orthographischen Form eines oder mehrere Symbole auf, die phonetisch nicht als eigene Laute realisiert werden (zur Diskussion solcher „stummer Phoneme“ im Französischen cf. Schane 1968 und Love 1981). Da jedoch genau diese Eigenschaft oftmals zu spezifischen Aussprache Fehlern bei Nicht-Muttersprachlern führt (siehe unten Kap. 7.4.1), ist es erforderlich, sie bereits bei der Zuordnung der Ebenen G und P korrekt zu erfassen.

Das folgende Beispiel zeigt eine automatische Zuordnung, wie sie vom Programm für den französischen Ortsnamen *Questembert* [kɛstãbɛʁ] durchgeführt wird, bei dem ein solches „stummes Phonem“ im Auslaut erscheint:

G-EBENE	qu	e	s	t	em	b	e	r	t
P-EBENE	k	ɛ	s	t	ã	b	ɛ	ʁ	-

Erläuterung: Die Sequenz <qu> repräsentiert im Französischen ein einzelnes Phonem [k]. Im Falle der französischen Nasalvokale wird eine Repräsentationsbeziehung zwischen der Graphemsequenz und dem Nasalvokal [ã] angenommen. Dies ist motiviert durch die Annahme, daß die Nasalität des Vokals durch den nachfolgenden Konsonanten bedingt ist und somit eine funktionale Einheit zwischen beiden Segmenten besteht (siehe hierzu auch die dazugehörige phonologische Regel in Kapitel 6.3). Eine Besonderheit besteht weiterhin darin, daß das wortfinale Graphem <t> einer Leerstelle (symbolisiert durch das Zeichen ' - ') auf der phonetischen Ebene zugeordnet wird. Diese Leerstelle ist dabei keineswegs ein redundanter Platzhalter für eine unbesetzte Position in der Eingabesequenz, sondern kann als funktionales Element für die Modellierung spezifischer Aussprache Fehler nutzbar gemacht werden. Kapitel 7.4.1 illustriert dies anhand eines Beispiels.

Das Zuordnungs-Modul arbeitet derzeit mit einer Fehlerrate von 2,0% (Deutsch), 4,2% (Englisch), 1,7% (Französisch) und 0,9% (Italienisch). Die unterschiedlichen Fehlerraten sind dabei jedoch nur zum Teil auf die einzelsprachliche Komplexität der Graphem-Phonem-Relationen zurückzuführen. Sie hängen auch damit zusammen, daß zum Test des Moduls für verschiedene Sprachen Lexika von unterschiedlicher Größe vorlagen. So konnte das Modul für Italienisch nur anhand eines Lexikons von 300 Einträgen geprüft werden, während für Deutsch Lexika von ca. 2000 Einträgen herangezogen werden konnten. Eine mittlere Fehlerrate von 2,2% (über alle vier Sprachen), wie sie sich aus den obigen Einzelwerten ergibt, stellt jedoch in jedem Falle einen akzeptablen und für den vorliegenden Zweck ausreichenden Wert dar.

7.2.3 Akzentstufenhierarchie und inkrementelle Regelanwendung

Eine Besonderheit des Regelsystems liegt darin, daß für jedes Eingabewort mehrere Aussprachevarianten generiert werden und dabei die phonetischen Eingabesequenzen in zunehmendem Maße mit Aussprache Fehlern versehen werden. Ziel ist es dabei, das Modell der Akzentstufen, wie es in Kapitel 4.4 dargestellt wurde, in möglichst unmittelbarer Weise in das Regelsystem einfließen zu lassen.

Diesem Modell zufolge kann das Kontinuum akzentbehafteter Aussprachevarianten, welches von minimalen lautlichen Abweichungen von der zielsprachlichen Standardaussprache (bei Sprechern mit guten Kenntnissen) bis hin zu erheblichen Lautverschiebungen (bei Sprechern mit minimalen Kenntnissen) reicht, in eine Abfolge prototypischer Varianten unterteilt werden, in denen charakteristische Aussprache Fehler von Sprechern der jeweiligen Kenntnisstufe repräsentiert sind. Zur Rekapitulation des Modells sei hier noch einmal die bereits aus Kapitel 4 bekannte Abbildung gezeigt:

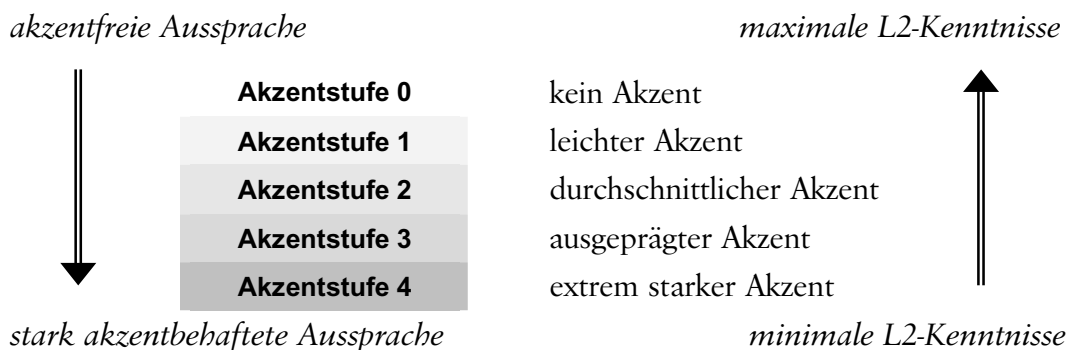


Abb. 7.3: Modell der Akzentstufen

Wie in Kapitel 4.4 erläutert wurde, basiert das derzeitige Modell auf 4 Akzentstufen, die ergänzt werden durch eine „Stufe 0“, welche der akzentfreien zielsprachlichen Standard-Aussprache entspricht. Akzentstufe 0 ist zwar als Teil einer regelbasierten Modellierung nicht von Interesse, da sie bereits in der kanonischen phonetischen Transkription des Eingabelexikons repräsentiert ist. Dennoch wird sie als potentielle Aussprachevariante unverändert in das Ausgabelexikon des Regelsystems aufgenommen, um die prinzipielle Möglichkeit einzuschließen, daß die L2-Sequenz vollständig akzentfrei nach den Regeln von L2 realisiert wird. Stufe 0 und Stufe 4 können somit als äußere Markierpunkte des Spektrums möglicher Varianten betrachtet werden.

Die eigentliche regelbasierte Modellierung umfaßt die Akzentstufen 1 bis 4. Jede Einzelregel ist mit einer Akzentstufe eindeutig markiert, mit der die Stärke des von dieser Regel generierten Akzents markiert ist. Das folgende einfache Regelpaar besagt beispielsweise, daß ein Ersatz des Vokals [æ] durch [ɛ] bereits auf einer niedrigen Akzentstufe 1 eintritt, während ein Ersatz des Konsonanten [θ] durch [s] erst auf Stufe 3 wirksam wird:

[æ] → [ɛ] (Stufe 1)

[θ] → [s] (Stufe 3)

Dabei gilt entsprechend der oben skizzierten inkrementellen Akzentstufung das Prinzip, daß höhere Akzentstufen die Anwendung von Regeln niedrigerer Stufen jeweils einschließen. Eine Ersetzung [æ] → [ɛ] erfolgt somit nicht nur auf Stufe 1, sondern auch auf den nachfolgenden Stufen 2 - 4. Die Akzentstufen 1 bis 4 weisen demnach die besondere Eigenschaft auf, daß in ihnen nicht vollständig verschiedene und einander ausschließende Fehler modelliert werden, sondern vielmehr eine *Zunahme* von Fehlern. Fehler, die auf Stufe 1 modelliert werden, bleiben auch in höheren Stufen erhalten, jedoch erweitert um zusätzliche Fehler, die typischerweise erst bei Sprechern mit geringen L2-Kenntnissen hinzutreten. Das Regelsystem ist demnach *inkrementell* aufgebaut: Sämtliche Aussprachefehler, die auf niedrigen Akzentstufen modelliert werden, werden in allen höheren Stufen übernommen. Durch die Markierung von Einzelregeln mit einer spezifischen Akzentstärke ist es möglich, die Konzeption diskreter Akzentstufen unmittelbar in die programmtechnische Realisierung einfließen zu lassen. Formal kann dies wie folgt dargestellt werden:

```

AL(G)    = zu generierende Akzentstufe
S        = Regelsatz der Sprachrichtung X → Y
Rule     = Einzelregel
AL(Rule) = spezifische Akzentstufe für Rule

for each Rule ∈ S
  if AL (Rule) <= AL (G) then
    APPLY (Rule)
end

```

Abb. 7.4: Formale Umsetzung der Akzentstufenhierarchie – höhere Ebenen schließen die Anwendung von Regeln niedrigerer Ebenen ein

Zur Illustration dieses Prinzips zeigt das folgende Beispiel die Programmausgabe des Regelsystems für den englischen Ortsnamen *Appleton Thorn* [æpɫtən θɔ:n] bei der Modellierung eines deutschen Akzents. In diesem Beispiel tritt das oben genannte Regelpaar [æ] → [ɛ] und [θ] → [s] auf den entsprechenden Akzentstufen in Kraft:

Appleton Thorn	[æpɫtən θɔ:n]	angewendete Regeln
Stufe 1	[ɛpɫtən θɔ:n]	Stufe 1
Stufe 2	[ɛpɫtən θɔ:n]	Stufe 1 und 2
Stufe 3	[ɛpɫtən sɔ:n]	Stufe 1 bis 3
Stufe 4	[aplətən tɔ:n]	Stufe 1 bis 4

Auf der ersten Akzentstufe wird die Lautersetzung [æ] → [ɛ] sowie die Substitution von [ɔ:] durch [ɔɹ] modelliert. Weiterhin wird die im Englischen in postvokalischer Position auftretende velare Realisierung des Phonems /l/ als [ɫ] („dark L“) durch die alveolare Realisierung [l] ersetzt. Auf Stufe 2 wird die englische Regel der Vokalreduktion (siehe hierzu auch unten 7.4.2) außer Kraft gesetzt, wodurch ein Ersatz des reduzierten Vokals [ə] durch den Vollvokal [ɔ] eintritt. Stufe 3 beinhaltet einen Ersatz des englischen Interdentals [θ] durch deutsches [s]. Auf Stufe 4 schließlich wird die deutsche Schriftaussprache vollständig auf den Ortsnamen angewendet, wodurch sich eine von der kanonischen Form stark abweichende Phonemisierung [aplətən tɔ:n] ergibt.

Die Akzentstufung, wie sie bis hierher skizziert wurde, kann auf vielfältige Weise modifiziert und flexibel den jeweiligen Erfordernissen angepaßt werden. Neben einer Erweiterung um (beliebig viele) zusätzliche Akzentstufen, welche problemlos vorgenommen werden kann, ist es auch möglich, nur *eine* Akzentstufe generieren zu lassen. Dies ist beispielsweise dann sinnvoll, wenn nur solche Aussprachevarianten in das Lexikon aufgenommen werden sollen, die eine erhebliche phonetische Abweichung von der kanonischen Form aufweisen (z.B. für ein Sprachsynthese-System, mit dem eine stark akzentbehafte Aussprache generiert werden soll). Weiterhin ist als eine künftige Erweiterung denkbar, Regeln für einzelne Akzentstufen mit a-priori-Wahrscheinlichkeiten zu versehen. Mittels dieser Methode kann das Modell der Akzentstufen möglicherweise genauer an eine spezifische Sprechergruppe adaptiert werden, indem die in der zu modellierenden Sprechergruppe vorherrschenden L2-Kennntnisstufen mit höheren Auftretenswahrscheinlichkeiten versehen werden.

7.3 Regeltypen

7.3.1 Kontextfreie phonetische Substitutionsregeln

Wie in Kapitel 6.2 dargelegt, besteht ein wesentliches Merkmal akzentgefärbter Aussprachevarianten darin, daß zielsprachliche Phoneme bzw. Allophone durch abweichende Laute „ersetzt“ werden. Diese allgemeine Feststellung gilt zunächst unabhängig von der Frage, *welche* L2-Laute einer Substitution unterworfen sind und *welche* Ersatzlaute für sie eintreten. Letzteres ist für eine formale Darstellung der entsprechenden Lautsubstitutionen ist nicht unmittelbar relevant; zunächst steht vielmehr die allgemeinere Frage im Vordergrund, über welche Mechanismen ein regelbasiertes System verfügen sollte, um die auftretenden Substitutionen zutreffend modellieren zu können.

Im vorgestellten System werden Aussprachevarianten, die auf phonemischen oder allophonischen Ersetzungen beruhen, grundsätzlich durch die Anwendung von Phonem-Abbildungsregeln modelliert. Regeln dieses Typs wurden im Verlauf dieser Arbeit bereits mehrfach besprochen. Die Regeln beruhen dabei stets auf der folgenden einfachen Basisform:

$$X \quad \rightarrow \quad Y$$

X bezeichnet hier ein Eingabesymbol und Y ein Ausgabesymbol, durch das X ersetzt wird. X ist im vorliegenden Zusammenhang ein L2-Laut (die Zielform), Y ist der Ersatzlaut, der an dessen Stelle eintritt. Bereits mittels solcher einfacher Zuordnungen kann eine zielsprachliche phonetische Sequenz in Ansätzen in eine „akzentbehaftete“ Sequenz transformiert werden. Die folgende Tabelle zeigt eine einfache, kontextfreie Lautzuordnung für die Sprachrichtung L1 Deutsch → L2 Englisch:

Englisch	Deutsch	Beispiel
[ɹ]	[ʀ]	<i>red</i> → [ʀet]
[ð]	[z]	<i>the</i> → [zə]
[θ]	[s]	<i>thorn</i> → [sɔ:n]
[æ]	[ɛ]	<i>hat</i> → [hɛt]
[əʊ]	[o:]	<i>rose</i> → [ʀo:s]

Mit Hilfe dieser fünf Zuordnungen ließe sich beispielsweise der englische Satz

There is a red rose on the hat [ðeəɪ ɪz ə .ɪed .ɹəʊz ɔn ðə hæʔ]

in die folgende phonetische Variante transformieren, die Merkmale eines starken deutschen Akzentes trägt:

There is a red rose on the hat [zeəʀ ɪz ə ʀed ʀo:z ɔn zə hɛʔ]

Bei derartigen Zuordnungen handelt es sich generell um die einfachste Form von Regeln, die im vorliegenden System angewandt wurden. Einige der in den Sprachdaten auftretenden Lautersetzungen können auf diese Weise adäquat formuliert werden. Dies gilt etwa für das deutsche Phonem /R/, dessen deutsche Realisierungsvarianten als uvularer Frikativ [ʀ] oder alveolarer Vibrant [r] von englischen Muttersprachlern zumeist durch den Approximanten [ɹ] ersetzt werden. Dieser Lautersatz ist relativ unabhängig vom lautlichen Kontext und kann somit als kontextunabhängige Ersetzungsregel [ʀ] → [ɹ] formuliert werden. Kontextunabhängige Lautersetzung kann darüber hinaus eingesetzt werden bei der Modellierung von Aussprachefehlern, die auf einer Äquivalenzklassifizierung (siehe Kapitel 2.4) beruhen: Derartige Fehler treten auf, wenn zu einem L2-Sprachlaut ein akustisch und perzeptiv ähnlicher, jedoch nicht identischer Laut im Inventar von L1 existiert (z.B. deutsch [ʀ] vs. englisch [ɹ]).

Generell ist kontextunabhängige Lautersetzung eher geeignet zur Modellierung moderater Lautverschiebungen, die bei Sprechern mit verhältnismäßig guten Kenntnissen der Zielsprache auftreten. Dieser Regeltyp ist jedoch in mindestens zweifacher Hinsicht unzureichend für eine Modellierung sämtlicher Fehler: Ers-

tens ist für eine Vielzahl von L2-Lauten keineswegs klar zu bestimmen, welches das optimale L1-Äquivalent ist, das als Ersatzlaut eintritt. Akustische oder artikulatorische Ähnlichkeit wird zwar oftmals als Faktor wirksam, ist jedoch keinesfalls ein verlässlicher Parameter, mittels dessen die lautlichen Abweichungen exakt prognostiziert werden können. Zweitens liefern die im Rahmen dieser Arbeit erhobenen Sprachdaten deutliche Evidenz dafür, daß phonetische und/oder orthographische Kontexte des ersetzten Lautes einen wesentlichen Einfluß auf die Wahl des Ersatzlautes ausüben. Solche Kontextfaktoren können die Anwendung einer spezifischen Substitution auslösen, restringieren oder vollständig blockieren. Es ist daher oftmals erforderlich, *Kontextbedingungen* für die Anwendung einzelner Regeln zu formulieren. Im nächsten Abschnitt ist dargestellt, in welcher Weise solche Restriktionen implementiert wurden.

7.3.2 Kontextbedingungen (positionsbedingte Substitutionen)

Häufig sind Ersetzungsregeln nur eingeschränkt gültig, d.h., die betreffenden Lautersetzungen werden nur dann wirksam, wenn der zu ersetzende Laut in einer spezifischen Umgebung auftritt. So ist beispielsweise eine Ersetzung des ungespannten Kurzvokals [ɪ] durch den gespannten Langvokal [i:], die bei deutschen Sprechern mit der Zielsprache Englisch zu beobachten ist, auf nur *eine* spezifische Position im Wort beschränkt. Sie tritt ausschließlich in wort-finaler oder morphem-finaler Stellung auf, z.B. in folgenden Beispielen:

Wort	kanonische Aussprache	deutscher Akzent
<i>party</i>	[pa:tɪ]	[paɪti:]
<i>Harry</i>	[hæ.ɪ]	[heri:]
<i>Kimberley</i>	[k ^h ɪmbəlɪ]	[k ^h ɪmbɔ̃li:]

Regeln für Lautersetzungen sind daher so zu formulieren, daß der lautliche Kontext berücksichtigt werden kann. Dabei kommen nicht nur Restriktionen in Frage, bei denen die Ersetzung auf einen einzigen Kontext beschränkt bleibt, wie im genannten Beispiel [ɪ] → [i:], sondern auch wechselnde Ersatzlaute, die vom jeweiligen Kontext bestimmt werden. Beispielsweise wird deutsch [a:] von englischen Sprechern in einigen Kontexten durch den Diphthong [eɪ], in anderen Kontexten dagegen durch [ɑ:] substituiert. So wird deutsch *Ravensberg* häufig als [ɹæɪvənsbɜ:g] wiedergegeben, wohingegen der Name *Rosenthal*, in dem der gleiche Vokal [a:] auftritt, als [ɹo:zəntɑ:l] realisiert wird. Eine entsprechende Regel, mit der diese Alternation erfaßt wird, könnte lauten: *Ersetze [a:] durch [eɪ], wenn ein Konsonant und mindestens ein Vokal folgen; ersetze [a:] durch [ɑ:] in allen anderen Kontexten.* Zugleich reflektiert die Distribution dieses Laut-

ersatzes eine spezifische Graphem-Phonem-Beziehung des Englischen, die hier auf das Deutsche übertragen wird (siehe dazu auch unten 7.3.4).

Um derartige Fälle zu erfassen, enthält ein großer Teil der bisher formulierten Regeln Kontextbedingungen, um ihren Anwendungsbereich einzugrenzen und übergenerierende Regeln zu vermeiden. Sofern der benötigte Kontext ausschließlich auf der phonetisch-phonologischen Ebenen liegt, kann bei der Formulierung derartiger Regeln auf etablierte Notationskonventionen der generativen Phonetik zurückgegriffen werden (cf. z.B. Kenstowicz & Kisseberth 1979, Kap. 9). Die Basisform dieser Regeln ist:

$$X_{L_2} \rightarrow Y_{L_1} \quad / \quad LC _ RC$$

Hierbei wird ein phonetisches Segment X_{L_2} (Element der kanonischen L2-Aussprache) durch ein Segment Y_{L_1} ersetzt, sofern die unmittelbaren linken (LC) und rechten (RC) Kontexte erfüllt sind, d.h. sofern LC und RC adjazente Segmente des Eingabesymbols X_{L_2} sind. Der Index L1 in Y_{L_1} kennzeichnet den Sachverhalt, daß hier ein Sprecher einer zu spezifizierenden L1 einen Ersatzlaut artikuliert; er kennzeichnet *nicht* notwendigerweise die Zugehörigkeit des Ersatzlautes zum L1-Lautinventar.

In Fällen, in denen eine Regel nicht nur für einzelne Segmente anwendbar ist, sondern für ganze Phonemklassen gilt, können X und Y ebenso wie die Kontexte LC und RC als phonetische Merkmalsbündel geschrieben werden. Das folgende Beispiel zeigt eine solche Notation für die im Deutschen vorliegende Regel der sog. ‚Auslautverhärtung‘ (Entstimmlichung von Obstruenten im Silbenauslaut):

$$\left[\begin{array}{l} + \text{obstruent} \\ + \text{voiced} \end{array} \right] \rightarrow \left[\begin{array}{l} + \text{obstruent} \\ - \text{voiced} \end{array} \right] \quad / \quad _ \#$$

Eine solche Notation kann als kompaktes Darstellungsinstrument zur Modellierung von Aussprachevarianten verwendet werden, bei denen Sprecher systematisch auftretende phonologische Regeln ihrer L1 auf L2 transferieren. Solche Varianten wurden in Kap. 6.3 als ein charakteristisches Kennzeichen fremdsprachlicher Akzentvarianten beschrieben (z.B. die Übertragung der deutschen ‚Auslautverhärtung‘ auf das Englische). Sie sind u.a. dadurch gekennzeichnet, daß sie oftmals für ganze Phonemklassen gelten; insofern ist eine Darstellung als Merkmalsbündel eine sinnvolle und oftmals ökonomischere Alternative zur Auflistung sämtlicher Einzelsegmente, auf die die betreffende Regel anzuwenden ist.

7.3.3 Formale Darstellung von Kontextspezifizierungen

Bei der Formulierung eines Regelkontextes können grundsätzlich die folgenden Typen unterschieden werden:

(i) Positive Kontextspezifizierung

Eine Ersetzung wird ausschließlich bei Vorliegen eines Kontexts Z vollzogen, andernfalls ist die Regel nicht anwendbar und es tritt kein Lautersatz ein. Formal ist dieser Regeltyp wie folgt darstellbar:

$$X \rightarrow Y / Z$$

Mittels dieses Regeltyps können Ersetzungen formuliert werden, die nur in spezifischen lautlichen Umgebungen zur Anwendung kommen. Das folgende Beispiel zeigt eine solche Regel für Sprachrichtung L1 Deutsch \rightarrow L2 Englisch, bei der kurzes, ungespanntes [ɪ] durch langes, gespanntes [i:] ersetzt wird:

$$[\text{ɪ}] \rightarrow [\text{i:}] / _ \#\#$$

Die Ersetzung $[\text{ɪ}] \rightarrow [\text{i:}]$ beschränkt sich dabei ausschließlich auf wortfinale Positionen. Dies wird mittels des Symbols $\#\#$ (= Wortgrenze) signalisiert, das hier als obligatorischer rechter Kontext angegeben ist.

(ii) Negative Kontextspezifizierung

In einigen Fällen ist es zugunsten einer möglichst ökonomischen Regelformulierung sinnvoller, anstelle einer positiven Spezifizierung diejenigen Kontexte auszuschließen, in denen die Regel *nicht* zur Anwendung kommt. Hier wird eine Ersetzung durch die Anwesenheit eines Kontextes Z blockiert, d.h. sie ist hier nicht anwendbar.

$$X \rightarrow Y / \neg Z$$

Die folgende (fiktive) Regel ersetzt [i:] durch [aɪ] in allen Kontexten mit Ausnahme eines nachfolgenden Lautes [g]:

$$[\text{i:}] \rightarrow [\text{aɪ}] / _ \neg [\text{g}]$$

Dieser Regeltyp läßt sich prinzipiell stets umkehren und als positiv spezifizierte Regel schreiben. Die Entscheidung, eine negativ spezifizierte Regel anstelle von

Regeltyp (i) zu formulieren, hängt vom Gültigkeitsbereich der Regel ab: Überwiegen die Fälle, in denen die Regel zur Anwendung kommt (*Default-Bedingung*), ist eine negative Formulierung oftmals ökonomischer.

(iii) Alternative Kontextspezifizierungen bei gleicher Substitution

Hier wird eine Ersetzung durch mehrere alternative Kontexte legitimiert, von denen jedoch nur einer erfüllt sein muß, um die Regel zur Anwendung zu bringen (dies entspricht einer Booleschen ODER-Verbindung). Der Ersatzlaut Y ist hier jedoch stets derselbe.

$$X \rightarrow Y \quad / \quad Z_1 \vee Z_2 \vee Z_3 \vee \dots Z_n$$

Dieser Regeltyp ist in der Praxis nur selten erforderlich; er kann jedoch sinnvoll sein, wenn eine Zusammenfassung alternativer Kontexte zu einem zu komplexen und damit unübersichtlichen und schlecht handhabbaren Ausdruck führt.

(iv) Alternative Kontextspezifizierungen bei variierender Substitution

Hier sind für X verschiedene Ersetzungen möglich: Ein zielsprachlicher Laut X wird im Kontext Z_1 durch den Laut Y_1 ersetzt, in Kontext Z_2 dagegen durch den Laut Y_2

$$X \rightarrow Y_1 \quad / \quad Z_1$$

$$X \rightarrow Y_2 \quad / \quad Z_2$$

Regelformulierungen dieses Typs werden benötigt, falls L2-Laute in Abhängigkeit von ihrer phonetischen Umgebung unterschiedlichen Substitutionen unterworfen sind. Hinsichtlich der Regelanordnung gilt hierbei als allgemeines Prinzip, daß stets die spezielleren (und möglicherweise nur singular auftretenden) Fälle zuerst aufgeführt und angewendet werden und der allgemeinste Fall als *Default-Bedingung* zuletzt angewendet wird, um zu verhindern, daß die *Default-Bedingung* die Eingabesymbole der spezifischeren Regeln vorzeitig eliminiert (cf. die sog. *Elsewhere Condition* bei Kiparsky 1973). Das Beispiel zeigt eine solche (vereinfachte) Regel für die Sprachrichtung L1 Englisch \rightarrow L2 Deutsch:

$$\begin{aligned} [\text{ʁ}] &\rightarrow [\text{ɪ}] \quad / \quad _ [+ \text{VOKAL}] \\ [\text{ʁ}] &\rightarrow \emptyset \quad (\textit{elsewhere}) \end{aligned}$$

Dieses Regelpaar besagt, daß die deutsche Standard-Realisierung des ‚r‘-Lautes, ein uvulares [ʁ], von englischen Sprechern durch den Approximanten [ɪ] ersetzt

wird. Eine Substitution findet hier ausschließlich in prävokalischer Position statt (Teil 1 der Regel). In allen anderen Positionen (Teil 2 der Regel) wird der Laut [ɛ] durch das Nullsegment \emptyset ersetzt, d.h. vollständig aus der Eingabesequenz getilgt.

7.3.4 Graphemabhängige Lautsubstitutionen

Wird das fremdsprachliche Material *gelesen*, so treten spezifische Aussprache Fehler auf, deren Ursache in der orthographischen Repräsentation des Materials liegt. Wie bereits in Kap. 6.4 ausgeführt, spielen hier insbesondere solche Fehler eine Rolle, die durch einen Transfer der muttersprachlichen Graphem-Phonem-Beziehungen auf die Zielsprache verursacht werden. Die vom Sprecher produzierten Ersatzlaute sind hierbei abhängig von der orthographischen Repräsentation des ersetzten Lautes. Hinsichtlich einer regelbasierten Modellierung dieses Fehlertyps bedeutet dies, daß sich entsprechende Regeln nicht länger als „autonome“ phonetische Substitutionen, sondern nur unter Rückgriff auf die graphematische Ebene und deren Relationen zur phonetischen Ebene formulieren lassen.

Ein einfaches Beispiel verdeutlicht diesen Sachverhalt: Bei englischen Sprechern mit der Zielsprache Deutsch ist häufig eine Substitution des Frikativs [v] durch den Approximanten [w] zu beobachten (z.B. im deutschen Namen *Wernitz*, realisiert als [weʌnɪts]). Mittels der bisher skizzierten Regeltypen ließe sich dies durch eine Einzelregel wie die folgende formulieren:

$$[v] \rightarrow [w] \quad / \quad (\text{optionaler Kontext})$$

Die Sprachdaten zeigen jedoch, daß eine Substitution $[v] \rightarrow [w]$ nur dann auftritt, wenn [v] orthographisch durch <w> repräsentiert ist. In allen anderen Fällen kommt die Regel *nicht* zur Anwendung, wie das folgende Beispielpaar zeigt:

(1) *Wernitz* [wɜ:nɪts] (Anwendung der Regel $[v] \rightarrow [w]$)

(2) *Ravensburg* [ra:vʌnsbuʁk] (keine Anwendung der Regel)

Die Ursache für diesen Fehler ist in dem Umstand zu sehen, daß englische Sprecher eine Schriftaussprache-Regel ihrer Muttersprache

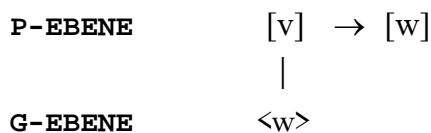
$$\langle w \rangle \rightarrow [w]$$

auf die Zielsprache Deutsch übertragen. Daher tritt Lautersetzung *nur* bei orthographischem <w>, nicht jedoch bei <v> auf. Für eine regelbasierte Modellierung folgt daraus, daß eine Kontextbedingung für die Regel $[v] \rightarrow [w]$ nicht

innerhalb der phonetischen Ebene formuliert werden kann, sondern auf orthographische Information zugreifen muß. Es muß eine Möglichkeit bereitgestellt werden, einzelne Lautsubstitution an bestimmte *graphemische Kontexte* bzw. *graphemische Repräsentationen* zu binden.

Als Alternative zu der zu Beginn dieses Kapitels beschriebenen Methode einer Graphem-Phonem-Umsetzung nach den Regeln der L1 des Sprechers wurde daher entschieden, die fraglichen Aussprachefehler (d.h. sämtliche Fehler, die durch Übertragung der Schriftausprache entstanden sind) mittels *graphemabhängiger Lautsubstitutionen* zu modellieren. In diesem Regeltyp werden Lautersetzungen an spezifische graphemische Repräsentationen des zu ersetzenden L2-Lautes gebunden. Hierzu wird zunächst mittels einer automatischen Graphem-Phonem-Zuordnung (beschrieben oben in 7.2.2) eine lineare Zuordnung zweier Ebenen G (Graphem) und P (Phonem) erzeugt, zwischen denen eine Repräsentationsbeziehung angenommen wird: Die einzelnen Segmente dieser Ebenen werden durch diesen Vorgang durch unmittelbare Assoziationen miteinander verbunden (siehe Kap. 7.2.2).

Im Anschluß können nun einzelne Regeln so formuliert werden, daß Ersetzungen auf einer der beiden Ebenen an Bedingungen auf der jeweils anderen Ebene geknüpft werden. Lautersetzungen (d.h. Ersetzungen auf der Ebene P) können somit z.B. durch Bedingungen restringiert werden, die ihrerseits nicht selbst auf der Ebene P angesiedelt sind. Für das oben dargestellte Beispiel englischer Sprecher, die das deutsche <w> phonetisch als [w] realisieren, kann eine solche Restriktion wie folgt dargestellt werden:



Die Verbindungslinie zwischen G-Ebene (Graphem) und P-Ebene (Phonem) zeigt an, daß der Laut in der Regeleingabe [v] orthographisch durch <w> repräsentiert sein muß, damit die Regel anwendbar ist. Diese Bindung ist obligatorisch, andernfalls kann die Regel nicht angewendet werden. Mit der obligatorischen Bindung der Lautersetzung [v] → [w] an die graphemische Repräsentation <w> wird sichergestellt, daß die Regel etwa im Falle von *Ravensburg* oder *Grevenbroich* nicht anwendbar ist, während sie andererseits auf Namen wie *Wernitz*, *Wiesbaden* oder *Würzburg* zutrifft.

Es hat sich in der praktischen Anwendung gezeigt, daß der Regeltyp der graphemabhängigen Lautsubstitution geeignet ist, um eine Vielzahl charakteristischer Aussprachefehler adäquat zu modellieren, bei denen die orthographische Repräsentation als Ursache wirksam wird. Das Spektrum reicht dabei von gering-

fügen Lesefehlern, bei denen die muttersprachlichen Schriftaussprache-Regeln nur teilweise auf L2-Wörter transferiert werden (= *lokaler Transfer*, siehe Kapitel 6.4.1), bis hin zu einer Modellierung eines nahezu vollständigen Transfers dieser Regeln auf die Zielsprache (= *globaler Transfer*), der in einer stark akzentgefärbter Aussprache resultiert. Gegenüber einer automatischen Phonemisierung zielsprachlicher Wörter mittels eines Graphem-Phonem-Umsetzers nach den Regeln von L1, wie sie zu Beginn dieses Kapitels beschrieben wurde, bietet diese Methode darüber hinaus den Vorteil, daß sie postlexikalisch, d.h. ohne die Notwendigkeit eines Eingriffs in das kanonische Lexikon angewendet werden kann.

7.4 Beispielanwendungen der Regeln

In den folgenden Abschnitten werden abschließend einige Beispiele präsentiert, wie die in den vorausgehenden Abschnitten genannten Regeltypen zur Modellierung charakteristischer Aussprachefehler verwendet werden können. Der Schwerpunkt der Beispiele liegt dabei auf solchen Regeln, in denen sich Elemente aller oben dargestellten Regeltypen finden. Da die Regeltypen in der praktischen Anwendung ohnehin meist in Kombination auftreten, entspricht dies aber durchaus der tatsächlichen praktischen Implementierung.

7.4.1 Rekonstruktion „stummer Phoneme“

Im bereits an früherer Stelle angeführten Beispiel des französischen Ortsnamens *Questembert* wurde das in der orthographischen Form wortfinal auftretende stumme <t> einem „Nullphonem“ (auch als „Leerstelle“ bezeichnet) zugeordnet. Damit ist gemeint, daß einem Element der orthographischen Ebene kein Element der phonetischen Ebene entspricht. Obgleich solche Elemente lautlich nicht realisiert werden, werden sie bei der automatischen Zuordnung von Graphem- und Phonem-Ebene nicht vollständig von der P-Ebene getilgt, sondern bleiben in Form eines Platzhalter-Symbols ‘-‘ erhalten.

Dieses spezifische Merkmal kann zur Modellierung solcher Aussprachefehler genutzt werden, bei denen ein graphemisch repräsentiertes, aber „stummes“ Phonem vom Sprecher artikuliert wird. Dieser Fehlertyp tritt beispielsweise häufig bei der Zielsprache Französisch auf, insbesondere bei Sprechern mit geringen Kenntnissen des Französischen. Hier finden sich Formen wie die folgenden:

Orthographisch	kanonische Aussprache	Akzent-Variante
<i>Fécamp</i>	[fekã]	[fekãp], [fekomp]
<i>Fougères</i>	[fugeɾ]	[fogeəs]
<i>Issoire</i>	[iswaɾ]	[iswaɾe]
<i>j'en avais ...</i>	[zẽnavɛ]	[zɛnavɛɾs]
<i>Langeais</i>	[lãʒɛ]	[lãgeɾs]
<i>Nevers</i>	[nøvɛɾ]	[nevɛɾs]
<i>Questembert</i>	[kɛstãbɛɾ]	[kɛstãbɛɾt]

Gemeinsam ist diesen Varianten, daß der wortfinale Konsonant oder Vokal, der in der kanonischen Aussprache stumm bleibt, von den Sprechern artikuliert wird. Dabei wird jedoch nicht ein beliebiger Laut eingeführt, sondern derjenige Laut, welcher durch das orthographische Symbol an dieser Position repräsentiert ist. Eine allein auf der phonetischen Ebene operierende Ersetzungsregel wäre nicht imstande, diesen Fehler zu modellieren, da erstens an der betreffenden Position *kein Laut vorhanden ist*, der zu ersetzen wäre (fehlendes Eingabesymbol), und zweitens der jeweils *eintretende Ersatzlaut* nur unter Bezugnahme auf die graphemische Repräsentation bestimmt werden kann.

Statt dessen ist hier eine zweistufige Regel mit Zugriff auf die Graphem-Ebene erforderlich. Als Eingabesymbol für diese Regel dient kein Phonemsymbol im üblichen Sinne, sondern die durch das Graphem-Phonem-Alignment eingeführte Leerstelle ‘-’. Ist eine solche Leerstelle vorhanden, so wird zunächst ermittelt, welches Graphemsymbol dieser Position zugeordnet ist. In einem zweiten Schritt wird derjenige Laut eingeführt, welcher durch dieses Graphemsymbol repräsentiert wird. Eine Regel, die für das oben genannte Beispiel *Questembert* [kɛstãbɛɾ] die Aussprachevariante [kɛstãbɛɾt] mit auslautendem [t] ableitet, läßt sich demzufolge schreiben als:

P-EBENE	[-]	→	[t]	/	__	##
G-EBENE	<t>					

Wie das Beispiel zeigt, wird hier die Leerstelle ‘-’ als Eingabesymbol der Regel verwendet und durch den Laut [t] ersetzt. Eine obligatorische Bedingung hierfür ist dabei jedoch, daß die Leerstelle einem Graphem <t> auf der Graphem-Ebene zugeordnet ist. Auf diese Weise kann der in der phonetischen Eingabesequenz zunächst nicht vorhandene Laut unter Bezugnahme auf die graphemische Struktur rekonstruiert werden.

7.4.2 Rekonstruktion von Reduktionsvokalen

Ein dem zuletzt dargestellten Fall verwandtes Modellierungsproblem tritt dort auf, wo die phonetische Oberflächenform Spuren einer Anwendung sprachspezifischer phonologischer Regeln aufweist, jedoch die zugrundeliegende Form (UR), aus der die Oberflächenform abgeleitet ist, nicht mehr offen zutage tritt. So verfügt beispielsweise das Englische über eine phonologische Regel der „Vokalreduktion“, durch die einige Vollvokale in bestimmten Positionen zu Schwa [ə] reduziert werden²². Muttersprachler des Deutschen und insbesondere des Französischen neigen dazu, diese phonologische Regel nicht anzuwenden und anstelle des [ə] einen Vollvokal zu produzieren. Die folgenden Beispiele illustrieren diesen häufig auftretenden Fehler:

Name	kanonische Aussprache L2	Akzent-Variante
<i>Abington</i>	[æbɪŋtən]	[a:bɪŋtən], [eɪbɪŋtən]
<i>London</i>	[lʌndən]	[lɒndən]
<i>Crooklands</i>	[kɹʊklənds]	[kɹʊklənds]
<i>Stanford</i>	[stænfəd]	[stænfɔ̃t]
<i>Aston</i>	[estən]	[æstən]

Da in der Eingabesequenz nur die phonetische Oberflächenform vorliegt, besteht das Modellierungsproblem darin, den zugrundeliegenden Vollvokal zu ermitteln, aus dem sich [ə] ableitet. Wiederum ist es nicht möglich, diese Erscheinung allein mittels Operationen auf der phonetischen Oberflächenform zu modellieren, da

- (1) sich der in Frage kommende Ersatzlaut nicht aus der Eingabesequenz (= phonetische Oberflächenform) ermitteln lässt und
- (2) *nicht jedes* Auftreten von Schwa [ə] durch die Regel erfaßt werden soll, sondern nur solche Fälle, in denen [ə] das Ergebnis eines Derivationsprozesses ist.

Aus diesen Gründen ist eine einfache Ersetzungsregel [ə] → [ɔ] hier nicht sinnvoll anwendbar. Eine einfache Lösung dieses Problems wäre möglich, wenn das Eingabelexikon neben der phonetischen Oberflächenform auch die zugrundeliegende phonologische Repräsentation (UR) enthielte. Wäre beispielsweise eng-

²² Aus Gründen der besseren Darstellbarkeit werden reduzierte Vokale im Englischen hier durchgängig als Schwa [ə] repräsentiert, auch wenn in einigen Kontexten eine Repräsentation reduzierter Silben als silbische Konsonanten [ŋ m ɹ ʃ] phonetisch adäquater wäre.

lich *Abington* mit einer UR /æbɪŋtɒn/ im Lexikon verzeichnet, so könnte eine Aussprachevariante [æbɪŋtən] hieraus leicht abgeleitet werden, indem die Anwendung der phonologischen Regel ‚Vokalreduktion‘ unterdrückt würde:

	UR	Regel Vokalreduktion	phonetisch
kanonische Form	/æbɪŋtɒn/	JA	[æbɪŋtən]
Akzentvariante	/æbɪŋtɒn/	NEIN	[æbɪŋtɒn]

Im vorliegenden Anwendungszusammenhang muß jedoch von dem Fall ausgegangen werden, daß Informationen über die zugrundeliegende Form nicht vorliegen. Es ist aber möglich, die fehlende Information aus der orthographischen Repräsentation abzuleiten und die entsprechende Aussprachevariante mittels einer graphemabhängigen Ersetzungsregel zu formulieren. Die folgende Regel generiert die gewünschte Zielform:

P-EBENE	[ə]	→	[ɔ]
G-EBENE	<o>		

Die Regel besagt, daß ein reduziertes Schwa in einen Vollvokal [ɔ] umzuwandeln ist, sofern es auf der Ebene G durch <o> repräsentiert ist. Mit dieser Bindung an die graphemische Repräsentation kann der betreffende Aussprachefehler zutreffend modelliert werden; die Regel generiert z.B. *Aston* → [æstɒn], *London* → [lɒndɒŋ] und viele weitere vergleichbare Varianten. Zu beachten ist dabei jedoch, daß analog zur genannten Regel weitere Regeln für *alle* Vollvokale zu formulieren sind, die einer Vokalreduktion unterworfen sein können. Eine Zusammenfassung zu einer einzigen allgemeinen Regel „Aufhebung der Vokalreduktion“ ist nicht möglich, da der Ersatzvokal stets von der spezifischen Repräsentation auf der Ebene G abhängt.

7.4.3 Graphemische Kontexte

Die Schriftaussprache-Regeln (Gesamtheit aller Graphem-Phonem-Beziehungen) einer Sprache sind nicht allein durch eine sprachenspezifische Menge unmittelbarer Assoziationen zwischen einzelnen Graphemzeichen und den damit repräsentierten Lauten verbunden, sondern insbesondere auch durch ihre *Kontextwerte*: Während sich einige Beziehungen zwischen Laut und Buchstabe in Form von

kontextunabhängigen Entsprechungen formulieren lassen (für das Deutsche z.B. die Regel „*das Graphem <x> wird [ks] ausgesprochen*“), muß ein großer Teil solcher Ausspracheregeln auf die umgebenden Grapheme, d.h. auf die Position des betreffenden Graphems innerhalb des Wortes Bezug nehmen.

Beispielsweise gilt im Deutschen, daß ein Vokal als Kurzvokal realisiert wird, wenn in der orthographischen Repräsentation zwei oder mehr Konsonanten folgen. Diese Regel gilt nicht nur bei orthographischen Doppelkonsonanten (Geminaten), z.B. in *Ha[mm]er*, *Schi[mm]el*, *Fe[ss]el*, sondern grundsätzlich bei allen nachfolgenden Konsonantengruppen, auch wenn es sich um verschiedene Konsonanten handelt (vgl. z.B. die Alternation der Vokallänge im Minimalpaar *Mode* - *Mo~~r~~de*). Die phonetische Vokallänge ist in diesen Fällen also nicht unmittelbar in einem einzelnen Graphemsegment repräsentiert, sondern kann erst durch die Kenntnis der nachfolgenden Grapheme eindeutig bestimmt werden²³.

Auch in akzentbehäfteter Aussprache treten phonetische Varianten auf, bei denen die beobachteten Lautersetzungen durch die *orthographische Umgebung* ausgelöst werden. So konnte beispielsweise bei Sprechern des Französischen mit der Zielsprache Deutsch eine auffallende Tendenz (relative Häufigkeit ca. 30%) beobachtet werden, Vokale vor orthographischen Doppelkonsonanten als *Langvokale* zu realisieren. Für deutsche Ortsnamen wie *A[mm]ensleben* oder *Niedermö[ll]ern* treten infolgedessen häufig Formen wie [a:mənsle:bən] oder [ni:dəmə:lən] auf. Die weiter oben genannte Regel der deutschen Schriftausprache, der zufolge orthographische Doppelkonsonanten stets einen vorhergehenden *Kurzvokal* markieren, wird somit von den betreffenden Sprechern nicht nur nicht befolgt, sondern vollständig invertiert. Dies läßt sich wie folgt darstellen:

(Korrekte) Regel des Deutschen: Vokalkürzung vor <CC>

[V] / ___ <CC> → V [-lang]

Von französischen Sprechern angewendete Regel:

[V] / ___ <CC> → V [+lang]

Mögliche phonetische Kontexte, die als Auslöser dieses Fehlers wirksam werden, konnten in den vorliegenden Sprachdaten nicht identifiziert werden. Für eine orthographische Motivation dieses Fehlers spricht zudem der Sachverhalt, daß dieser Fehler ausschließlich bei gelesenem Material auftrat und bei akustischen Stimuli nicht beobachtet werden konnte.

²³ Die im Deutschen zusätzlich auftretende Interaktion zwischen Vokallänge und Wortbetonung soll hier zur Vereinfachung unberücksichtigt bleiben.

Um solche Aussprachefehler modellieren zu können, kann wiederum auf den Regeltyp der graphemabhängigen Lautsubstitution zurückgegriffen werden. Seine Anwendung erfolgt hier jedoch in einer geringfügig anderen Weise als bisher dargelegt: Während in den bisherigen Beispielen für diesen Regeltyp die *unmittelbare Assoziation zwischen dem zu ersetzenden Laut und seiner graphemischen Repräsentation* als Anwendungsbedingung für die Regel herangezogen wurde, wird hier nun der *graphemische Kontext* relevant – d.h. die unmittelbare Umgebung auf der G-Ebene. Um beispielsweise den genannten Aussprachefehler französischer Sprecher – eine Vokallängung vor orthographischen Doppelkonsonanten – zu modellieren, kann folgende Regel formuliert werden:

P-EBENE	[a]	→	[a:]
G-EBENE	< >	/	— <CC>

Hier wird erstmals ein *graphemischer Kontext* definiert, welcher erfüllt sein muß, damit die Regel anwendbar ist. Die Regel besagt, daß ein Vokal [a] in einen Langvokal [a:] umgewandelt wird, sofern auf der Ebene G ein Doppelkonsonant <CC> als rechter Kontext vorliegt. Da die graphemische Repräsentation des zu ersetzenden Lautes [a] für diese Regel irrelevant ist, kann sie unspezifiziert bleiben. Dies ist hier durch die leeren Klammern < > angezeigt. Die Regel generiert die gewünschte Zielform in allen Kontexten, bei denen auf der orthographischen Ebene ein Doppelkonsonant folgt.

7.4.4 Sprachspezifische Grapheme

In Kapitel 6.4.2 wurden Abweichungen der Graphem-Inventare von L1 und L2 als eine mögliche Fehlerquelle der L2-Aussprache dargestellt: Verfügt die Zielsprache über sprachspezifische Graphemzeichen oder Diakritika, welche dem Sprecher nicht bekannt sind, treten hierdurch häufig Aussprachefehler beim Lesen von L2-Material auf, die auf einer Anwendung von „Ersatz-Graphemzeichen“ beruhen (graphemische Reinterpretation).

Eine häufig verwendete Strategie der Sprecher besteht dabei darin, die betreffenden Symbole aus L2 durch die jeweils *ähnlichsten* Grapheme aus L1 zu ersetzen und sie entsprechend ihres Lautwertes in L1 zu artikulieren (siehe dazu ausführlich Kap. 6.4). So werden die deutschen „Umlaute“ <ä, ü, ö> von Sprechern, deren Muttersprache nicht über diese Grapheme verfügt, durch ihre unmarkierten Basisgrapheme <a, o, u> ersetzt und dementsprechend phonemisiert. Beispielsweise kann eine Realisierung von deutsch *Düsseldorf* als [dus|dɔ:f] oder [dʌs|dɔ:f] durch englische Sprecher auf einen solchen „Graphem-Ersatz“ zurück-

geführt werden: Das L2-Graphem <ü> wird vom Sprecher durch das aus seiner L1 bekannte <u> ersetzt und dementsprechend phonemisiert. Wie bereits in den vorherigen Beispielen kann auch dieser Fehler nicht durch eine Regel modelliert werden, die allein auf der phonetischen Ebene operiert. So würde eine Lautersatzungs-Regel

$$[Y] \rightarrow [u]$$

jedes Vorkommen von [Y] erfassen. Damit würden jedoch nicht nur erwünschte Fälle wie *Düsseldorf*, sondern z.B. auch Wörter wie *Chlorophyll* oder Ortsnamen wie *Stadtkyll* erfaßt und mit den Aussprachevarianten [kloʁofʊl] oder [ʃtatkul] phonemisiert. Eine solche Übergenerierung ließe sich auch dadurch nicht verhindern, daß die Regel mittels Kontextbedingungen auf einen *phonetischen Kontext* eingeschränkt würde, da die Ursache des Fehlers auf der Ebene der Orthographie angesiedelt ist.

Derartige Fehler können jedoch gezielt modelliert werden, indem bestimmte Lautsubstitutionen an die relevanten sprachspezifischen Grapheme gebunden werden und nur dann zur Anwendung kommen, wenn ein solches Graphem vorliegt. Die folgende Regel erfaßt beispielsweise den oben genannten Fall von *Düsseldorf*, bei dem <ü> als [u] wiedergegeben wird:

P-EBENE	[Y]	→	[u]
G-EBENE	<ü>		

Diese Regel generiert zwar die gewünschte Form [dʊs|dʊ:f], bleibt jedoch in dem Falle unwirksam, daß [Y] durch ein anderes Graphem als <ü> repräsentiert ist (z.B. durch <y>). Somit weist die Regel exakt die Eigenschaften auf, die zur Modellierung des hier zur Diskussion stehenden Merkmals eines „englischen Akzents“ gefordert sind.

Regeln dieses Typs wurden für mehrere untersuchte Zielsprachen formuliert. Sie modellieren die in den Sprachdaten beobachteten Varianten, die durch eine Interferenz der Grapheminventare verursacht werden, in der gewünschten Weise. Anwendbar ist dieser Regeltyp auf eine Reihe weiterer hier nicht untersuchter Zielsprachen, in denen Sonderzeichen und sprachspezifische Diakritika auftreten. Dies ist insbesondere dann sinnvoll, wenn die Diakritika deutliche Phonemspürungen markieren, z.B. <ø, å> gegenüber <o, a> in skandinavischen Sprachen. Zu beachten ist bei der Formulierung derartiger Regeln jedoch stets, daß der gewählte Ersatzlaut in hohem Maße von der L1 des Sprechers abhängt: Dabei tritt zumeist derjenige Laut als Ersatz ein, der nach den Schriftaussprache-Regeln

der L1 des Sprechers durch das „Ersatzgraphem“ repräsentiert wird (siehe Kap. 6.4.2).

7.5 Regeltypen: Vorläufiges Fazit

Viele der in Kapitel 6 (und an anderen Stellen dieser Arbeit) dargestellten Fehlertypen sind zum gegenwärtigen Zeitpunkt im erstellten Regelsystem berücksichtigt und programmtechnisch implementiert.

Zur besseren Übersicht sind in der folgenden Tabelle 7.3 (folgende Seite) noch einmal sämtliche Fehlertypen zusammengetragen, die bisher erörtert wurden. Zugleich ist dabei gekennzeichnet, welche dieser Fehlertypen in der bisherigen Implementierung des Regelsystems berücksichtigt sind. Fehlertypen, deren Modellierung bisher nicht implementiert wurde, sind für mögliche künftige Erweiterungen des Regelsystems vorgesehen. Teilweise mußten diese Fehlertypen in der bisherigen programmtechnischen Umsetzung ausgespart werden, weil notwendige Datenbestände nicht im erforderlichen Umfang vorlagen. Als Beispiel sei der in Kapitel 6.5.1 dargelegte Umstand genannt, daß lexikalische Morpheme der Zielsprache von Aussprache Fehlern oft weniger betroffen sind als onomastische oder unikale Morpheme und aus diesem Grunde von einigen Lautersetzungen ausgenommen werden sollten. Eine technische Modellierung des sprachlichen Wissens, welches diesem Phänomen zugrunde liegt, könnte nur unter Einbeziehung entsprechend präparierter und vorsegmentierter Morphemlexika erfolgen. Diese müßten dabei erstens für jede untersuchte Sprache vorliegen und sollten zweitens einen erheblichen Umfang aufweisen²⁴, um sinnvoll eingesetzt werden zu können. Ähnliches gilt für die Einbeziehung von Informationen über die Silbenstruktur bei der Modellierung von Fehlern (siehe Kapitel 6.2.6). Voraussetzung hierfür ist die Verfügbarkeit umfangreicher Lexika mit einer bereits im Lexikon codierten syllabischen Segmentierung sämtlicher Einträge oder – als Alternative dazu – eines Programms zur automatischen Syllabifizierung, die für mehrere Sprachen gleichermaßen zuverlässig arbeitet.

²⁴ Was als ein Lexikon mit „erheblichem Umfang“ gelten kann, hängt letztlich vom Anwendungskontext ab. Für den hier diskutierten Zusammenhang kann aber ein Umfang von mindestens 5000 Lexikoneinträgen – pro untersuchter Sprache – als ein realistischer Richtwert angenommen werden. Lexika von geringerem Umfang wären hier vermutlich kaum nutzbringend einsetzbar.

Fehlertyp	Untertypen	siehe Kapitel	Implementiert
(1) Segmentaler Lautersatz <ul style="list-style-type: none"> • Substitutionen • Elisionen • Einfügungen 	a) <i>ohne Kontextbeschränkungen</i> b) <i>mit Kontextbeschränkungen</i>	6.2	<i>ja</i>
(2) Interferenzen durch phonologische Regeln	a) <i>Transfer von L1-Regeln auf L2</i> b) <i>Nichtanwendung von L2-Regeln</i>	6.3	<i>ja</i>
(3) Transfer von Graphem-Phonem-Beziehungen	a) <i>globaler Transfer</i> b) <i>lokaler Transfer</i>	6.4.1	<i>ja</i>
(4) Interferenz durch abweichende Graphem-Inventare	a) <i>Graphem-Ersatz, Nichtbeachtung von Diakritika</i> b) <i>abweichende phonetische Funktion von Diakritika</i>	6.4.2	<i>ja</i>
(5) Transfer der L1-Silbenstruktur	-	6.2.6 6.2.7	<i>teilweise</i>
(6) Drittsprachen-Interferenz	-	6.6	<i>nein</i>
(7) Übergeneralisierungen	-	6.6	<i>teilweise</i>
(8) Einflüsse der Morphemstruktur	-	6.5	<i>nein</i>
(9) Idiosynkratische Fehler	<i>Versprecher, Abbrüche, Lesefehler u.v.m.</i>	-	<i>nein</i>

Tabelle 7.3: Zusammenfassung der dargestellten Fehlertypen und bisherige programmtechnische Implementierung

Wenn auch derartige Ressourcen bei der Durchführung dieser Arbeit nicht zur Verfügung standen, so bedeutet dies nicht, daß die entsprechenden Fehlertypen nicht prinzipiell einer regelbasierten Modellierung zugänglich sind. Vielmehr können dem Regelsystem jederzeit weitere Module zur Modellierung dieser Fehler hinzugefügt werden.

Hierbei ist jedoch grundsätzlich von der Möglichkeit auszugehen, daß die Kombination mit den bereits existierenden Modulen nicht die Gestalt einer rein *additiven* Verbindung annimmt, sondern daß vielfältige Interaktionen zwischen den Modulen zu berücksichtigen sind. So ist beispielsweise denkbar, daß Fehler, deren Ursache auf der Ebene der Silbenstruktur zu suchen ist, auf verschiedene Weise auf die phonetisch-segmentale Ebene zurückwirken und dadurch einzelne segmentale Fehler (Lautsubstitutionen) auslösen oder verstärken, aber auch aufheben oder anderweitig modifizieren können. In künftigen Erweiterungen des Regelsystems werden solche Interaktionen um so mehr in den Mittelpunkt des Interesses rücken, je mehr zusätzliche linguistische Ebenen in die Modellierung von Aussprache Fehlern einbezogen werden.

8 Evaluierung der Regeln

8.1 Kriterien einer Evaluierung

8.1.1 Systembezogene oder linguistische Evaluierung?

Die vorliegende Untersuchung ist einerseits motiviert durch Fragestellungen, die sich im Zusammenhang mit sprachtechnologischen Anwendungen und ihrer technischen Implementierung ergeben. Andererseits soll mit der vorgestellten Methode der regelbasierten Modellierung von akzentbehafteten Aussprachevarianten aber nicht ausschließlich ein pragmatisch-technischer Lösungsansatz entwickelt werden, sondern es wird darüber hinaus auch die Anforderung gestellt, Sprecherverhalten adäquat zu beschreiben. In diesem Spannungsfeld von linguistischer Deskription einerseits und praktischen sprachtechnologischen Anforderungen andererseits kann eine Evaluierung des entwickelten Ansatzes prinzipiell auf zwei Wegen erfolgen:

- (i) Eine Evaluierung der *linguistischen Adäquatheit* mit dem Ziel, die Tragfähigkeit des Gesamtkonzepts und die Gültigkeit der entwickelten Regeln in Bezug auf das zu modellierende sprachliche Verhalten von Sprechern zu prüfen.
- (ii) Eine *systembezogene* Evaluierung mit dem Ziel, die entwickelten Regeln in einem Spracherkennungs- oder Sprachsynthesystem zu implementieren und ihre Auswirkungen auf die Performanz eines bestimmten Systems zu untersuchen.

Trotz der sprachtechnologischen und damit anwendungsbezogenen Orientierung der vorliegenden Studie ist eine systembezogene Evaluierung mit Nachteilen verbunden. So ist es z.B. bei einer Implementierung des Regelsystems in einem spezifischen Spracherkennungssystem erforderlich, die verwendeten phonetischen Kategorien zunächst den erkennerspezifischen Kategorien anzupassen. Typischerweise sind die in heutigen, auf sog. *Hidden-Markov-Modellen* (HMM) basierenden Spracherkennern verwendeten Phoneminventare weniger umfangreich als die in dieser Arbeit verwendeten, verhältnismäßig detailreichen phonetischen Transkriptionen. Dies ist dadurch begründet, daß ein großer Teil potentieller Aussprachevariation bereits mittels statistischer Methoden auf der Ebene der akustischen Phonem-Modelle erfaßt und modelliert werden kann.

Vorbedingung für einen Performanztest mit einem solchen System wäre es daher, die mittels der Regelsätze automatisch generierten Transkriptionen zunächst auf das erkenner-spezifische Inventar zu reduzieren. Damit würden jedoch phonetische Detailinformationen, die bewußt in das Regelsystem eingeführt wurden, zum Zweck der Evaluierung nachträglich wieder eliminiert.

Performanztests, bei denen die Evaluierung methodisch und praktisch an ein spezifisches sprachtechnologisches System gekoppelt ist, laufen zudem stets Gefahr, die jeweils systemspezifische Eignung eines Ansatzes zu evaluieren, nicht jedoch ihre generelle linguistische Plausibilität und Adäquatheit – und damit ihre prinzipiell systemübergreifende Eignung. In sprachtechnologischen Anwendungen korrespondiert eine hohe Systemperformanz nicht immer mit dem Kriterium linguistischer Adäquatheit: So konnte beispielsweise gezeigt werden, daß beim Einsatz eines automatischen Spracherkennungs-Systems zur automatischen phonetischen Transkription eine hohe Erkennungsrate des Systems keineswegs zwingend eine hohe Transkriptionsqualität nach sich zieht (cf. Kessens & Strik 2001). Zwischen dem aus technischer Sicht maßgeblichen Kriterium der Systemperformanz (Erkennungsrate) und der linguistischen Adäquatheit der Systemausgabe (z.B. einer automatischen phonetischen Transkription) besteht demzufolge nicht immer eine unmittelbare Korrelation. Bezogen auf die vorliegende Arbeit bedeutet dies, daß die Ergebnisse einer systembezogenen Evaluierung voraussichtlich in hohem Maße systemspezifisch sein werden und nur begrenzte Aussagekraft bezüglich der Qualität der symbolisch-phonetischen Modellierung haben.

Aus diesen Gründen wurde eine systembezogene Evaluierung – d.h. eine Integration des Regelsystems in ein bestehendes sprachtechnologisches System mit anschließender Untersuchung der Systemperformanz – als nicht optimal erachtet. Statt dessen soll bei der Evaluierung das Kriterium der linguistischen Adäquatheit der Regelsätze im Vordergrund stehen. Kernfragen sind hierbei, ob und in welchem Maße

- (a) die Regeln sprecherübergreifende Aussprachecharakteristika, die nicht aus der bisherigen Sprachdatensammlung abgeleitet sind, realistisch modellieren können, d.h. auch für neue Sprecher Gültigkeit besitzen;
- (b) die Regeln trotz des begrenzten Vokabulars, anhand dessen sie gewonnen wurden, auch vokabularübergreifend (d.h. für neues Vokabular) plausible Aussprachevarianten generieren.

Hierzu wurde eine Methode entwickelt, die es erlaubt, den Grad der Approximation der automatisch generierten Varianten an tatsächlich auftretende Aussprachevarianten zu ermitteln und zu quantifizieren. Dieser Approximationsgrad liefert – unabhängig von einer spezifischen sprachtechnologischen Anwendung – ein Maß für die linguistisch-phonetische Plausibilität des Gesamtansatzes (Entsprechung zwischen empirischen Sprachdaten und regelbasierter Modellierung).

8.1.2 „Fehler“ bei einer Modellierung von Fehlern

Bei einer Modellierung von *Aussprachefehlern*, wie sie in der vorliegenden Arbeit angestrebt wird, ist eine Evaluierung der Ergebnisse stets mit dem bereits in früheren Kapiteln angesprochenen Problem verbunden, daß zunächst keine Referenzform existiert, die als verbindlicher Maßstab herangezogen werden kann. Während beispielsweise die Performanz eines automatischen Graphem-Phonem-Umsetzers daran gemessen werden kann, wie exakt (d.h. mit welcher Wortfehlerrate oder Phonemfehlerrate) die korrekte kanonische Transkription eines Vokabulars generiert werden kann (cf. Pols 1991; Damper et al. 1999; Belhoula et al. 1997), existiert ein solches Evaluierungskriterium in der vorliegenden Studie zunächst nicht. Eine klare Trennung von „korrekten“ und „fehlerhaften“ Systemausgaben erscheint hier problematisch, da die mittels der Regeln zu generierende Zielform eher einem lautlichen Prototyp mit unscharfen Grenzen entspricht als einer eindeutig definierten Lautsequenz. Dies erschwert die Bestimmung eindeutiger Kriterien, wann ein „Fehler“ des Regelsystems vorliegt.

Dennoch erscheint es durchaus möglich, die automatisch generierten Aussprachevarianten einer Evaluierung zu unterziehen. An die Stelle einer Unterscheidung von (a) korrekten und (b) fehlerhaften Systemausgaben könnte bei einer Modellierung von Aussprachefehlern, wie sie hier angestrebt wird, das Kriterium der Plausibilität treten. Generell läßt sich sagen, daß eine automatisch generierte Variante dann plausibel ist, wenn diese in der Sprechergruppe, deren Aussprache modelliert werden soll, tatsächlich auftreten könnte. Dies bedeutet nicht, daß sie zwangsläufig auftreten *wird*: Der Regelsatz hat keinen exakt prognostischen Charakter, sondern kann nur *potentiell* auftretende Aussprachen möglichst gut repräsentieren. Plausibilität in diesem Sinne ist aber ein äußerst dehnbares, „unscharfes“ Kriterium, denn die Entscheidung darüber, ob eine automatisch generierte Aussprachevariante auftreten *könnte*, ist letztlich nicht eindeutig zu treffen. Die Tatsache allein, daß sie in einem begrenzten Korpus von Sprachdaten bisher nicht aufgetreten ist, läßt keinerlei Schlüsse zu, ob sie nicht künftig in einem anderen oder erweiterten Korpus auftreten wird. Gerade im Falle akzentbehafteter Aussprache treten mitunter phonetische Varianten auf, die jeglicher sprachwissenschaftlicher Prognose entgegenlaufen; es kann aus dieser Sicht also beinahe jede Variante als „potentiell möglich“ gelten.

Das Kriterium der Plausibilität ist daher enger zu fassen. Vorgeschlagen wird daher, daß ein Regelsatz dieses Kriterium dann erfüllt, wenn

- (1) die durch ihn generierten Aussprachevarianten innerhalb einer untersuchten Sprechergruppe mit hoher Frequenz auftreten;
- (2) Varianten, die nur selten – z.B. bei nur einer geringen Anzahl von Sprechern – auftreten, von den Regeln nicht oder nur als zusätzliche, fakultative Aussprachevarianten generiert werden;

(3) Varianten, die phonetisch vollständig „abwegig“ sind, *niemals* generiert werden. Eine solche Variante würde beispielsweise vorliegen, wenn für das deutsche Wort *gestern* [gɛstɐn] eine Aussprachevariante [ga:stin] oder [kɛstɛnn] vorliegt. Derartige idiosynkratische Varianten treten zwar durchaus vereinzelt in den Sprachdaten auf, ihre gezielte Modellierung erscheint jedoch kaum aussichtsreich.

Die in diesem Kapitel dargelegte Evaluierungs-Strategie bezieht sich allein auf die ersten beiden dieser Kriterien. Dennoch ist auch Kriterium (3) durchaus relevant für eine weitere Optimierung des Regelsystems: Aufgrund von Fehlern in der Regelgestaltung bzw. der ausführenden Computerprogramme generieren regelbasierte Systeme mitunter Varianten, die vollständig unmotiviert bzw. linguistisch sinnlos erscheinen. Solche Fehler treten aber bei praktisch *jeder* regelbasierten Generierung von phonetischen Varianten (z.B. auch bei automatischen Graphem-nach-Phonem-Umsetzern) in gewissem Umfang auf – eine Teilaufgabe der Optimierung solcher Systeme liegt also immer darin, diesen Fehlertyp weitestmöglich zu eliminieren. Da diese Fehler jedoch meist programmtechnischer Natur sind, werden sie hier nicht weiter behandelt²⁵.

Im vorliegenden Zusammenhang ist dagegen vor allem der Sachverhalt von Interesse, daß die Plausibilität der Regeln letztlich nur dann erfolgreich geprüft werden kann, wenn als Maßstab nicht potentielle Varianten, sondern tatsächlich beobachtete Aussprachevarianten herangezogen werden. Das im folgenden dargestellte Evaluierungsverfahren gründet sich daher vorrangig auf dieses Kriterium: Die Entsprechung zwischen (1) *regelbasiert erzeugten Varianten* und (2) den *tatsächlich von Sprechern produzierten Varianten* wurde anhand eines phonetischen Distanzmaßes quantifiziert und an verschiedenen Sprechergruppen geprüft. Die so ermittelten Ergebnisse liefern nicht nur ein wichtiges Maß für die generelle Tragfähigkeit des entwickelten Ansatzes, sondern auch ein wertvolles Instrument zur Weiterentwicklung der Regelsätze.

²⁵ Die bei Schaden (2004) dargestellten skizzierten phonetischen Lexika, die mittels der in dieser Arbeit entwickelten Regelsätze erzeugt wurden, bieten sich für eine solche Form der Evaluierung an. Bereits durch eine sorgfältige Sichtung der hier generierten Varianten und eine Kennzeichnung von „abwegigen“ phonetischen Formen ist es leicht möglich, fehlerhafte Regeln zu identifizieren und auf diesem Wege die Regelsätze zu optimieren. Es handelt sich hierbei jedoch stets nur um eine *ergänzende* Evaluierungsstrategie.

8.2 Beschreibung des Evaluierungsverfahrens

8.2.1 Ermittlung der besten Approximation

Die in Kap. 4.4 beschriebene Akzentstufung und die darauf aufbauende Generierung mehrerer alternativer Aussprachevarianten pro Worteinheit dient einer umfassenden Abdeckung potentieller Varianten und damit einer Modellierung interindividueller Variabilität innerhalb der untersuchten Domäne. Ziel ist es dabei jedoch nicht, *jede* potentielle Aussprachevariante exakt zu modellieren. Ein solches Ziel erscheint prinzipiell nicht realistisch, da es voraussetzen würde, daß die in einer Sprechergruppe produzierten Varianten eindeutig prognostizierbar wären. Statt einer exakten Reproduktion und Prognose von Aussprachevarianten wird daher angestrebt, häufig auftretende und für eine Sprechergruppe charakteristische Varianten bestmöglich zu *approximieren*, d.h. eine möglichst optimale Annäherung zu erzielen.

Eine zentrale Anforderung ist dabei, daß die regelbasiert generierten Formen die in dieser Sprechergruppe auftretenden Varianten phonetisch adäquater beschreiben als eine unmodifizierte Standard-Transkription. Dies ist grundsätzlich dann der Fall, wenn sie den Aussprachevarianten der Sprecher *ähnlicher* sind als die kanonische Transkription. Um genau dies ermitteln zu können, wurde eine Methode erarbeitet und programmtechnisch implementiert, mit deren Hilfe phonetische Ähnlichkeit auf der Ebene der phonetischen Transkription quantifiziert werden kann. Kern dieser Methode ist ein Algorithmus bzw. ein Modul zur automatisierten Berechnung der *Distanz* bzw. *Ähnlichkeit* zwischen zwei Phonemsequenzen. Mittels einer solchen Distanzberechnung auf der Ebene der phonetischen Transkription kann ermittelt werden, ob und unter welchen Bedingungen die automatisch generierten Aussprachevarianten eine im Vergleich zur kanonischen Aussprache höhere Approximation an die tatsächlich in einer Sprechergruppe auftretenden Varianten erzielen können. Abb. 8.1 illustriert dieses Verfahren:

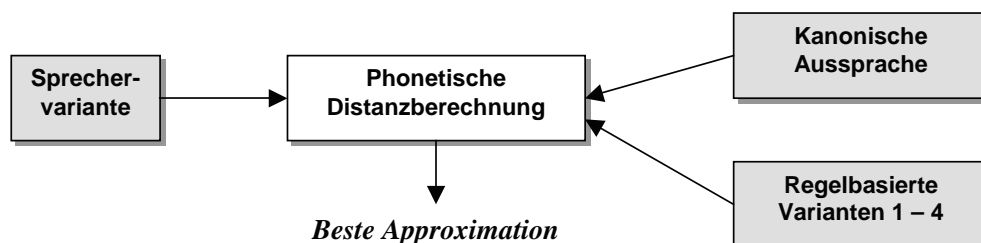


Abb. 8.1: Evaluierung mittels phonetischer Distanzberechnung

Entsprechend dem Ansatz einer prototypischen Akzentstufung ist dabei keine exakte Übereinstimmung der Aussprachevarianten mit den automatisch generierten Varianten gefordert, sondern vielmehr eine möglichst *optimale Approximation* durch eine (oder mehrere) der derzeit vier prototypischen Akzentstufen. Diese Approximation gibt schließlich Auskunft darüber, ob das zu modellierende Sprecherverhalten durch die Regeln mit hinreichender Genauigkeit erfaßt und beschrieben wird.

8.2.2 Definition eines phonetischen Distanzmaßes: Edit-Distanz

Algorithmen zur phonetischen Distanzberechnung werden in der (computer-)linguistischen Forschung erst in jüngerer Vergangenheit eingesetzt und wurden z.B. im Bereich der Dialektologie, aber auch in sprachtechnologischen Kontexten erfolgreich angewandt. So stellt z.B. Nerbonne (2003) Arbeiten vor, in denen phonetische Distanzalgorithmen zur Quantifizierung von Distanzen zwischen niederländischen Dialekten eingesetzt wurden, und stellt die Ergebnisse in Beziehung zu traditionellen dialektologischen Arbeiten. Hierbei treten interessante Übereinstimmungen mit früheren Studien, aber auch abweichende neue Daten hinsichtlich der Distribution von Dialektarealen und dem Verlauf von Isoglossen im niederländischen Sprachraum auf. Arbeiten wie diese stellen einen theoretisch wie forschungspraktisch vielversprechenden Schnittpunkt zwischen traditioneller linguistischer Feldforschung und moderner Computerlinguistik dar.

Grundsätzlich sollte ein phonetisches Distanzmaß bei einem Vergleich zweier Phonemsequenzen zu einem Ergebnis wie dem folgenden gelangen:

Orthographisch	Referenzform	Variante	Phonetische Distanz ²⁶
<i>Bamberg</i>	[bambɛɣk]	[bambɛɣk]	niedrig
<i>Berlin</i>	[bɛɣli:n]	[bɛ:ɣli:n]	mittel
<i>Essen</i>	[ɛsən]	[ɛsɛ:n]	hoch

Leichte phonetische Abweichungen sollten sich demnach in einem nur geringen Distanzwert niederschlagen, während größere phonetische Unterschiede zwischen den Sequenzen in einem hohen Wert resultierten. Die Ermittlung eines solchen Distanzwertes setzt sich prinzipiell aus zwei Teilschritten zusammen, welche in der Praxis ineinandergreifen:

²⁶ Die Werte *niedrig*, *mittel*, *hoch* haben hier nur illustrativen Charakter; an ihre Stelle treten später numerische Werte.

- (1) **Zeichenfolgenvergleich** – Bestimmung eines Distanzwertes zwischen zwei Symbolsequenzen A und B (globaler Vergleich);
- (2) **Segmentaler Vergleich** – Bestimmung der Distanz zwischen einzelnen Segmenten A_i und B_i an definierten Positionen der Sequenzen A und B auf der Basis von Ähnlichkeiten zwischen A_i und B_i (lokaler Vergleich).

In der Literatur wurden verschiedene Varianten der phonetischen Distanzberechnung vorgestellt (eine Übersicht gibt z.B. Kondrak 2003), die jedoch im Wesentlichen auf ähnlichen Prinzipien des *Dynamic Programming* (DP) beruhen, welche mit den erforderlichen phonetisch-phonologischen Erweiterungen versehen wurden.

Auch die im folgenden vorgestellte Methode zur phonetischen Distanzberechnung beruht grundsätzlich auf diesen Prinzipien. Dabei wird eine etablierte Methode des globalen Vergleichs von Symbolsequenzen kombiniert mit einer Methode, mittels derer lokale Abstände (Distanzen) auf phonetischer Grundlage ermittelt werden können. Der in dieser Studie verwendete Algorithmus zur Ermittlung der phonetischen Distanz basiert auf der sog. *Edit-Distanz* (auch *Levenshtein-Distanz*; cf. Levenshtein 1966). Hierbei handelt es sich um ein Abstandsmaß zwischen im Prinzip beliebigen Symbolsequenzen, welches in verschiedenen wissenschaftlichen Disziplinen Anwendung gefunden hat (cf. Kruskal 1983) und seit relativ kurzer Zeit auch im Rahmen linguistischer Forschung eingesetzt wird.

Der Basis-Algorithmus der Edit-Distanz ermittelt den Abstand zwischen zwei Zeichenketten A und B , indem die Anzahl von *Substitutionen* (*substitutions* = S), *Einfügungen* (*insertions* = I) und *Löschungen* (*deletions* = D) ermittelt wird, welche minimal erforderlich sind, um A in B zu konvertieren. Die Edit-Distanz zwischen A und B ist demnach zu verstehen als diejenige Sequenz elementarer Operationen des Typs S , I oder D , welche mit dem geringstmöglichen Aufwand die Konvertierung von A nach B vollzieht. Das folgende Beispiel zeigt eine solche Sequenz von Elementaroperationen für die Zeichenketten *hunger* und *fingers*.

Sequenz	Operation	Wert
hunger		
funger	ersetze <h> durch <f>	2
finger	ersetze <u> durch <i>	2
fingers	füge <s> ein	1
fingers		
Edit-Distanz		$\Sigma = 5$

Tabelle 8.1: Konvertierung von *hunger* zu *fingers*,
Ermittlung der Edit-Distanz

Als Distanz zwischen den Zeichenketten wird somit der Gesamtwert 5 ermittelt. Zu beachten ist dabei die unterschiedliche Gewichtung der Basisoperationen: Während Einfügungen und Löschungen jeweils mit dem Wert 1 in die Berechnung eingehen, erhalten Substitutionen den Wert 2. Dies ist motiviert durch die Annahme, daß eine Substitution im Prinzip als sequentielle Kombination einer Einfügung und einer Löschung betrachtet werden kann ($S = D + I$). Entsprechend wird z.B. für Minimalpaare wie *loben* - *leben* eine Edit-Distanz von 2 ermittelt, da hier genau eine Substitution ($S = 2$) für die Konvertierung erforderlich ist.

Weiterhin ist zu beachten, daß die oben dargestellte Sequenz von Operationen nicht die einzig mögliche Abfolge von S , I und D ist, die den minimalen Distanzwert generiert. Es existieren stets alternative „Pfade“ zur Konvertierung von Sequenz A in Sequenz B , die ein identisches Ergebnis für die minimale Edit-Distanz liefern. Jedoch hat sich - trotz verschiedener Varianten des Algorithmus, die in der Literatur zu finden sind - ein Standard-Algorithmus zur Berechnung des „optimalen Pfads“ etabliert, der sich formal wie folgt darstellen läßt:

```

function MIN-EDIT-DIST (target, source) returns min-distance
n ← LENGTH (target)
m ← LENGTH (source)
Create a distance matrix distance[n+1,m+1]
distance[0,0] ← 0
for each column i from 0 to n do
    for each row j from 0 to m do
        distance[i, j] ← MIN (
            distance[i - 1, j] + ins-cost (targeti),
            distance[i - 1, j - 1] + subs-cost (sourcej, targeti),
            distance[i, j - 1] + del-cost (sourcej)
        )
    end
return min-distance = distance [n+1, m+1]

```

Abb. 8.2 : Algorithmus zur Berechnung der Edit-Distanz zwischen zwei Zeichensequenzen *target* und *source* (übernommen aus: Jurafsky & Martin 2000: 156)

Hierbei wird eine zweidimensionale Matrix *distance* [*n*, *m*] aus den zu vergleichenden Zeichensequenzen initialisiert, die dann schrittweise durchlaufen wird. Dabei wird jeweils das lokale Minimum aus den möglichen Operationen *substitution*, *insertion* oder *deletion* ermittelt, d.h. diejenige Operation, welche einen minimalen lokalen Distanzwert liefert. Aus der Akkumulation dieser lokalen Distanzen kann schließlich die Gesamtdistanz zwischen den Sequenzen abgelesen werden; sie entspricht dem letzten Wert der Matrix *distance* [*n*, *m*].

Für einen Vergleich phonetischer Varianten eignet sich die Edit-Distanz grundsätzlich auch deshalb, weil der Algorithmus es erlaubt, Abstände zwischen Symbolsequenzen unterschiedlicher Länge zu berechnen und diese Differenzen bei der Ermittlung eines Distanzwertes zu berücksichtigen. Da akzentbehaftete Aussprachevarianten häufig durch Einfügungen und Elisionen phonetischer Segmente gekennzeichnet sind, ist dies eine wesentliche Anforderung an das gesuchte Distanzmaß. So zeigt das folgende Beispiel Distanzwerte zwischen der kanonischen Aussprache von dt. *Hamburg* [hambʊɾk] und einigen fiktiven Aussprachevarianten, die z.T. Einfügungen und Elisionen enthalten:

MIN-EDIT-DIST	[hambʊʁk]	[ha:mbʊʁk]	[ha:mʊʁk]	[hambyʁg]	[ambyʁg]
[hambʊʁk]	0.0 (ident.)	2.0	3.0	4.0	5.0

Tabelle 8.2: Edit-Distanz MIN-EDIT-DIST für potentielle Aussprachevarianten von *Hamburg*

Entsprechend dem oben skizzierten Algorithmus, bei dem jede Einfügung oder Elision eines Segmentes mit einer Erhöhung der Distanz um den Wert 1 verbunden ist, erhöht sich auch in diesem Beispiel die Gesamtdistanz, wenn die betreffende Aussprachevariante eine geringere Anzahl von Lauten aufweist (wie z.B. [ambyʁg]). Längendifferenzen zwischen den zu vergleichenden Sequenzen werden also stets in die Distanzberechnung einbezogen.

8.2.3 Phonetische Gewichtung durch Merkmalsysteme

Trotz der genannten Vorteile weist der bisher beschriebene Algorithmus für den vorliegenden Zusammenhang ein entscheidendes Defizit auf. Dieses liegt darin begründet, daß zur Berechnung lokaler Distanzen innerhalb der Matrix *distance* $[n,m]$ für sämtliche Operationen (*S*, *D*, *I*) stets ein *konstanter* Wert verwendet wird. Im oben dargestellten Standard-Algorithmus MIN-EDIT-DIST wurde für Löschungen und Einfügungen der Wert 1 und für Substitutionen der Wert 2 angenommen. In der vorliegenden Arbeit wurde dagegen für jede der drei Operationen *S*, *D*, und *I* der Wert 1 als Grundlage der Berechnung verwendet²⁷.

Mit der Anwendung eines konstanten Wertes ist jedoch ein wesentliches Merkmal phonetischer Sequenzen nicht adäquat erfaßt. Das Hauptproblem besteht darin, daß die Edit-Distanz in ihrer unmodifizierten Form mögliche *Ähnlichkeiten* zwischen den ersetzten Segmenten nicht berücksichtigt. So wird für Substitutionen zweier Segmente a_i und b_i stets eine Konstante verwendet, die unabhängig von der Beschaffenheit von a_i und b_i und möglichen Ähnlichkeiten zwischen den Segmenten ist. Ein solches Vorgehen ist nur dann adäquat, wenn sämtliche Segmente, aus denen sich die Sequenzen zusammensetzen können, äquidistant sind, d.h. keinerlei Abstufungen bezüglich ihrer Ähnlichkeiten bestehen. Bei phonetischen Zeichensequenzen ist dies jedoch gerade nicht der Fall: Paare von Phonemen weisen stets einen spezifischen Grad artikulatorischer, akustischer oder auditiver Ähnlichkeit auf. So ist beispielsweise eine Substitution des Vokals [e:] durch [ɛ:] als weniger gravierend einzustufen als die Substitution desselben Vokals durch [i:] oder [a:]. Entsprechend ist die Gesamtdistanz zwi-

²⁷ Hierbei handelt es sich um eine gebräuchliche Variante des Algorithmus (cf. Jurafsky & Martin 2000: 154/55).

schen den phonetischen Sequenzen [vɛ:zəl] und [vi:zəl] geringer als z.B. zwischen [vɛ:zəl] und [vi:zəl]. Daher ist der genannte Algorithmus der Edit-Distanz um eine *phonetische Gewichtung* zu erweitern, welche die Ähnlichkeit der einzelnen Segmente in die Distanzberechnung einbezieht (*segmentales Ähnlichkeitsmaß*).

Als Grundlage der Ermittlung segmentaler Ähnlichkeit kann auf etablierte Beschreibungsverfahren der Phonologie zurückgegriffen werden. In phonologischen Analysen haben sich *phonologische Merkmale* als ein ökonomisches und sprachenübergreifend einsetzbares Beschreibungsinstrument bewährt. Jedes Phonem ist dabei definiert als ein Vektor von Merkmalen aus einer a priori definierten, endlichen Menge phonologischer Merkmale. So lassen sich beispielsweise die deutschen Phoneme /u:/ und /a/ als Bündel folgender sechs phonologischer Merkmale beschreiben²⁸:

/u:/ [+vokalisch], [+lang], [+rund], [+geschlossen], [+hoch], [-vorn]

/a/ [+vokalisch], [-lang], [-rund], [-geschlossen], [-hoch], [-vorn]

Zur Charakterisierung des betreffenden Phonems ist jedes dieser Merkmale mit einem Wert versehen (der sog. *Merkmalsausprägung* oder *Merkmalspezifizierung*). Im Beispiel sind die Merkmale – einem gebräuchlichen Darstellungsmodus folgend – positiv oder negativ spezifiziert; es handelt sich also um *binäre* Merkmale. In der Literatur wurden darüber hinaus auch skalare (graduell gestufte) phonetische Merkmalsysteme vorgeschlagen, bei denen einzelne Merkmale einen Wert auf einer mehrstufigen Skala annehmen können (z.B. Vieregge et al. 1984 sowie das System von Ladefoged 1975, welches von Somers 1998 auch zur phonetischen Distanzberechnung eingesetzt wurde). Eine vergleichende Übersicht verschiedener Merkmalsysteme findet sich bei Kondrak (2003) sowie in Heeringa (2004, Kap. 3).

Zwei Eigenschaften sind an diesem Analyseverfahren besonders hervorzuheben: Erstens wird hierbei nicht länger das Phonem als atomare lautliche Einheit angesehen, sondern eine interne Strukturierung auf subphonemischer Ebene angenommen, welche durch die Merkmale und deren Kombination definiert ist. Zweitens kann das Phonemsystem einer Sprache vollständig beschrieben werden als ein System phonologischer Oppositionen, welche in der betreffenden Sprache als bedeutungsunterscheidende lautliche Elementardifferenzen wirksam werden. Das folgende Beispiel verdeutlicht das Verfahren anhand der Klassifizierung einiger Vokalphoneme des Deutschen mittels der oben eingeführten sechs phonologischen Merkmale:

²⁸ Die Liste der hier gezeigten Merkmale ist *nicht* vollständig und dient nur der Illustration; ein vollständiges Merkmalsystem für das Phoneminventar des Deutschen findet sich z.B. bei Meinhold & Stock (1982: 141).

Phonem	/u:/	/o:/	/ε/	/ɪ/	/a/
Merkmalsprofil					
[± vokalisch]	+	+	+	+	+
[± lang]	+	+	-	-	-
[± rund]	+	+	-	-	-
[± geschlossen]	+	+	-	-	-
[± hoch]	+	-	-	+	-
[± vorn]	-	-	+	+	-

Tabelle 8.3: Darstellung der Vokalphoneme /u:/ /o:/ /ε/ /ɪ/ /a/ als binäre Merkmalsmatrix

Wie die Tabelle zeigt, kann jedes der Merkmale positiv [+] oder negativ [-] spezifiziert sein. Hieraus ergibt sich für jedes Phonem ein spezifisches *Merkmalsprofil*. Wären die genannten sechs Merkmale stets in einer festgelegten Reihenfolge angeordnet, so wäre dieses Profil allein mittels binärer Sequenzen darstellbar, durch die die einzelnen Phoneme eindeutig voneinander unterschieden sind: So ließe sich /u:/ mittels der Sequenz [+ + + + + -] und /ε/ als [+ - - - - +] schreiben.

Eine solche Merkmalspezifizierung kann nun im vorliegenden Zusammenhang zur Ermittlung der phonetisch-segmentalen Distanz bzw. Ähnlichkeit nutzbar gemacht werden. Der gesuchte segmentale Distanzwert $pDist(a, b)$ zwischen zwei Phonemen a und b entspricht dabei der *Summe der abweichenden Merkmalspezifizierungen* von a und b : Je größer dieser Wert ist, um so höher ist die phonetische Distanz zwischen den fraglichen Segmenten. Die folgende Tabelle illustriert dieses Verfahren anhand der Vokalphoneme /u:/ und /a/ auf der Basis der oben genannten sechs Merkmale:

Phonem	/u:/	/a/	segmentale Distanz $diff(a, b)$
[± vokalisch]	+	+	0
[± lang]	+	-	1
[± rund]	+	-	1
[± geschlossen]	+	-	1
[± hoch]	+	-	1
[± vorn]	-	-	0
			$\Sigma = 4$

Tabelle 8.4: Berechnung phonetisch-segmentaler Distanz am Beispiel /u:/ ~ /a/ auf der Basis 6 phonologischer Merkmale

Der Tabelle zufolge wird als segmentale Distanz zwischen /a/ und /u:/ der Wert 4 ermittelt. Dieser Wert entspricht der Anzahl der differierenden Merkmalswerte $diff(a, b)$. Je höher dieser Wert ist, desto größer ist die berechnete Distanz der fraglichen Phoneme.

Der in dieser Arbeit verwendete Algorithmus zur Ermittlung der phonetisch-segmentalen Distanz verwendet grundsätzlich das oben dargelegte Prinzip. Dabei werden derzeit 25 Merkmale als Deskriptoren für Phonemsymbole verwendet, die in der oben skizzierten Weise als binäre Merkmale definiert sind. Bei den Merkmalen handelt es sich zum größten Teil um artikulatorische Klassen, die auch in etablierten Merkmalsystemen zu finden sind. Dennoch folgt das Merkmalsystem keiner spezifischen Theorie, sondern ist im wesentlichen durch die praktische Zielsetzung einer segmentalen Distanzberechnung motiviert. Dies hat z.B. zur Folge, daß das Kriterium der Beschreibungsökonomie, welches phonologischen Merkmalsystemen üblicherweise zugrunde liegt, nicht im Vordergrund stand: Während phonologische Modelle in der Regel danach streben, auf der Basis möglichst weniger lautlicher Oppositionen (= Merkmale) eine vollständige Beschreibung zu etablieren, wurden hier verhältnismäßig zahlreiche Merkmalsklassen aufgestellt.

Dabei sind auch Kategorienbildungen möglich, die nicht ohne weiteres durch artikulatorische oder akustische Eigenschaften der betreffenden Laute begründet sind. So wurde beispielsweise ein Merkmal [± r_type] (= rhotiziert, *rhotic*) definiert, in dem verschiedene Realisierungen des /r/-Phonems zusammengefaßt sind. Obgleich für diese Laute – realisiert als [r], [r̥], [ʀ], [ʁ], [ʁ̥], [ɹ] u.a. – keinerlei gemeinsame akustische oder artikulatorische Basis benannt werden kann, erscheint eine Klassenbildung dennoch aus verschiedenen Gründen motiviert (cf. hierzu Wiese 2003; Lindau 1985). Darüber hinaus wurden die verwendeten

Merkmale mit einer hierarchischen *Gewichtung* versehen, mittels derer Hauptklassen wie [vokalisch], [obstruent] [sonorant] und Unterklassen wie [lang], [velar], [nasal], [aspiriert] gekennzeichnet werden. Hauptklassen erhalten dabei eine höhere Gewichtung als Unterklassen. Eine solche zusätzliche Skalierung ist sinnvoll, damit zu den Unterklassen gehörige Merkmale mit geringerer Stärke in die segmentale Distanzberechnung eingehen. So ist gewährleistet, daß geringfügige artikulatorische Differenzen einen niedrigeren Distanzwert erhalten. Eine fehlende Merkmalsgewichtung kann prinzipiell dazu führen, daß z.B. eine Differenz zweier Phoneme im Merkmal [\pm aspiriert] einer Differenz von [\pm vokalisch] äquivalent ist, was u.U. zu unplausiblen segmentalen Distanzen führt (siehe hierzu auch Kondrak 2003: 283).

Die derzeitig implementierte Merkmalsgewichtung weist jedem Merkmal einen Wert zu, mit dem es in die Berechnung der segmentalen Distanz eingeht. Eine mögliche künftige Erweiterung dieses Gewichtungsverfahrens liegt im Aufbau und der Implementierung einer vollständigen Merkmalshierarchie, wie sie z.B. bei Ladefoged (1999) entworfen wird. Hierbei werden artikulatorische Merkmale in Form von mehrfach verzweigenden Bäumen dargestellt, bei denen Ober- und Unterklassen auf mehreren Ebenen als hierarchisches System gegliedert sind.

Grundsätzlich ist das verwendete Merkmalsystem offen für Erweiterungen oder Reduktionen. So ist es möglich, verschiedene in der Literatur vorgeschlagene Merkmalsysteme (z.B. Jakobson & Halle 1956; Ladefoged 1975; Vieregge et al. 1984; Kondrak 2003) zu implementieren und die Ergebnisse von Distanzberechnungen auf der Basis dieser Systeme zu vergleichen. Die programmtechnische Umsetzung des Distanzmaßes sieht diese Möglichkeit vor. Grundsätzlich aber bleibt das hier zugrunde gelegte Verfahren der Berechnung der Distanz phonetischer Sequenzen auch dann erhalten, wenn unterschiedliche Merkmalsysteme verwendet werden.

8.3 Implementierung des Verfahrens

Die Methode der phonetischen Distanzberechnung setzt sich aus verschiedenen Teilschritten zusammen, die in der praktischen Realisierung und programmtechnischen Implementierung ineinander greifen. Diese Schritte sind im folgenden dargestellt. Hiermit ist zugleich eine zusammenfassende Darstellung des verwendeten Distanzmaßes gegeben.

Schritt (1): Ermittlung der phonetisch-segmentalen Distanz

Die phonetisch-segmentale Distanz zwischen zwei Phonemen $pDist(a, b)$ wird ermittelt, indem die Summe der differierenden Merkmalsausprägungen $diff(feature(a_i), feature(b_i))$ unter Berücksichtigung der jeweiligen Merkmalsgewichtung $weight(feature)$ berechnet wird. Dies ist in folgender Formel zusammengefaßt:

$$pDist(a, b) = \frac{1}{N} \sum_{i=0}^N diff(feature_i(a), feature_i(b)) \cdot weight(feature_i)$$

Der Wert N entspricht hierbei der Gesamtanzahl der verwendeten Merkmale, auf die die Summe der abweichenden Merkmale bezogen wird. So würde z.B. bei einer Summe von 8 Merkmalsdifferenzen bei insgesamt $N = 36$ Merkmalen der Distanzwert $8/36 = 0,22$ ermittelt, wohingegen bei einem reduzierten Merkmalsinventar von $N = 20$ sich bei gleicher Anzahl von 8 Merkmalsdifferenzen ein höherer Distanzwert von $8/20 = 0,4$ ergeben würde (Beispiele jeweils ohne Merkmalsgewichtung).

Im nächsten (jedoch nicht obligatorischen) Schritt wurde der Distanzwert in einen *Ähnlichkeitswert* transformiert, bei dem 1,0 die maximale Ähnlichkeit (= Identität) und 0,0 die maximale Distanz kennzeichnet. Dieser Berechnung zufolge ergeben sich beispielsweise für Paare der Vokalreihe /a:/ /a/ /u:/ /y:/ /e:/ /au/ /o:/ folgende Werte:

	/a:/	/a/	/u:/	/y:/	/e:/
/a:/	1,0	0,88	0,62	0,55	0,47
/au/	0,66	0,5	0,47	0,55	0,47
/o:/	0,62	0,47	0,85	0,78	0,71

Tab 8.5: Phonetisch-segmentale Ähnlichkeit für die Vokalphoneme

/a:/ /a/ /u:/ /y:/ /e:/ /au/ /o:/
(0 = minimale Ähnlichkeit; 1 = maximale Ähnlichkeit)

Schritt (2): Ermittlung der Gesamtdistanz zweier Phonemsequenzen

Die Berechnung der Gesamtdistanz zweier Phonemsequenzen erfolgt mittels einer durch einen phonetischen Merkmalsvergleich erweiterten Edit-Distanz. Dabei wird die phonetisch-segmentale Distanz $pDist(a, b)$ als Gewichtungsfaktor in den oben dargestellten Algorithmus der Edit-Distanz integriert. Ausgehend

von einer Berechnung der Edit-Distanz, bei der sämtliche Substitutionen mit der Konstante $subs-cost = 1$ in die Distanzberechnung eingehen, ergibt sich aus dem bisher Gesagten folgende Modifikation des Algorithmus:

Standard Edit-Distanz: $subs-cost(a, b) = 1$

Phonetisch gewichtete Edit-Distanz: $subs-cost(a, b) = 1 * pDist(a, b)$

Derzeit ist diese Gewichtung auf die Operation der *Substitutionen* beschränkt. Für künftige Erweiterungen des Algorithmus ist es aber durchaus denkbar, auch Einfügungen und Löschungen mit einer segmentalen Gewichtung zu versehen. So erscheint es beispielsweise in bestimmten Kontexten sinnvoll, daß segmentale Einfügungen eine geringere Gewichtung erhalten, wenn das eingefügte Segment den unmittelbar angrenzenden Segmenten phonetisch ähnlich ist.

Schritt (3): Normierung der Distanz

Der ermittelte Distanzwert wird schließlich normiert, d.h. auf einen Basis-Referenzwert bezogen. Dies ist erforderlich, weil die (unmodifizierte) Edit-Distanz einen positiven Ganzzahlwert liefert, der die für das Sequenzpaar (a, b) maximal mögliche Distanz nicht als Bezugsgröße berücksichtigt. Dies würde dazu führen, daß z.B. für die drei folgenden Paare von Zeichenketten

- | | | |
|-----|--------------|--------------|
| (1) | ABCD | FGHI |
| (2) | XYZ-ABCD | XYZ-FGHI |
| (3) | XYZ-ABCD-XYZ | XYZ-FGHI-XYZ |

eine identische Edit-Distanz $D = 4$ ermittelt wird (oder $D = 8$ bei einer Konstante $subs-cost = 2$). Dies widerspricht jedoch der Intuition, daß die Distanz für das Paar (1) höher sein sollte, da ja hier *sämtliche* Symbole der Sequenzen voneinander verschieden sind - und damit also die maximal mögliche Distanz für zwei Sequenzen dieser Länge erreicht ist -, während in (2) und (3) Teilübereinstimmungen bestehen, welche in Form einer reduzierten Distanz zu berücksichtigen sind.

Dieses Defizit kann mittels einer Normierung unter Berücksichtigung der maximal möglichen Edit-Distanz D_{max} behoben werden, auf den die ermittelte Anzahl der Operationen S , D , und I zu beziehen ist. Bei $D_{max}(a, b)$ handelt es sich um die maximale Edit-Distanz, welche für zwei Sequenzen der Länge von a und b auftreten kann. Im Falle von ABCD und FGHI ist dieser Wert D_{max} erreicht, da keine Übereinstimmung zwischen den beiden Sequenzen besteht. Bei den Paaren (ii) und (iii) dagegen ist D_{max} aufgrund der Teilübereinstimmungen nicht erreicht. Bei einer Konstante $subs-cost = 1$ entspricht der Wert für D_{max} der

Länge der längeren der beiden zu vergleichenden Sequenzen. Dies ist die minimale Edit-Distanz zweier Sequenzen a und b , die keinerlei Übereinstimmung aufweisen.

Im Fall des Sequenzpaares (1) beträgt dieser Wert $D_{\max} = 4$, in Sequenzpaar (2) $D_{\max} = 8$ und in Sequenzpaar (3) $D_{\max} = 12$. Zur Normierung des Distanzwertes wird nun die ermittelte Edit-Distanz $D(a, b)$ auf den jeweiligen für das Zeichenfolgenpaar gültigen Wert $D_{\max}(a, b)$ bezogen:

$$D_{\text{norm}}(a, b) = \frac{D(a, b)}{D_{\max}(a, b)}$$

Nach dieser Berechnung ergeben sich nun folgende, modifizierte Werte für die oben genannten Beispiele (1), (2) und (3):

$$\begin{aligned} (1) \quad D_{\text{norm}}(\text{ABCD}, \text{FGHI}) &= 4/4 = 1,00 \\ (2) \quad D_{\text{norm}}(\text{XYZ-ABCD}, \text{XYZ-FGHI}) &= 4/8 = 0,50 \\ (3) \quad D_{\text{norm}}(\text{XYZ-ABCD-XYZ}, \text{XYZ-FGHI-XYZ}) &= 4/12 = 0,33 \end{aligned}$$

Auf diese Weise erfolgt eine Normierung der Distanzwerte, bei der der Wert 1 als obere Schwelle möglicher Distanzwerte eintritt. Die so normierten Distanzwerte reflektieren mögliche Teilübereinstimmungen zwischen den Zeichenketten in der gewünschten Weise.

Schritt (4): Transformation in einen Ähnlichkeitswert

Schließlich wurde - u.a. aus Gründen der besseren Vergleichbarkeit verschiedener Distanzmaße - der ermittelte Distanzwert für die gesamten Zeichenketten in einen *Ähnlichkeitswert* transformiert (wie bereits bei der Berechnung der phonetisch-segmentalen Distanz). Die Ähnlichkeit zwischen zwei Zeichenketten wird dabei mit Kommazahlen im Bereich [0 .. 1] angegeben. Hierbei bezeichnet der Wert 1,0 eine maximale Ähnlichkeit (= Identität); niedrigere Werte kennzeichnen eine geringere Ähnlichkeit zwischen den Phonemsequenzen; der Wert 0 bezeichnet die „maximale Differenz“. So ergeben sich beispielsweise bei einer Distanzberechnung zwischen der kanonischen Aussprache von *Hamburg* [hambʊɾk] und den potentiellen Aussprachevarianten [hɑ:mʊɾk], [hambʊɾg], [ambʊɾg] folgende Werte für die phonetisch gewichtete Ähnlichkeit (bezeichnet als *PSimilarity*):

<i>Hamburg</i>	[hambʊɤk]	[ha:mʊɤk]	[hambyɤg]	[ambyɤg]
[hambʊɤk]	1,0 (ident.)	0,924	0,768	0,673

Tabelle 8.6: Phonetisch gewichtete Ähnlichkeit *PSimilarity* zwischen kanonischer Aussprache und Aussprachevarianten von *Hamburg*

Dabei ist zu beachten, daß die „maximale Differenz“ 0 zwischen zwei phonetischen Sequenzen in der Praxis nie erreicht wird, da durch die phonetisch-segmentale Distanzberechnung stets eine minimale segmentale Ähnlichkeit zwischen den ersetzten Segmenten wirksam wird, die den Ähnlichkeitswert über die Schwelle von 0 hebt.

8.4 Anwendung des Verfahrens

Das in den vorausgehenden Abschnitten dargestellte Distanzmaß bildet die wichtigste Grundlage des Evaluierungsverfahrens, mit dem die Plausibilität und linguistische Adäquatheit der entwickelten Regelsätze geprüft wurde. Der beschriebene Algorithmus zur Berechnung einer phonetisch gewichteten Distanz wurde programmtechnisch implementiert und in eine Evaluierungsprozedur integriert, so daß große Teile der Evaluierung automatisiert erfolgen können. Die Evaluierung ist dabei stets sowohl auf eine spezifische Sprechergruppe als auch auf ein spezifisches Test-Lexikon (Vokabular) bezogen. Dabei wird für jeden Lexikoneintrag und jeden Sprecher die phonetische Distanz zwischen der vom Sprecher produzierten Aussprachevariante und (i) der kanonischen Aussprache sowie (ii) den automatisch generierten Varianten ermittelt. Ergebnis dieser Berechnung ist die *beste Approximation*, d.h. die Ermittlung der Aussprachevariante mit der geringsten Distanz zur Sprechervariante. Die beste Approximation wird hierbei entweder (i) durch die kanonische Form oder (ii) durch eine der 4 regelbasierten Varianten erreicht.

Dies ist im folgenden am Beispiel des englischen Ortsnamens *Ashford* illustriert, bei dem die phonetischen Ähnlichkeiten zwischen einer Sprechervariante, der kanonischen Aussprache und den regelbasierten Varianten der Akzentstufen 1 bis 4 (*Level 1 ... Level 4*) berechnet wurden:

Sprechervariante: [ɛʃfət]		PSimilarity (0 ≤ N ≤ 1)
kanonisch:	[æʃfəd]	0,470
Level 1:	[ɛʃfət]	0,848
Level 2:	[ɛʃfɔ̃t]	0,909 ← beste Approximation
Level 3:	[aʃfɔ̃t]	0,744
Level 4:	[aʃfɔ̃t]	0,744

Der Wert *PSimilarity* gibt die phonetischen Ähnlichkeiten an, die mittels des oben skizzierten Algorithmus berechnet wurden. Aus ihnen wird die beste Approximation ermittelt. Im Beispiel erreicht die Akzentstufe 2 mit *PSimilarity* = 0,909 die höchste Approximation an die Sprechervariante.

In der vollständigen Evaluierungsprozedur wurde die phonetische Approximation jedoch nicht auf einzelne Sprecher, sondern stets auf eine ganze Sprechergruppe bezogen, um die sprecherübergreifende Validität der Regeln zu quantifizieren. Weiterhin ist die Evaluierungsprozedur stets auf ein spezifisches Test-Lexikon bezogen, für das die Ergebnisse gelten. Das Verfahren zur Bestimmung der sprechergruppenspezifischen phonetischen Adäquatheit der Regeln durch Ermittlung der phonetischen Distanz kann zusammenfassend wie folgt skizziert werden:

- (1) Basis der Evaluierung ist ein kanonisch transkribiertes phonetisches Lexikon mit N Einträgen. Die Einträge entsprechen dem in der Sprachdatensammlung verwendeten Material.
- (2) Zu jedem Worteintrag W des zur Evaluierung herangezogenen phonetischen Lexikons werden mittels des Regelsystems vier alternative Aussprachevarianten $V_1(W) \dots V_4(W)$ automatisch generiert. Die im Lexikon angegebene kanonische Transkription für W wird als $V_0(W)$ bezeichnet.
- (3) Für jeden Eintrag W werden die manuell erstellten phonetischen Transkriptionen der von den Sprechern produzierten Varianten als Referenzform herangezogen. Bei 20 Sprechern würden sich hierbei 20 Aussprachevarianten von W ergeben, die als $S_1(W) \dots S_{20}(W)$ bezeichnet werden.
- (4) Für jede Sprechervariante $S_1(W) \dots S_{20}(W)$ wird die phonetische Distanz D zu jeder der automatisch generierten Varianten $V_1(W) \dots V_4(W)$ sowie zur kanonischen Aussprache $V_0(W)$ ermittelt. Die Transkription mit der minimalen phonetischen Distanz zur sprecherspezifischen Aussprache $S_n(W)$ gilt als die *optimale Approximation*. Dabei handelt es sich entweder um die kanonische Aussprache $V_0(W)$ oder um eine der vier automatisch generierten Varianten $V_1(W) \dots V_4(W)$.

(5) Hier wird nun eine binäre Entscheidung getroffen: Weist eine der Varianten $V_1(W) \dots V_4(W)$ eine niedrigere Distanz zur Sprechervariante $S_n(W)$ auf als die kanonische Aussprache $V_0(W)$, so sind die regelbasierten Varianten hinsichtlich ihrer phonetischen Adäquatheit der kanonischen Aussprache überlegen. Dabei ist es unerheblich, welche der Varianten $V_1(W) \dots V_4(W)$ als optimale Approximation ermittelt wurde; entscheidend ist vielmehr die Frage, *ob* mittels der Regeln eine höhere Approximation erzielt werden kann.

(6) Schritt (5) wird für jede Sprechervariante $S_1(W) \dots S_{20}(W)$ wiederholt. Dabei wird die relative Häufigkeit ermittelt, mit der die regelbasiert generierten Varianten eine gegenüber der kanonischen Aussprache höhere Approximation erzielen. Dieser Wert gibt Auskunft über die beste Approximation für W , bezogen auf die gesamte untersuchte Sprechergruppe.

(7) Die Schritte (3) bis (6) werden für jedes W im Lexikon durchgeführt. Im Anschluss wird – hier bezogen auf das gesamte Lexikon – ermittelt, mit welcher Häufigkeit die regelbasierten Varianten eine höhere Approximation erzielen als die kanonische Aussprache. Dieser Wert schließlich gibt Auskunft über die Adäquatheit der Regeln, bezogen auf ein spezifisches Lexikon und eine Sprechergruppe.

Formal lässt sich der Test, bezogen auf eine Sprechergruppe und ein spezifisches Lexikon, demnach wie folgt darstellen:

<i>testlex</i>	Lexikon (Vokabular des Tests)
$W_1 .. W_i$	Lexikoneinträge, $W \in testlex$
<i>spkr_group</i>	Sprechergruppe des Tests
$S_1 .. S_n$	Sprecher, $S \in spkr_group$
$V_0 (W)$	kanonische Transkription von W
$V_1 .. V_4 (W)$	Automatisch generierte Aussprachevarianten für Wort W
GENERATE_VARIANTS	Funktion zur automatischen Generierung von Aussprachevarianten
COMPUTE_DISTANCE	Funktion zur Berechnung der phonetischen Distanz
GETMINIMUM	Funktion zur Ermittlung des Minimums aus mehreren Werten
for each $W \in testlex$ do	
$V_1 .. V_4 (W) \leftarrow GENERATE_VARIANTS (W)$	
for each $S \in spkr_group$ do	
for each $V_0 .. V_4 (W)$ do	
$D \leftarrow COMPUTE_DISTANCE (S(W), V_i(W))$	
$DMIN \leftarrow GETMINIMUM (D (V_1) .. D (V_4))$	
if $DMIN < D (S_i, V_0)$ then	
<i>variants++</i>	
else	
<i>canonical++</i>	

Abb. 8.3: Darstellung des Testverfahrens als Pseudo-Code

Die in dieser Testprozedur akkumulierten Werte *canonical* und *variants* geben an, mit welcher absoluten Häufigkeit die kanonische Transkription bzw. eine der vier regelbasierten Varianten die beste Approximation erzielt haben. Diese Werte werden schließlich in eine relative Häufigkeit (%) transformiert, die sowohl auf ein spezifisches Vokabular als auch auf eine spezifische Sprechergruppe bezogen ist. Alle in den folgenden Abschnitten dargestellten Ergebnisse beruhen sämtlich auf dieser Testprozedur. Als Ergebnis ist hierbei stets die relative Häufigkeit der besten Approximation, bezogen auf (a) eine Sprechergruppe und (b) ein Lexikon angegeben.

8.5 Ergebnisse der Evaluierung

8.5.1 Sprechergruppen und Sprachmaterial

Das oben skizzierte Verfahren wurde zur Evaluierung auf zwei Ebenen angewandt: (a) für eine sprecherbezogene Evaluierung, in der die Abdeckung von interindividueller (d.h. sprecherübergreifender) Variation durch die Regeln ermittelt wird, und (b) für eine vokabularbezogene Evaluierung, bei der geprüft wird, ob und wie weit die entwickelten Regeln auf neues Vokabular übertragbar sind. Beide Ebenen konnten auf der Basis derselben Evaluierungsprozedur bearbeitet werden und sind daher im folgenden gemeinsam dargestellt. Eine Evaluierung wurde für drei verschiedene Sprechergruppen durchgeführt:

Gruppe A: Sprecher, deren Aussprachevarianten bereits zur Erstellung der Regeln herangezogen wurden (pro Muttersprache 16 bis 25 Sprecher). Zwar handelt es sich hierbei um jene Sprecher, aus deren Aussprachevarianten die Regeln abgeleitet wurden, wodurch die Gefahr der Zirkularität gegeben ist. Jedoch liefert der hier ermittelte Wert einen wichtigen Referenz-Wert, der idealerweise auch mit neuen Sprechergruppen erreicht werden sollte.

Gruppe B: Sprecher, die nicht zur Regelerstellung herangezogen wurden (6 Sprecher pro Muttersprache). Die Zusammensetzung der Gruppe ist hinsichtlich Alter und Fremdsprachenkenntnissen vergleichbar mit Gruppe A.

Gruppe C: Muttersprachler Italienisch (10 Sprecher), die bei der Regelerstellung bisher nicht berücksichtigt wurden. Die Zusammensetzung der Gruppe unterscheidet sich von Gruppe A und B durch das höhere Alter der Sprecher (45-65 Jahre) und geringe Fremdsprachenkenntnisse.

Da ein wesentliches Interesse der Evaluierung darin bestand, ob und in welchem Maße die Regelsätze für neue, bisher nicht berücksichtigte Sprecher Gültigkeit besitzen (sprecherübergreifende Generalisierbarkeit der Regeln), sind im folgenden nur die Ergebnisse für Gruppen B und C dargestellt, d.h. für solche Sprecher, welche in die Regelerstellung nicht eingeflossen sind. Auf eine detaillierte Vorstellung der Ergebnisse für Gruppe A wird hier zugunsten einer gestrafften Darstellung verzichtet. Zusammenfassend sei jedoch festgehalten, daß die Testergebnisse für Gruppe A weitgehend vergleichbar mit denen der Gruppe B sind, wobei erwartungsgemäß die Approximation durch die regelbasierten Varianten in vielen Sprachrichtungen (geringfügig) höher liegt als in Gruppe B.

8.5.2 Ergebnisse Sprechergruppe B

Das Testdaten-Set der Gruppe B ist von besonderem Interesse hinsichtlich der Frage, ob die Regeln auch für neue Sprecher plausible Varianten generieren. Die Ergebnisse sind also nicht nur bezüglich der Frage nach der Abdeckung *interindividueller Variabilität* von Bedeutung, sondern darüber hinaus auch hinsichtlich der Einschätzung des Potentials der Regeln zur *Prädiktion von Aussprachevariation*. Dabei ist die Sprecherdistribution in Gruppe B in Bezug auf Alter und L2-Kenntnisse mit Gruppe A vergleichbar.

Die folgende Tabelle zeigt die Ergebnisse für 6 Sprachrichtungen, die sich aus der Kombination der Muttersprachen **Englisch, Deutsch, Französisch** und **Italienisch** mit den Zielsprachen **Englisch** und **Deutsch** ergeben.

L1	L2					
	Englisch			Deutsch		
Englisch				<i>ct_read</i> (6)	<i>ct_repeat</i> (6)	<i>sent_read</i> (6)
kanonisch				32.2%	57.8%	50.0%
Level 1				45.0%	35.3%	34.0%
Level 2				15.9%	4.7%	13.5%
Level 3				4.3%	0.4%	2.2%
Level 4				2.7%	1.9%	0.3%
var_gesamt				67.8%	42.2%	50.0%
Deutsch	<i>ct_read</i> (7)	<i>ct_repeat</i> (7)	<i>sent_read</i> (7)			
kanonisch	37.4%	42.3%	54.9%			
Level 1	41.5%	45.1%	36.5%			
Level 2	14.6%	7.8%	6.3%			
Level 3	5.8%	4.4%	1.3%			
Level 4	0.7%	0.3%	1.0%			
var_gesamt	62.6%	57.7%	45.1%			
Französisch	<i>ct_read</i> (6)	<i>ct_repeat</i> (6)	<i>sent_read</i> (6)	<i>ct_read</i> (6)	<i>ct_repeat</i> (6)	<i>sent_read</i> (6)
kanonisch	40.1%	61.3%	57.6%	29.8%	50.4%	54.0%
Level 1	28.4%	27.7%	24.7%	50.2%	43.0%	22.5%
Level 2	17.1%	5.1%	11.6%	14.7%	6.6%	18.6%
Level 3	11.3%	5.1%	4.9%	5.3%	0.0%	4.8%
Level 4	3.1%	0.8%	1.2%	0.0%	0.0%	0.0%
var_gesamt	59.9%	38.7%	42.4%	70.2%	49.6%	46.0%
Italienisch	<i>ct_read</i> (6)	<i>ct_repeat</i> (6)	<i>sent_read</i> (6)	<i>ct_read</i> (6)	<i>ct_repeat</i> (6)	<i>sent_read</i> (6)
kanonisch	29.8%	50.0%	55.9%	29.4%	41.2%	57.6%
Level 1	33.3%	31.0%	31.9%	19.8%	28.4%	13.7%
Level 2	22.9%	8.1%	8.2%	25.2%	17.9%	18.0%
Level 3	8.9%	9.3%	1.8%	22.5%	11.3%	9.7%
Level 4	5.0%	1.6%	2.1%	3.1%	1.2%	1.0%
var_gesamt	70.2%	50.0%	44.1%	70.6%	58.8%	42.4%

Tabelle 8.7: Ergebnisse der Evaluierung Sprechergruppe B (6 Sprecher pro L1, 7 Sprecher bei L1 Deutsch), Zielsprachen Englisch und Deutsch

ERLÄUTERUNG DER DARSTELLUNG IN DEN TABELLEN

Die vorausgehende und die später folgende Tabelle (Sprechergruppe C) zeigen die Ergebnisse für verschiedene Sprachrichtungen in einer nach Muttersprache (L1) und Zielsprache (L2) aufgeschlüsselten Matrix. Je Sprachrichtung sind die Ergebnisse zweifach gegliedert:

(1) **Nach Vokabular und Aufnahmeszenario.** Die Sprachaufnahmen gliederten sich in jeweils drei Teile: (a) *Ortsnamen, gelesen*; (b) *Ortsnamen, Wiederholung akustischer Prompts* sowie (c) *kurze Sätze, gelesen*. Die Tabelle gibt die Ergebnisse für die in diesen drei Szenarien entstandenen Sprachaufnahmen wieder: *ct_read* (= Ortsnamen, gelesen), *ct_repeat* (= Ortsnamen, Wiederholung akustischer Prompts, und *sent_read* (= Sätze, gelesen). Die Zahlen in Klammern geben die Anzahl der Sprecher an, die in die Testprozedur einbezogen wurden.

(2) **Nach Akzentstufen.** Die prozentualen Häufigkeiten der besten Approximation sind angegeben für (a) die kanonische Aussprache, (b) die einzelnen automatisch generierten Akzentstufen (*Level 1 .. 4*) und (c) die Gesamtheit aller automatisch generierten Akzentstufen (*var_gesamt*). Hieraus kann abgelesen werden, mit welcher Häufigkeit die regelbasiert generierten Varianten die beste Approximation erzielen konnten. Darüber hinaus ist abzulesen, welchen Anteil einzelne Akzentstufen hieran aufweisen.

8.5.3 Ergebnisse bei Ortsnamen, isolierte Wörter

Betrachtet man die Ergebnisse zunächst für das Szenario *ct_read* (Ortsnamen, gelesen), so ist zu beobachten, daß bei allen Sprachrichtungen in deutlich mehr als 50% aller Fälle die beste Approximation mittels der regelbasierten Varianten erreicht wurde (Wert *var_gesamt*). Dieser Wert variiert jedoch in Abhängigkeit von der behandelten Sprachrichtung. So wurde für die Sprachrichtung L1 Italienisch → L2 Englisch sowie L1 Französisch → L2 Deutsch in 70,2% der Fälle die höchste Approximation mittels der Varianten erzielt, während der Wert bei der Sprachrichtung L1 Französisch → L2 Englisch lediglich bei 59,9% liegt.

Neben Unterschieden im Grad der Ausarbeitung einzelner Regelsätze ist auch die Zusammensetzung der jeweiligen Sprechergruppe für derartige Differenzen verantwortlich. Grundsätzlich gilt dabei, daß bei insgesamt guten L2-Kenntnissen innerhalb der Testgruppe tendenziell eine höhere Approximation durch die kanonische Variante erwartbar ist. So erreichen z.B. die Kenntnisse der Zielsprache Englisch bei den französischen Sprechern der Gruppe B einen relativ hohen Mittelwert von 3,0 (max. 4,0).

Bei allen Sprachrichtungen wird die beste Approximation durch Varianten niedriger Akzentstufen erreicht (Stufe 1 und 2), während hohe Akzentstufen nur sehr geringe Anteile erzielen. Dies ist jedoch keinesfalls als Indiz dafür zu werten, daß die Varianten der Akzentstufen 3 und 4 völlig inadäquat sind. Vielmehr steht dieses Ergebnis in engem Zusammenhang mit der Sprecherdistribution der Datensammlung: Da ein überproportionaler Anteil an Sprechern mit verhältnismäßig guten Kenntnissen der Zielsprachen vorliegt, sind extrem ausgeprägte Akzentvarianten, wie sie in den hohen Akzentstufen modelliert werden, in den Testdaten unterrepräsentiert. Es ist jedoch zu erwarten, daß in Untersuchungen mit Sprechern geringer Kenntnisstufen die Varianten der oberen Akzentstufen eine höhere Approximation erzielen. So haben informelle Tests gezeigt, daß sich die Werte für die beste Approximation um bis zu 15% zugunsten der automatisch generierten Varianten verschieben, wenn die Evaluierungsprozedur nur für Teilgruppen von Sprechern mit relativ geringen L2-Kenntnissen durchgeführt wird. Hierbei ist auch zu beobachten, daß der Anteil der höheren Akzentstufen an der besten Approximation ansteigt.

Obwohl in der Leseaufgabe *ct_read* und der Wiederholungsaufgabe *ct_repeat* identisches Vokabular verwendet wurde, erreichen die Regeln bei der Wiederholungsaufgabe generell eine geringere Approximation als im Szenario *ct_read*. Diese Differenz ist u.a. dadurch zu erklären, daß verschiedene Aussprachefehler, die in der Leseaufgabe mit hoher sprecherübergreifender Frequenz und Regelmäßigkeit auftreten, deutlich reduziert oder vollständig eliminiert sind, wenn die kanonische Aussprache akustisch dargeboten wird und diese von den Sprechern nachgesprochen wird. Eine anhand der Sprachdaten exemplarisch durchgeführte quantitative Untersuchung verschiedener Aussprachefehler hat diese Tendenz bestätigt. Aussprachefehler sind demnach in diesem Szenario generell weniger ausgeprägt; entsprechend erreicht die kanonische Transkription hier häufig eine höhere Approximation als die regelbasierten Varianten.

Dieser Sachverhalt ist jedoch grundsätzlich im Modell der Akzentstufung berücksichtigt, da (1) phonetische Varianten mit geringer ausgeprägten Aussprachefehlern von jeweils niedrigeren Akzentstufen abgedeckt werden und (2) die kanonische Aussprache in den regelbasiert generierten Lexika stets als eine mögliche Aussprachevariante (neben den 4 Akzentstufen) erhalten bleibt.

8.5.4 Ergebnisse bei neuem Vokabular

Die bisher erstellten Regelsätze basieren auf einem relativ begrenzten Korpus von Ortsnamen und ihrer Realisierung durch Sprecher verschiedener Muttersprachen. Um zu Aussagen über ihre Generalisierbarkeit zu gelangen, ist es sinnvoll, die erstellten Regeln mit neuem Vokabular zu testen. Hierzu wurde das Vokabular der dritten Teilaufgabe der Sprachaufnahmen herangezogen (Szenario *sent_read*).

Es handelt sich hierbei nicht um Eigennamen, sondern um Standard-Vokabular der jeweiligen Zielsprachen.

In allen Sprechergruppen weichen die Resultate z.T. deutlich von den Ergebnissen bei Ortsnamen ab. Bei allen Sprachrichtungen ist zu beobachten, daß die regelbasierten Varianten im Szenario *sent_read* eine niedrigere Approximation erreichen als im Falle der Ortsnamen. Hierfür sind u.a. folgende Gründe zu nennen:

- Sprecher mit durchschnittlichen bis guten Kenntnissen der Zielsprache erzielen bei Standard-Vokabular tendenziell eine größere Annäherung an die zielsprachliche Standard-Aussprache als bei Eigennamen. Dies ist zum einen damit zu erklären, daß es sich um erlerntes und memoriertes Vokabular handelt, das grundsätzlich weniger anfällig für eine fehlerhafte Aussprache ist. Diese Beobachtung steht im Einklang mit sog. *Dual-Route-Modellen* des Lesens bzw. der Worterkennung (vgl. Coltheart 1978; Seidenberg 1999 und Kap. 6.5 dieser Arbeit): Diesen Modellen zufolge wird im Leseprozess die Aussprache nicht ausschließlich durch Graphem-Phonem-Zuordnungen (regelbasiert), sondern auch auf der Grundlage größerer Einheiten (Wort- oder Subworteinheiten) erschlossen. Es erscheint plausibel, daß im Falle von Eigennamen die regelbasierte Strategie deutlich überwiegt, weshalb es hier zu spezifischen Aussprachefehlern kommt.
- Bei der Aussprache zusammenhängender Sätze (im Gegensatz zu isolierten Wörtern) treten Phänomene auf, welche in den bisherigen Regeln noch nicht berücksichtigt sind, etwa phonetische Assimilationseffekte an Wortgrenzen, visuelle Worterkennungsfehler (Lesefehler) u.v.m.
- Die Sprachdaten weisen wortklassenspezifische Differenzen bei Aussprachefehlern auf. So treten einige Lautersetzungen, die in Eigennamen häufig zu beobachten sind, innerhalb von Funktionswörtern (etwa Konjunktionen, Artikel) nur selten auf. Eine Korrelation von Aussprachefehlern mit Faktoren wie Worthäufigkeit, Wortlänge oder morphologischer Komplexität kann als wahrscheinlich gelten; diese Hypothese ist jedoch genauer zu untersuchen, bevor entsprechende Angaben in das Regelsystem integriert werden.

Diese - vorläufig nur als Hypothesen formulierten - Ursachen bieten wichtige Ansatzpunkte für eine Weiterentwicklung der Regeln. So sind weitere Untersuchungen anhand von zielsprachlichem Standard-Vokabular im Satzzusammenhang zweckmäßig, um die Regeln für diese Domäne zu verbessern. Hierbei wäre es z.B. denkbar, bestimmte Einzelregeln in Abhängigkeit von der Wortklasse des betroffenen Wortes zu modifizieren bzw. ihre Anwendung in bestimmten Kontexten zu blockieren. So wäre es beispielsweise möglich, das Akzentstufenmodell dahingehend zu nutzen, daß die von den Regeln generierte

dahingehend zu nutzen, daß die von den Regeln generierte „Akzentstärke“ für bestimmte Wortklassen erhöht oder herabgesetzt wird.

8.5.5 Ergebnisse Sprechergruppe C

Die Evaluation für Sprechergruppe C gibt Aufschluß darüber, welche Ergebnisse bei Sprechern zu erwarten sind, deren Kenntnisse der Zielsprache(n) nur gering ausgeprägt sind. Dies ist deshalb von großem Interesse hinsichtlich des Gesamtansatzes, weil mittels der Akzentstufen grundsätzlich eine möglichst gute Repräsentation von Sprechern *aller* Kenntnisstufen erreicht werden soll. Gruppe C zeigt, welche Approximation für eine in der bisherigen Studie unterrepräsentierte Gruppe erreicht wird. Die Ergebnisse sind in der folgenden Tabelle für die Zielsprachen Englisch und Deutsch dargestellt.

L1	L2			L2		
	Englisch			Deutsch		
Italienisch	<i>ct_read</i> (10)	<i>ct_repeat</i> (10)	<i>sent_read</i> (10)	<i>ct_read</i> (10)	<i>ct_repeat</i> (10)	<i>sent_read</i> (10)
kanonisch	29.5%	45.8%	54.8%	11.7%	27.3%	40.3%
Level 1	27.4%	27.9%	31.6%	13.3%	20.3%	7.9%
Level 2	23.5%	7.2%	6.8%	33.3%	31.7%	27.7%
Level 3	11.4%	15.8%	4.6%	29.8%	17.7%	21.6%
Level 4	8.1%	3.3%	2.2%	11.9%	3.0%	2.5%
var_gesamt	70.5%	54.2%	45.2%	88.3%	72.7%	59.7%

Tabelle 8.8: Ergebnisse der Evaluierung, Sprechergruppe C, Zielsprachen Englisch und Deutsch

Für beide Zielsprachen – Englisch und Deutsch – wird mittels der regelbasierten Varianten zumeist eine gegenüber der kanonischen Form deutlich höhere Approximation erzielt. Insbesondere im Szenario *ct_read* der Zielsprache Deutsch liegt der Anteil der Varianten bei über 88% und erreicht damit den bisher höchsten Wert.

Zu erklären ist dies u.a. durch den (weiter oben bereits angedeuteten) Sachverhalt, daß bei geringeren L2-Kenntnissen der Anteil der höheren Akzentstufen an der besten Approximation steigt. So ist bei der Zielsprache Deutsch im Vergleich zu Sprechergruppe B in Tabelle 8.7 eine deutlichere Konzentration der besten Approximation bei Stufe 2 und 3 zu beobachten. Dies reflektiert die geringeren L2-Kenntnisse innerhalb der Sprechergruppe C. Die relative Abstufung zwischen den drei Szenarien *ct_read*, *ct_repeat* und *sent_read* ist dagegen ähnlich wie in Gruppe B und spiegelt erneut die Verhältnisse, wie sie oben bereits skizziert wurden: Erstens sind Fehler im Wiederholungs-Szenario tendenziell weniger

ausgeprägt, weshalb der Anteil der regelbasierten Varianten hier absinkt; zweitens wurden die Regeln bisher nicht für zusammenhängende Sätze optimiert, so daß die Approximation für das Szenario *sent_read* geringer ist.

Abschließend sei kurz die stets auffallend geringe Approximation durch Varianten der Akzentstufe 4 (*Level 4*) diskutiert. Auf dieser Akzentstufe werden Aussprachevarianten modelliert, bei denen zielsprachliche Wörter fast vollständig nach den Schriftaussprache-Regeln der Muttersprache des Sprechers phonemisiert werden. Hierbei entstehen stark abweichende Varianten wie etwa deutsch [aplətɔn] für engl. *Appleton* oder [fʁɛːjʊs] für franz. *Fréjus*. Derartige Varianten treten in der Praxis verhältnismäßig selten auf; dennoch wurden sie bewußt in das Spektrum regelbasierter Varianten aufgenommen, um die obere Grenze akzentbehafteter Varianten zu markieren und auch solche Formen abzudecken. Akzentstufe 4 repräsentiert diesen Ausnahmefall und erreicht daher erwartungsgemäß nur selten eine optimale Approximation für die Sprecher des vorliegenden Sprachkorpus.

8.6 Möglichkeiten einer weiterführenden Evaluierung

Die dargestellte Evaluierung weist ein insgesamt positives Ergebnis auf, in dem sich zeigt, daß charakteristische Aussprachefehler sprecherübergreifend generalisierbar und mittels phonologischer Regeln formulierbar sind. Trotz sprachrichtungsspezifischer Unterschiede kann als Resultat festgehalten werden, daß die regelbasierten Varianten auch bei neuen Sprechern vielfach eine gegenüber der kanonischen Aussprache höhere Approximation erzielen.

Die Ergebnisse zeigen jedoch auch, daß eine Weiterentwicklung der bestehenden Regelsätze erforderlich ist, vor allem im Hinblick auf folgende Teilbereiche:

- (i) Sprecher mit stark ausgeprägtem Akzent und die Regeln der oberen Akzentstufen 3 und 4;
- (ii) die vokabularübergreifende Gültigkeit der Regeln, insbesondere außerhalb der Domäne der Eigennamen.

Somit liefert die Evaluierung des regelbasierten Ansatzes nicht nur quantifizierbare Angaben über die Performanz des Regelsystems, sondern gibt gleichzeitig wichtige Hinweise auf Prioritäten bei der künftigen Weiterentwicklung.

Zusätzlich zu der hier dargestellten Evaluierung wurden die Regeln dadurch geprüft, indem sie auf phonetische Lexika von je 1000 neuen Ortsnamen der Sprachen Englisch, Deutsch und Französisch (Schaden 2004) angewendet wurden. Obwohl für diese Korpora keine Evaluierung im hier beschriebenen Sinne

durchgeführt wurde, vermitteln sie einen grundlegenden Eindruck davon, welche Ergebnisse bei der Anwendung der Regeln auf neues, umfangreicheres Vokabular zu erwarten sind. Anhand derartiger Testkorpora können unplausible regelbasierte Aussprachevarianten identifiziert und die entsprechenden Regeln modifiziert werden. Hierin ist ein wichtiger Ansatzpunkt für eine weitere, ergänzende Evaluierungsstrategie zu sehen.

9 Zusammenfassung: Möglichkeiten und Grenzen einer Modellierung von Akzenten

9.1 Perspektiven einer Weiterentwicklung

Mit zunehmendem Einsatz sprachtechnologischer Systeme wie Sprachsynthese, automatischer Spracherkennung oder ihrer Kombination in sogenannten „Natürlichsprachlichen Dialogsystemen“ steigt allgemein der Bedarf an Adaptionismethoden für solche Nutzer, deren Sprachverhalten nicht der jeweils als Maßstab zugrundegelegten sprachlichen Norm entspricht.

In der vorliegenden Arbeit wurde eine Methode entwickelt, um fremdsprachlich akzentgefärbte Aussprachevarianten mittels phonologischer Regelsätze auf der Ebene der phonetischen Transkription zu modellieren. Das vorgestellte Konzept zur Modellierung phonetischer Variabilität innerhalb der Domäne fremdsprachlicher Akzente kann als Verfahren der lexikalischen Adaption für sprachtechnologischer Systeme eingesetzt werden. Hierbei stehen vor allem zwei Anwendungskontexte im Mittelpunkt:

- Integration in ein automatisches Spracherkennungssystem zur Adaption an nicht-muttersprachliche Sprecher. Das Regelsystem kann hier als Verfahren der lexikalischen Adaption an spezifische Nutzergruppen verwendet werden.
- Integration in ein Sprachsynthesystem zur Generierung einer synthetischen Stimme mit fremdsprachlichem Akzent (als Komponente einer sog. „personalisierten Stimme“).

Die Ergebnisse der im vorausgehenden Kapitel dargestellten Evaluierung zeigen, daß charakteristische Aussprachefehler mittels des vorgestellten Ansatzes adäquat modelliert werden können. So können die regelbasiert generierten Varianten vielfach eine höhere phonetische Approximation an die von Sprechern produzierten akzentgefärbten Aussprachevarianten erzielen als die kanonische Aussprache. Damit ist ein wesentliches Ziel der vorliegenden Arbeit erreicht: Es wurde nicht nur die grundsätzliche Möglichkeit einer regelbasierten Modellierung fremdsprachlicher Akzente nachgewiesen und zugleich programmtechnisch umgesetzt, sondern auch ermittelt und quantifiziert, ob und in welchem Maße die innerhalb einer Sprechergruppe auftretende inter-individuelle Variabilität durch die regelbasierten Varianten und das Modell der Akzentstufen adäquat modelliert

werden kann. Das entwickelte Evaluierungsverfahren stellt dabei ein wichtiges Instrument zur Weiterentwicklung der Regeln dar, da künftige Regelmodifikationen mittels dieses Verfahrens einem unmittelbaren Test unterzogen werden können.

Mit der vorgestellten Arbeit wurde Grundlagenwissen für sprachtechnologische Systeme erarbeitet und operationalisiert, das in verschiedenen Anwendungszusammenhängen (automatische Spracherkennung, Sprachsynthese u.a.) zum Einsatz kommen kann. Lexikalische Adaptionstechniken, welche in der automatischen Spracherkennung und -synthese bisher lediglich zur Modellierung intralingualer Variation angewandt wurden, wurden dabei ausgeweitet auf den Bereich sprachenübergreifender Variation. Dabei ist das entwickelte Verfahren fundiert durch linguistische Überlegungen und Modelle, die in der bisherigen Forschung noch nicht in Beziehung zu sprachtechnologischen Fragen gesetzt wurden.

Dabei wurde besonderer Wert auf die Konzeption des Regelsystems als einer systemunabhängigen, generischen Wissensquelle gelegt. Das entwickelte Regelsystem wurde daher so gestaltet, daß es in gängigen Spracherkennungs- oder Sprachsynthesystemen ohne grundlegende Eingriffe in die Programmarchitektur integriert werden und zum Einsatz kommen kann. Weitere – in dieser Arbeit nicht thematisierte – mögliche Anwendungsgebiete des entwickelten Ansatzes liegen in der Generierung von regionalen bzw. dialektalen Aussprachevarianten für sprachtechnologische Systeme (vergleichbare Arbeiten für das Englische cf. Fitt & Isard 1999). Der Aufbau des Regelsystems erlaubt einen solchen Einsatz ohne wesentliche Modifikationen. Weiterreichende Einsatzmöglichkeiten ergeben sich auch für das in Kapitel 8 entwickelte phonetische Distanzmaß, das überall dort eingesetzt werden kann, wo ein quantitativer Vergleich von Phonemsequenzen sinnvoll ist, etwa bei der Evaluierung der Performanz von Graphem-nach-Phonem-Umsetzern, in dialektphonologischen Studien oder auch im Bereich von Sprachverständlichkeitstests.

9.2 Grenzen der Untersuchung

Diese Arbeit bewegt sich im interdisziplinären Schnittfeld von Sprachwissenschaft und Sprachtechnologie. Dieses Schnittfeld ist nicht selten auch ein Spannungsfeld: Die praktisch-technischen Erfordernisse der Sprachtechnologie legen mitunter Lösungswege nahe, die aus linguistischer, sprachpsychologischer oder kommunikationswissenschaftlicher Perspektive nicht als die plausibelsten erscheinen, während andererseits die Sprachwissenschaft zumeist keine unmittelbar technisch umsetzbaren Modelle und Methoden bereitstellt (und dies auch nicht primär anstrebt), die für sprachtechnologische Entwicklungsarbeiten genutzt werden

könnten. Einige der heute gängigen und erfolgreich angewendeten Methoden der Sprachtechnologie mögen aus Sicht einer auf Erklärungsadäquatheit und konsistente Theoriebildung bedachten Sprachwissenschaft, Sprachpsychologie oder Kommunikationswissenschaft lediglich als *ad hoc*-Lösungen ohne weiterreichenden explanativen Wert erscheinen. So kann etwa der seit den 1980er Jahren währende Erfolg statistischer Methoden auf dem Gebiet der automatischen Spracherkennung nur schwerlich damit begründet werden, daß der Mensch bei der kognitiven Verarbeitung von Lautsprache tatsächlich vorwiegend nach statistischen oder probabilistischen Prinzipien verfähre und sich hieraus die Überlegenheit einer solchen Modellierungstechnik geradezu „natürlich“ ableite.

Dennoch können sich derartige Lösungen – zumindest vorläufig – als technisch überlegen erweisen. Dies rührt zum Teil daher, daß viele Prozesse menschlicher Sprachverarbeitung (noch) nicht in dem Maße verstanden sind, daß eine technische Modellierung auf der Basis des vorhandenen Wissens möglich wäre. In diesem Falle erlangen technische Verfahren gerade dadurch ihre Berechtigung, daß mit ihnen *nicht* versucht wird, menschliches Sprachverhalten von Grund auf zu rekonstruieren, sondern es lediglich in eingrenzbaaren Teilaspekten zu *simulieren*. Die primäre Anforderung an sprachtechnologische Systeme ist in diesem Falle nicht ihre Übereinstimmung mit den Mechanismen menschlicher Sprachverarbeitung, sondern ihre Funktionalität und praktisch-technische Verwertbarkeit.

Darüber hinaus spricht jedoch auch manches dafür, daß bei der menschlichen Sprachverarbeitung Prozesse involviert sind, die sich einem technischen Zugang ganz entziehen. Eine technische Modellierung setzt stets voraus, daß das zu modellierende Phänomen zunächst analytisch in seine Bestandteile zergliedert wird, um es auf dieser Grundlage einer Operationalisierung zugänglich zu machen – diesem methodischen Ansatz folgt grundsätzlich auch die vorliegende Arbeit. Dabei sollte jedoch stets bedacht werden, daß in einem Forschungskontext wie dem vorliegenden die Basiseinheiten der Analyse nicht notwendigerweise denjenigen Einheiten entsprechen, die tatsächlich für einen Sprecher oder Hörer im Vollzug sprachlicher Äußerungen relevant sind. So wies Ungeheuer (1968/72) – in einem anderen thematischen Zusammenhang – auf das Vorliegen einer grundlegenden Dichotomie zwischen einer „extrakommunikativen“ und einer „kommunikativen“ Betrachtungsweise von Sprache hin und verwies darauf, daß die Analyse-Einheiten der Linguistik (wie sie auch in der vorliegenden Studie verwendet wurden) keineswegs immer deckungsgleich mit den Einheiten der sprachlichen Praxis seien.

Ähnliche Überlegungen finden sich bereits deutlich früher bei Hermann Paul (1909), der das Problem einer Gleichsetzung von Sprachstruktur und „lebendiger Rede“ wie folgt umriß:

Es ist ein grosser Irrtum, wenn man meint, daß, um den Klang eines Wortes in seiner Eigentümlichkeit zu erfassen, so daß eine Erregung der damit assoziierten Vorstellungen möglich wird, die einzelnen Laute, aus denen sich das Wort zusammensetzt, zu deutlichem Bewusstsein gelangen müssten. Es ist sogar, um einen ganzen Satz zu verstehen, nicht immer nötig, daß die einzelnen Wörter ihrem Klange und ihrer Bedeutung nach zum Bewusstsein kommen. Die Selbsttäuschung, in der sich die Grammatiker bewegen, rührt daher, daß sie das Wort nicht als einen Teil der lebendigen, rasch vorüberauschenden Rede betrachten, sondern als etwas Selbständiges, über das sie mit Musse nachdenken, so daß sie Zeit haben, es zu zergliedern. (Paul 1909: 50)

Diese grundsätzliche Überlegung zum Zusammenhang von sprachwissenschaftlicher Analyse und sprachlichem Handlungsvollzug, die hier in allgemeiner Weise vorgetragen wird, ist zweifellos auch für die heutige *Sprachtechnologie* (und damit auch für die vorliegende Arbeit) in spezieller Weise relevant: Forschungsaktivitäten auf diesem Gebiet bringen oftmals einen Gegenstand hervor, welcher der alltäglich-vertrauten „lebendigen Rede“ nicht immer ähnlich ist. Wenn in den vorausgehenden Kapiteln etwa die Rede davon war, daß fremdsprachliche Akzente als „Substitution von Phonemen“ beschreibbar und formalisierbar sind, so geschah dies stets in dem Bewußtsein, daß es sich um eine notwendige Reduktion handelt, um den Gegenstandsbereich in einer Weise zu erfassen, die mit den derzeitigen Methoden und technischen Voraussetzungen der maschinellen Sprachverarbeitung kompatibel ist.

Das zuletzt Gesagte kann ohne weiteres auch auf das in dieser Arbeit diskutierte Thema fremdsprachlicher Akzente bezogen werden: Denn im alltäglichen Umgang mit Sprache treten für einen Sprecher oder Hörer möglicherweise ganz andere Eigenschaften fremdsprachlicher Akzente als besonders relevant, auffällig oder charakteristisch hervor als diejenigen phonetischen Aspekte, die in dieser Arbeit thematisiert wurden. Aus diesem Grund sind auch linguistische Ansätze stets mit Skepsis zu betrachten, die durch das Bestreben gekennzeichnet sind, aus linguistisch-phonetischen Eigenschaften der Sprachform unmittelbare Schlüsse auf die damit zu erreichenden Verständigungsleistungen zu ziehen. In der Literatur zur Phonetik und Phonologie des Zweitsprachenerwerbs sind beispielsweise immer wieder Hinweise auf „praktische Verständnisprobleme“ zu finden, welche nach Meinung der Autoren *unmittelbar* auf defiziente Phonemrealisierungen zurückzuführen seien. Demzufolge würden die Verständnisprobleme, mit denen sich ein nicht-muttersprachlicher Sprecher in seinem sprachlichen Alltag möglicherweise konfrontiert sieht, proportional zur Anzahl der Phonemverschiebungen oder -verwechslungen ansteigen, die der Sprecher produziert. Derartige Überlegungen gründen sich auf die Idee einer unmittelbaren Kopplung von Sprachform und Sprachfunktion; sie gehen explizit oder implizit von der Annah-

me sprachlicher „Wohlgeformtheit“ von Äußerungen als notwendiger und hinreichender Bedingung des Verstehens und des erfolgreichen sprachlichen Handlungsvollzugs aus.

Tatsächlich aber erlaubt die situative Einbettung von Sprache in der Praxis immer auch weitreichende Lautverschiebungen, -verwechslungen und -elisionen, ohne daß damit die sprachliche Verständigung ernsthaft gefährdet wäre. Dies gilt auch und gerade für Sprechweisen, welche durch einen fremdsprachlichen Akzent gekennzeichnet sind. Die im Rahmen dieser Arbeit vorgestellten Modelle und Überlegungen sagen also zunächst nichts darüber aus, wie sich die beschriebenen und modellierten phonetischen „Abweichungen“ sprachpraktisch auswirken. Ein solches Thema wäre aus einer anderen bzw. deutlich erweiterten wissenschaftlichen Perspektive zu behandeln.

Statt dessen ist das in der vorliegenden Arbeit vorgestellte Modell in weiten Teilen als ein „extrakommunikatives“ zu verstehen, in dem gewisse perspektivische Eingrenzungen und Vereinfachungen (vorübergehend) in Kauf genommen werden. Sowohl der Grad als auch die Art dieser Vereinfachungen orientiert sich dabei stets an den aktuell gegebenen sprachtechnologischen Rahmenbedingungen (d.h. am technisch *Machbaren*) sowie auch an praktischen Erfordernissen (d.h. am technisch *Erwünschten*). Eine solche Reduktion beinhaltet stets, daß verschiedene Fragen, die zweifellos in engem Zusammenhang zum hier behandelten Thema stehen, gezielt ausgeblendet werden. Somit ist zwangsläufig auch mit einer relativ detaillierten Deskription, Analyse und Modellierung fremdsprachlicher Akzente, wie sie in dieser Arbeit durchgeführt wurde, letztlich nur ein eng umgrenzter Teilaspekt des gesamten Themenkomplexes erfaßt.



10 Literatur²⁹

- Altenberg, E. P.; Vago, R. M. (1987): "Theoretical implications of an error analysis of second language phonology production". In: Ioup, G.; Weinberger, S. H. (eds.): *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, 148-164.
- Amdal, I.; Korkmazskiy, F.; Surendran, A. (2000): "Joint pronunciation modelling of non-native speakers using data-driven methods". *Proceedings International Conference on Spoken Language Processing (ICSLP 2000)*, Peking, *.
- Barry, W. J. (1989): "Perception and production of English vowels by German learners: instrumental-phonetic support in language teaching". *Phonetica* 46, 155-168.
- Barry, W. J.; Fourcin, A. J. (1992): "Levels of labelling". *Computer Speech and Language* 6, 1-14.
- Bausch, K.-R.; Kasper, G. (1979): „Der Zweitsprachenerwerb: Möglichkeiten und Grenzen der ‚großen‘ Hypothesen“. *Linguistische Berichte* 64, 3-35.
- Beebe, L. (1987): "Myths about interlanguage phonology". In: Ioup, G.; Weinberger, S. H. (eds.): *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, 165-175.
- Belhoula, A. (1996): *Ein regelbasiertes Verfahren zur maschinellen Graphem-nach-Phonem-Umsetzung von Eigennamen in der Sprachsynthese*. Düsseldorf: VDI Verlag.
- Belhoula, A.; Kugler, M.; Krüger, R.; Rühl, H.-W. (1997): "Evaluation of a TTS-system intended for the synthesis of names". In: van Santen, J. P. H.; Sproat, R.; Olive, J.; Hirschberg, J. (eds.): *Progress in Speech Synthesis*. Berlin, Heidelberg, New York: Springer, 529-540.
- Benarousse, L.; Geoffrois, E.; Grieco, J.; Series, R.; Steeneken, H.; Stumpf, H.; Swail, C.; Thiel, D. (2001): "The NATO native and non-native (N4) speech corpus". *Proceedings Workshop on Multilingual Speech and Language Processing*, Aalborg, *.
- Best, C. (1994): "The emergence of native-language phonological influences in infants: A perceptual assimilation model". In: Goodman, J.; Nusbaum, H.C. (eds.): *The Development of Speech Perception*. Cambridge, Mass.: MIT Press, 167-224.

²⁹ Anmerkung zum Literaturverzeichnis: Einige zitierte Texte lagen nicht in gedruckter Fassung, sondern nur in elektronischer Form vor (z.B. Konferenzbände auf CD-ROM oder einzelne im Internet verfügbare Artikel, Vorabdrucke etc.). Bei diesen Texten liegen zum Teil keine Seitenangaben vor. Ist dies der Fall, so sind diese Texte anstelle einer Seitenangabe mit dem Symbol * gekennzeichnet.

- Beulen, K.; Ortman, S.; Eiden, A.; Martin, S.; Welling, L.; Overmann, J.; Ney H.: (1998): "Pronunciation modelling in the RWTH large vocabulary speech recognizer". *Proceedings Modeling Pronunciation Variation for Automatic Speech Recognition*, Rolduc, *.
- Bierwisch, M. (1972): „Schriftstruktur und Phonologie“. *Probleme und Ergebnisse der Psychologie* 43, 21-44.
- Bird, S. (2002): "Phonology". In: Mitkov, R. (ed.): *Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press. 3-24.
- Boas, F. (1889): "On alternating sounds". *American Anthropologist* 2.1. (1889), 47-53.
- Böhm, A. (1993): *Maschinelle Sprachausgabe deutschen und englischen Textes*. Aachen: Shaker (zugl.: Dissertation Universität Bochum, Institut für Allgemeine Elektrotechnik und Akustik).
- Bohn, O.-S.; Flege, J. E. (1990): "Perception and production of a new vowel category by adult second language learners". In: *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech*. Amsterdam: University of Amsterdam Press, 37-56.
- Bongaerts T.; van Summeren C.; Planken B.; Schils E. (1997): "Age and ultimate attainment in the pronunciation of foreign languages". *Studies in Second Language Acquisition* 19/4, 447-465.
- Boysson-Bardies, D. de; Sagart, L.; Durand, C. (1984): "Discernible differences in the babbling of infants according to target language". *Journal of Child Language* 11, 1-15.
- Boysson-Bardies, D. de; Sagart, L.; Halle, P.; Durand, C. (1986): "Acoustic investigations of cross-language variability in babbling". In: Lindblom, B; Zetterström, R. (eds.): *Precursors to Early Speech*. New York: Stockton Press, 113-126.
- Boysson-Bardies, D. de; Halle, P.; Sagart, L.; Durand, C. (1989): "A cross-linguistic investigation of vowel formants in babbling". *Journal of Child Language* 8, 511-524.
- Brière, E. J. (1968): *A Psycholinguistic Study of Phonological Interference*. The Hague, Paris: Mouton.
- Brown, G.; Anderson, A.; Shillcock, R.; Yule, G. (1984): *Teaching Talk*. Cambridge: Cambridge University Press.
- Bühler, K. (1934/82): *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Stuttgart: Fischer [Ungekürzter Nachdruck der Ausgabe Jena 1934].
- Busch, D. (1982): "Introversion-extroversion and the EFL proficiency of Japanese students". *Language Learning* 32 (1), 109-132
- Byrne, W.; Knodt, E.; Khudanpur, S.; Bernstein, J. (1998): "Is automatic speech recognition ready for non-native speech? A data collection effort and initial experiments in modeling conversational hispanic English". *Proceedings ESCA Workshop Speech Technology in Language Learning (STiLL)*, Marholmen, Sweden, *.

- Chambers, J. K.; Trudgill, P. (1980): *Dialectology*. Cambridge: Cambridge University Press.
- Chomsky, N. (1957): *Syntactic Structures*. Den Haag: Mouton.
- Chomsky, N. (1959): "Review of Skinner's *Verbal Behavior*." *Language* 35 (1959), 325-339.
- Chomsky, N. (1986): *Knowledge of Language: Its Nature, Origin, and Use*. New York u.a.: Praeger.
- Chomsky, N.; Halle, M. (1968): *The Sound Pattern of English*. New York: Harper & Row.
- Church, K. (2003): "Speech and language processing: Where have we been and where are we going". *Proceedings Eurospeech 2003*, Genf, 1-4.
- Coltheart, M. (1978): "Lexical access in simple reading tasks". In: Underwood, G. (ed.): *Strategies of Information Processing*. New York: Academic Press, 151-216.
- Corder, S. P. (1967/74): "The significance of learners' errors". *International Review of Applied Linguistics (IRAL)* 4 (1967) 161-170. Nachgedr. in und zitiert nach: Richards, J. C. (ed.): *Error analysis: Perspectives on Second Language Acquisition*. London: Longman, 19-27.
- Cremelie, M.; Martens, J. P. (1995): "On the use of pronunciation rules for improved word recognition". *Proceedings Eurospeech 1995*, Madrid, 1747-1750.
- Cremelie, N.; ten Bosch, L. (2001): "Improving the recognition of foreign names and non-native speech by combining multiple grapheme-to-phoneme converters". *Proceedings ISCA ITRW Workshop Adaptation Methods for Speech Recognition*, Sophia Antipolis, France, *.
- Cucchiari, C.; Strik, H. (2003): "Automatic phonetic transcription: An overview. *Proceedings 15th International Conference of Phonetic Sciences (ICPhS 2003)*, Barcelona, 347-350.
- Dahlbäck, N.; Swamy, S.; Nass, C.; Arvidsson, F.; and Skågeby, J. (2001): "Spoken interaction with computers in a native or non-native language - same or different?" *Proceedings 8th TC.13 IFIP International Conference on Human-Computer Interaction (INTERACT 2001)*, Tokyo, *.
- Damper, R. I.; Marchand, Y.; Adamson M. J; Gustafson, K: (1999): "Evaluating the pronunciation component of text-to-speech systems for English: A performance comparison of different approaches". *Computer Speech and Language* 13, 155-176.
- Dickerson, L. J. (1975): "The learner's interlanguage as a system of variable rules". *TESOL Quarterly*, Vol. 9, No. 4, 401-408.
- Draxler, C. (2000): "Speech databases". In: van Eynde, F.; Gibbon, D. (eds.): *Lexicon Development for Speech and Language Processing*. Dordrecht: Kluwer, 169-206.
- Dulay, H. C.; Burt, M. K. (1973): "Should we teach children syntax?" *Language Learning* 23, 245-258.

- Dulay, H. C.; Burt, M. K. (1974): "You can't learn without goofing: An analysis of children's second language 'errors'". In: Richards, J. C. (ed.): *Error analysis: Perspectives on Second Language Acquisition*. London: Longman, 95-123.
- Dutoit, T. (1997): *An Introduction to Text-to-Speech Synthesis*. Dordrecht u.a.: Kluwer.
- Eklund, R.; Lindström, A. (1999): "How foreign are 'foreign' speech sounds? Implications for speech recognition and speech synthesis". Proceedings Workshop *Multilingual Interoperability in Speech Technology (MIST)*, Leusden, The Netherlands, 1999, *.
- Eklund, R.; Lindström, A. (2001): "Xenophones: An investigation of phone set expansion in Swedish and implications for speech recognition and speech synthesis". *Speech Communication* 35 (1-2), 81-102.
- Ferreiros, J.; Macías-Guarasa, J.; Pardo, J.M.; Villarrubia, L. (1998): "Introducing multiple pronunciations in Spanish speech recognition systems." Proceedings Workshop *Modeling Pronunciation Variation for Automatic Speech Recognition*, Kerkrade, 1998, 29-34.
- Fitt, S. (1995): "The pronunciation of unfamiliar native and non-native town names". *Proceedings Eurospeech 1995*, Madrid, Vol. 3, 2227-30.
- Fitt, S. (1997): "The generation of regional pronunciations of English for speech synthesis". *Proceedings Eurospeech 1997*, 2447-50.
- Fitt, S. (1998): *Processing Unfamiliar Words. A Study in the Perception and Production of Native and Foreign Place Names*. Ph.D. Thesis, University of Edinburgh.
- Fitt, S.; Isard, S. (1998): "Representing the environments for phonological processes in an accent-independent lexicon for synthesis of English". *Proceedings International Conference on Spoken Language Processing (ICSLP 1998)*, *.
- Fitt, S.; Isard, S. (1999): "Synthesis of regional English using a keyword lexicon". *Proceedings Eurospeech 1999*, Budapest, 823-826.
- Flach, G. (1995): "Modelling pronunciation variability for special domains". *Proceedings Eurospeech 1995*, Madrid, 1743-46.
- Flege, J. E. (1980): "Phonetic approximation in second language acquisition". *Language Learning* 30, 117-134.
- Flege, J. E. (1981): "The phonological basis of foreign accent". *TESOL Quarterly* 15, 443-455.
- Flege, J. E. (1987): "Effects of equivalence classification on the production of foreign language speech sounds". In: James, A; Leather, J. (eds.): *Sound Patterns in Second Language Acquisition*. Dordrecht: Foris, 9-39.
- Flege, J. E. (1988a): "The production and perception of foreign language speech sounds". In: Winitz, H. (ed.): *Human Communication and Its Disorders. A Review*. Volume I. Norwood, N.J.: Ablex, 224-401.
- Flege, J. E. (1988b): "Factors affecting degree of perceived foreign accent in English sentences". *Journal of the Acoustical Society of America* 84, 1, 70-79.

- Flege, J. E. (1992): "Speech learning in a second language". In: Ferguson, C.; Menn, L.; Stoel-Gammon, C. (eds.): *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press, 565-604.
- Flege, J. E. (1995): "Second language speech learning: Theory, findings and problems". In: Strange, W. (ed.): *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore, MD: York Press. 233-277.
- Flege, J. E. (1997): "English vowel production by Dutch talkers". In: James, A.; Leather, J. (eds.): *Second-Language Speech. Structure and Process*. Berlin, New York: Mouton de Gruyter, 11-52.
- Flege, J. E.; Hillenbrand, J. (1987): "Limits on phonetic accuracy in foreign language speech production". In: Ioup, G.; Weinberger, S. H. (eds.): *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, 176-203.
- Fleischer, W. (1992): „Onomastische Strukturen in der deutschen Sprache der Gegenwart“. In: W. Fleischer: *Name und Text: Ausgewählte Studien zur Onomastik und Stilistik*. Hrsg. v. I. Barz. Tübingen: Niemeyer, 43-57.
- Fries, C. C. (1945): *Teaching and Learning English as a Foreign Language*. Ann Arbor. University of Michigan Press.
- Fröhlich, W. D. (1997): *Wörterbuch Psychologie*. 21. bearb. u. erw. Aufl. (1. Aufl. 1968). München: dtv.
- Furui, S. (2001): "From read speech recognition to spontaneous speech understanding". *Proceedings 6th Natural Language Processing Pacific Rim Symposium*, Tokyo, 19-25.
- Gass, S. M.; Selinker, L. (eds.) (1992): *Language Transfer in Language Learning*. Philadelphia, PA: John Benjamins.
- Gerosa, M; Giuliani, D. (2004): "Preliminary investigations in automatic recognition of English sentences uttered by Italian children". *Proceedings of NLP and Speech Technologies in Advanced Language Learning Systems Symposium*, Venice, 9-12.
- Gibbon, D.; Moore, R.; Winski R. (1997): *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter.
- Gimson, A.C. (1980): *An Introduction to the Pronunciation of English*. Third Edition, London: Edward Arnold Publishers.
- Glück, H. (1993): „Die geschriebene Sprachform im Fremdspracherwerb. Am Beispiel einiger Probleme ägyptischer Germanistikstudenten mit dem Vokalismus und den Wortakzentstrukturen im Deutschen“. In: Baurmann, J.; Günther, H.; Knoop, U. (Hrsg.): *homo scribens*, Tübingen: Niemeyer, 319-340
- Goldsmith, J. (1976): *Autosegmental phonology*. Doctoral thesis, Massachusetts Institute of Technology, Cambridge/MA [publiziert als: Goldsmith, J. (1979): *Autosegmental phonology*. New York: Garland Press].

- Guiora, A.; Beit-Hallahmi, B.; Brannon, R.; Dull, C.; Scovel, T. (1972): "The effects of experimentally induced changes in ego states on pronunciation ability in a second language: An exploratory study". *Comprehensive Psychiatry* 13, 421-428.
- Gustafson, J. (1994): "ONOMASTICA - Creating a multi-lingual dictionary of European names". *FONETIK '94, Papers from the 8th Swedish Phonetics Conference*, Lund, Sweden, 66-69.
- Hammarström, G. (1966): *Linguistische Einheiten im Rahmen der modernen Sprachwissenschaft*. Berlin, Heidelberg, New York: Springer.
- Haugen, E. (1953): *The Norwegian Language in America*. Vols. I and II. Philadelphia: University of Pennsylvania Press.
- Heeringa, W. J. (2004): *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Groningen: University Library Groningen.
- Heyde, A. (1979): *The Relationship between Self-Esteem and the Oral Production of a Second Language*. Doctoral dissertation. Ann Arbor: University of Michigan.
- Hirschfeld, U. (1983): „Ergebnisse des Sprachvergleichs Spanisch-Deutsch im Bereich der Phonologie und Phonetik“. In: *Deutsch als Fremdsprache* 3, 169-174.
- Hudson, R. (1980): *Sociolinguistics*. Cambridge: Cambridge University Press.
- Ikeno, A.; Pellom, B.; Cer, D.; Thornton, A.; Brenier, J.; Jurafsky, D.; Ward, W.; Byrne, W. (2003): "Issues in recognition of Spanish-accented spontaneous English". *Proceedings ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, Tokyo, April 2003, *.
- International Phonetic Association (1999): *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Ioup, G.; Weinberger, S. H. (eds.) (1987): *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Cambridge, MA: Newbury House Publishers.
- Jakobson, R.; Halle, M. (1956): *Fundamentals of Language*. Den Haag: Mouton.
- James, C. (1971): "The exculpation of contrastive linguistics". In: Nickel, G. (ed.): *Papers in Contrastive Linguistics*. Cambridge: Cambridge University Press, 53-68.
- James, A.; Kettemann, B. (eds.) (1983): *Dialektphonologie und Fremdspracherwerb*. Tübingen: Narr.
- Jamieson, D. G. (1995): "Techniques for training difficult non-native speech contrasts". *Proceedings 13th International Congress of Phonetic Sciences (ICPhS 95)*, Stockholm, Vol. 4, 100-107.
- Jannedy, S; Möbius, B. (1997): "Name pronunciation in German text-to-speech synthesis". *Proceedings Fifth Conference on Applied Natural Language Processing*, Association for Computational Linguistics, Washington/D.C., 49-56.

- Jilka, M. (2000): *The Contribution of Intonation to the Perception of Foreign Accent. Identifying Intonational Deviations by Means of F0 Generation and Resynthesis*. Dissertation, Universität Stuttgart.
- Johansson, F. A. (1973): *Immigrant Swedish Phonology: A Study in Multiple Contact Analysis*. Lund: CWK Gleerup.
- Jurafsky, D; Martin, J. (2000): *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.
- Kadmon, N. (ed.) (1998): *Glossary of Toponymic Terminology*. Version 4. Working Paper E/CONF.91/L.13, *Seventh UN Conference on the Standardization of Geographical Names*, New York, 1998, *.
- Kaplan, R. M.; Kay, M. (1994): "Regular models of phonological rule systems." *Computational Linguistics* 20 (3), 331-378.
- Kemp, T. (1996): „Regelbasiert generierte Aussprachevarianten für Spontansprache“. *Proceedings KONVENS 96*, Bielefeld. Mouton de Gruyter, *.
- Kenstowicz, M.; Kisseberth, Ch. (1979): *Generative Phonology. Description and Theory*. New York: Academic Press.
- Kenstowicz, M. (1994): *Phonology in Generative Grammar*. Oxford: Blackwell Publications.
- Kessens, J. M.; Strik, H. (2001): "Lower WERs do not guarantee better transcriptions". *Proceedings Eurospeech 2001*, Aalborg, Denmark, Vol. 3, 1721-1724.
- Kim, H.-K. (2000): *The Interlanguage Phonology of Korean Learners of English: A Computational Implementation of Optimality Theoretic Constraints*. Ph.D. Dissertation, Georgetown University. Washington, DC.
- Kiparsky, P. (1973): "Elsewhere in phonology". In: Anderson, S.R.; Kiparsky, P. (eds.): *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston, 93-106.
- Klein, H.-W. (1973): *Phonetik und Phonologie des heutigen Französisch*. 4. durchges. Aufl., München: Hueber.
- Kondrak, G. (2003): "Phonetic alignment and similarity". *Computers and the Humanities* 37, 273-291.
- Koß, G. (2002): *Namenforschung. Eine Einführung in die Onomastik*. 3. aktualis. Auflage. Tübingen: Niemeyer.
- Krashen, S. D. (1981): *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
- Kruskal, J. (1983): "An overview of sequence comparison". In: Sankoff, D. & Kruskal, J. (eds.): *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*. Reading/Mass.: Addison-Wesley, 1-44.

- Kuhl, P. K.; Iverson, P. (1995): "Linguistic experience and the 'perceptual magnet effect'". In: Strange, W. (ed.): *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*. Timonium, MD: York Press, 121-154.
- Ladefoged, P. (1975): *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
- Ladefoged, P. (1999): "Linguistic phonetic descriptions". In: W.J. Hardcastle; J. Laver (eds.): *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 589-618.
- Lado, R. (1957): *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.
- Lamel, L. F.; Schiel F.; Fourcin, A.; Mariani, J.; Tillmann, H. G. (1994): "The Translanguage English Database (TED)". *Proceedings International Conference on Spoken Language Processing (ICSLP 1994)*, Yokohama, 1795-1798.
- Laver, J. (1994): *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Leather, Jonathan (1999): "Second language speech research: an introduction". In J. Leather (ed.): *Phonological Issues in Language Learning*. Oxford: Basil Blackwell, 1-58.
- Lehtinen, G.; Safra, S. (1998): "Generation and selection of pronunciation variants for a flexible word recognizer". *Proceedings Workshop Modeling Pronunciation Variation for Automatic Speech Recognition*, Kerkrade, 67-72.
- Lenneberg, E. (1972): *Biologische Grundlagen der Sprache*. Frankfurt/Main: Suhrkamp (Orig.: *Biological foundations of language*, New York, 1967).
- Levenshtein, V. I. (1966) : "Binary codes capable of correcting deletions, insertions and reversals". *Cybernetics and Control Theory* 10, 707-710.
- Lindau, M. (1985): "The story of /r/". In: Fromkin, V. (ed.): *Phonetic Linguistics. Essays in Honor of Peter Ladefoged*. Orlando u.a.: Academic Press, 157-168.
- Lindström, A.; Eklund, R. (1999): "How foreign are 'foreign' speech sounds? Implications for speech recognition and speech synthesis." *Proceedings Workshop Multilingual Interoperability in Speech Technology (MIST)*, Leusden, The Netherlands, 1999, *.
- Lippmann, R.P. (1997): "Speech recognition by humans and machines". *Speech Communication* 22 (1997), 1-15.
- Livescu, K.; Glass, J. (2000): "Lexical modeling of non-native speech for automatic speech recognition". *Proceedings ICASSP 2000*, Istanbul, *.
- Logan, J. S.; Lively, S. E.; Pisoni, D. B. (1991): "Training Japanese listeners to identify English /r/ and /l/: A first report". *Journal of the Acoustical Society of America* 89, 874-886.
- Long, M. H. (1990): "Maturation constraints on language development". *Studies in Second Language Acquisition* 12/3, 251-85.

- Love, N. (1981): *Generative Phonology. A Case-Study from French*. Amsterdam: John Benjamins. (Linguisticæ Investigationes, Supplementa 4).
- Lyons, J. (1977): *Semantics*. Volume 1. Cambridge: Cambridge University Press.
- Major, R. C. (1997): "L2 acquisition, L1 loss, and the critical period hypothesis". In: James, A. & Leather, J. (eds.): *Second-language Speech. Structure and Process*. Berlin, New York: Mouton de Gruyter, 147-159.
- Major, R. C. (2001): *Foreign Accent. The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum.
- Mangold, M. (1990): *DUDEN Aussprachewörterbuch. Wörterbuch der deutschen Standardaussprache*. 3. Aufl., Mannheim, Wien, Zürich: Dudenverlag.
- McAllister, R. (1997): "Perceptual foreign accent: L2 users' comprehension ability". In: James, A; Leather, J. (eds.): *Second-Language Speech. Structure and Process*. Berlin, New York: Mouton de Gruyter, 119-132.
- McLaughlin, B. (1989): "The three phases (faces?) of second language research". In: Dietrich, R.; Graumann, C.F. (eds.): *Language Processing in Social Context*. Amsterdam: Elsevier Publ., 211-231.
- Meinhold, G.; Stock, E. (1982): *Phonologie der deutschen Gegenwartssprache*. Zweite durchges. Aufl.. Leipzig: VEB Bibliographisches Institut.
- Mengel, A. (1993): "Transcribing names - a multiple choice task: mistakes, pitfalls and escape routes". *Onomastica Research Colloquium 1993*, London, *.
- Menzel, W.; Atwell E.; Bonaventura P.; Herron D.; Howarth P.; Morton R.; Souter C. (2000): "The ISLE corpus of non-native spoken English". *Proceedings 2nd International Conference on Language Resources and Evaluation (LREC 2000)*, Athen, Vol. 2, 957-963.
- Miller, C. A. (1998): *Pronunciation Modeling in Speech Synthesis*. Ph.D. Dissertation, University of Pennsylvania.
- Mohri, M. (1997): "Finite-state transducers in language and speech processing". *Computational Linguistics* 23 (2), 231-238.
- Munro, M. J.; Derwing, T. M.; Burgess, C. S. (2003): "The detection of foreign accent in backwards speech". *Proceedings 15th International Conference of Phonetic Sciences (ICPhS 03)*, Barcelona, 535-538.
- Nemser, W. (1971a): "Approximative systems of foreign language learners". *International Review of Applied Linguistics (IRAL)* 9, 115-123.
- Nemser, W. (1971b): "The predictability of interference phenomena in the English speech of native speakers of Hungarian". In: Nickel, G. (ed.): *Papers in Contrastive Linguistics*. Cambridge: Cambridge University Press, 89-102.
- Nemser, W. (1971c): *An experimental study of phonological interference in the English of Hungarians*. Bloomington: Indiana University.
- Nerbonne, J. (2003): "Linguistic variation and computation". *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2003)*, Budapest, 3-10.

- Nerbonne, J.; Heeringa, W.; Hout, E. van den; Kooi, P. van der; Otten, S.; Vis, W. van de (1996): "Phonetic distance between Dutch dialects". In: Durieux, G.; Daelemans, W.; Gillis, S. (eds.): *CLIN VI, Papers from the sixth CLIN Meeting*, Center for Dutch Language and Speech, University of Antwerp, 185-202.
- Neufeld, G. G. (1977): "Language learning ability in adults: A study on the acquisition of prosodic and articulatory features". *Working Papers in Bilingualism* 12, 4-60.
- Neufeld G. G. (2001): "Non-foreign-accented speech in adult second language learners: Does it exist and what does it signify?" *ITL Review of Applied Linguistics* 133/134, 185-206.
- Newton, B. (1972): *The Generative Interpretation of Dialect*. Cambridge: Cambridge University Press.
- Nickel, G. (ed.) (1971): *Papers in Contrastive Linguistics*. Cambridge: Cambridge University Press.
- Nikolov M. (2000): "The critical period hypothesis reconsidered: Successful adult learners of Hungarian and English". *International Review of Applied Linguistics (IRAL)* 38, 109-124.
- Onomastica Consortium (1995): "The Onomastica Interlanguage Pronunciation Lexicon." *Proceedings Eurospeech 1995*, Madrid, Vol. 1, 829-832.
- Paap, K. R.; Noel, R. W.; Johansen, L. S. (1992): "Dual-route models of print to sound: Red herrings and real horses". In: Frost, R.; Katz, L. (eds.): *Orthography, morphology, and meaning*. Amsterdam: North-Holland, 293-318.
- Paul, H. (1909): *Prinzipien der Sprachgeschichte*. 4. Aufl., Halle/Saale: Max Niemeyer.
- Petyt, K. M. (1980): *The Study of Dialect*. London: André Deutsch.
- Pfister, B.; Romsdorfer, H. (2003): "Mixed-lingual text analysis for polyglot TTS synthesis". *Proceedings Eurospeech 2003*, Genf, 2037-2040.
- Polivanov, E. (1931) : «La perception des sons d'une langue étrangère». *Travaux du Cercle Linguistique de Prague* 4; In: *Le Cercle de Prague* (Change, 3). Paris, 1969. 111-114.
- Pols, L. C. W. (1991): "Quality assessment of text-to-speech synthesis by rule". In: Furui, S.; Sondhi, M. M. (eds.): *Advances in speech signal processing*. New York: Marcel Dekker Inc., Kap. 13, 387-416.
- Prince, A.; Smolensky, P. (1993): *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms., Rutgers University and University of Colorado at Boulder.
- Rabiner, L. (1999): "Speech recognition in machines". In: R. Wilson & F. K. Keil (eds.): *MIT Encyclopedia of Cognitive Science*. Cambridge, MA: MIT Press, 785-787.
- Raux, A; Eskenazi, M. (2004): "Non-native users in the Let's Go!! spoken dialogue system: Dealing with linguistic mismatch". *HLT/NAACL 2004*, Boston, MA, *.

- Rein, K. (1983): *Einführung in die kontrastive Linguistik*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Rochet, B. L. (1995): "Perception and production of L2 speech sounds by adults." In: Strange, W. (ed.): *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*. Timonium, MD: York Press, 379-410.
- Rubach, J. (1984): "Rule typology and phonological interference." In: Eliasson, S. (ed.): *Theoretical issues in contrastive phonology*. Heidelberg: Groos, 37-50.
- Sanders, E.; van den Heuvel, H. (2001): "Speaker recruitment for speech databases". *Proceedings PRASA 2001*, Franschhoek, South Africa, 109-114.
- Schaden, S. (2001): „Spracherkennung bei akzentgefärbten Aussprachevarianten“. *Fortschrittsberichte der Akustik - DAGA 2001*. Oldenburg, 98-99.
- Schaden, S. (2002a): "A database for the analysis of cross-lingual pronunciation variants of European city names". *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Vol. 4, 1277-1283.
- Schaden, S. (2002b): „Regelbasierte Generierung fremdsprachlich akzentgefärbter Aussprachevarianten“. In: *Elektronische Sprachsignalverarbeitung*. Tagungsband der 13. Konferenz ESSV 2002. (*Studientexte zur Sprachkommunikation*, Band 24, Hrsg. v. Rüdiger Hoffmann), Dresden, 289-293.
- Schaden, S. (2003a): "Rule-based lexical modelling of foreign-accented pronunciation variants". *Proceedings of the Research Notes Session of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL '03)*. Budapest, Conference Companion Volume, 159-162.
- Schaden, S. (2003b): "Generating non-native pronunciation lexicons by phonological rules". *Proceedings 15th International Conference of Phonetic Sciences (ICPhS 2003)*, Barcelona, 2545-2548.
- Schaden, S. (2003c): "Non-native pronunciation variants of city names as a problem for speech technology applications". In: *Text, Speech and Dialogue. Proceedings of the Sixth International Conference TSD 2003* (České Budějovice, Czech Republic). Berlin, Heidelberg, New York: Springer [Lecture Notes in Artificial Intelligence 2807], 229-236.
- Schaden, S. (2004): "CrossTowns: Automatically generated phonetic lexicons of cross-lingual pronunciation variants of European city names". *Proceedings Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisboa, Portugal, 1395-1398.
- Schane, S. A. (1968): *French phonology and morphology*. Cambridge/Mass.: MIT Press.
- Schlobinski, P. (1996): *Empirische Sprachwissenschaft*. Opladen: Westdeutscher Verlag.
- Schultz, T; Waibel, A. (2001): "Language independent and language adaptive acoustic modeling for speech recognition". *Speech Communication* 35 (1-2) 31-51.

- Scovel, T. (1969): "Foreign accents, language acquisition and cerebral dominance". *Language Learning* 19, 245-254.
- Scovel, T. (1988): *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech*. New York: Newbury House/Harper & Row.
- Scovel, T. (2000): "A critical review of the critical period research". *Annual Review of Applied Linguistics* 20, 213-223.
- Seidenberg, M.S. (1999): "Visual word recognition". In: R. Wilson & F. K. Keil (eds.): *MIT Encyclopedia of Cognitive Science*. Cambridge, MA: MIT Press, 869-871.
- Selinker, L. (1972): "Interlanguage". *International Review of Applied Linguistics (IRAL)* 10, 209-231. Nachgedr. in und zitiert nach: Richards, J.C. (ed.) (1974): *Error Analysis: Perspectives on Second Language Acquisition*. London: Longman, 31-54.
- Sheldon, A.; Strange, W. (1982): "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception". *Applied Psycholinguistics* 3, 243-261.
- Singh, R.; Ford, A. (1986): "Interphonology and phonological theory." In: James, A.; Leather, J. (eds.): *Sound Patterns in Second Language Acquisition*. Dordrecht: Foris, 163-172.
- Snow, C. (1988): "Relevance of the notion of a critical period to language acquisition". In: Bornstein, M. (ed.): *Sensitive Periods in Development: Interdisciplinary Perspectives*. Hillsdale, NJ: Erlbaum, 183-210.
- Somers, H. (1998): "Similarity metrics for aligning children's articulation data". *Proceedings 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL '98)*, Montreal, 1227-1232.
- Stølen, M. (1987): "The effect of affect on interlanguage phonology". In: Ioup, G.; Weinberger, S. H. (eds.): *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Cambridge, MA: Newbury House Publishers, 389-400.
- Strik, H. (2001): "Pronunciation adaptation at the lexical level". *Proceedings ISCA Workshop 'Adaptation Methods for Speech Recognition'*, Sophia Antipolis, France, 123-131.
- Strik, H.; Cucchiaroni, C. (1999): "Modeling pronunciation variation for ASR: overview and comparison of methods". *Speech Communication* 29, 225-246.
- Szulc, A. (1973): „Die Haupttypen der phonischen Interferenz“. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 26, 111-119.
- Tarone, E. (1978): "The phonology of interlanguage". In: Richards, J. (ed.): *Understanding Second and Foreign Language Learning*. Rowley, MA, Newbury House, 15-33.
- Teixeira, C.; Trancoso, I.; Serralheiro, A. (1997): "Recognition of non-native accents". *Proceedings Eurospeech 1997*, Rhodos, 2375-2378.

- Ternes, E. (1976): *Probleme der kontrastiven Phonetik*. Hamburg: Buske Verlag (Forum Phonetikum 13).
- Ternes, E. (1999): *Einführung in die Phonologie*. 2. Auflage (1. Aufl. 1987). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Tesch, G. (1978): *Linguale Interferenz. Theoretische, terminologische und methodische Grundfragen zu ihrer Erforschung*. Tübingen: Narr.
- Tomokiyo, L. M.; Burger, S. (1999): "Eliciting natural speech from non-native users: Collecting speech data for LVCSR". *Proceedings ACL-IALL Joint Workshop on Computer Mediated Language Assessment and Evaluation in NLP*, Maryland, *.
- Tomokiyo, L. M. (2000a): "Handling non-native speech in LVCSR: A preliminary Study". *Proceedings EUROCALL/CALICO/ISCA Workshop Integrating Speech Technology in Language Learning (INSTIL)*, Dundee, *.
- Tomokiyo, L. M. (2000b): "Acoustic and lexical modeling of non-native speech in LVCSR". *Proceedings International Conference on Spoken Language Processing (ICSLP 2000)*, Peking, *.
- Trancoso, I.; Viana, M.C. (1995): "Issues in the pronunciation of proper names: The experience of the ONOMASTICA project". *Proceedings Workshop on Integration of Language and Speech*, Moskau, *.
- Trancoso, I.; Viana, M.C.; Mascarenhas, M.I.; Teixeira, C.J. (1999): "On deriving rules for nativised pronunciation in navigation queries". *Proceedings Eurospeech 1999*, Budapest, Vol. 1, 195-198.
- Trubetzkoy, N. S. (1939/89): *Grundzüge der Phonologie*. Göttingen: 1989 [Originalausgabe: *Travaux du Cercle Linguistique de Prague* 7 (1939)].
- Uebler, U.; Boros, M. (1999): "Recognition of non-native German speech with multilingual recognizers". *Proceedings Eurospeech 1999*, Budapest, *.
- Uebler, U. (2000): *Multilingual Speech Recognition*. Berlin: Logos-Verlag.
- Ungeheuer, G. (1968/72): „Kommunikative und extrakommunikative Betrachtungsweisen in der Phonetik“. Bonn, 1968 (IPK-FB 68/4). Nachgedr. in: Ungeheuer, G.: *Sprache und Kommunikation* (Forschungsberichte des IKP 13), Hamburg: Buske, 1972, 37-50.
- Ungeheuer, G. (1969/77): „Das Phonemsystem der deutschen Hochlautung“. In: Siebs, T. (1969): *Deutsche Aussprache. Reine und gemäßigte Hochlautung mit Aussprachewörterbuch*. 19. Aufl., Berlin, 27-42. Nachgedruckt in: Ungeheuer, G. (1977): *Materialien zur Phonetik des Deutschen* (Forschungsberichte des IKP 61). Hamburg: Buske, 63-78.
- van Bael, C.P. J.; Binnenpoorte, D.; Strik, H.; van den Heuvel, H. (2003): "Validation of phonetic transcriptions based on recognition performance". *Proceedings Eurospeech 2003*, Genf, 1545-1548.
- van Compernelle, D. (1999): "Speech recognition by goats, wolves, sheep and non-natives". *Proceedings Workshop Interoperability in Speech Technology (MIST)*. Leusden, Niederlande, 1999, *.

- van den Heuvel, H. (1994): "Pronunciation of foreign names by Dutch grapheme-to-phoneme converters". *Proceedings of the Department of Language and Speech*, University of Nijmegen, 99-105.
- van Leeuwen, D. A.; Orr, R. (1999): "Speech recognition of non-native speech using native and non-native acoustic models". *Proceedings Workshop Multilingual Interoperability in Speech Technology (MIST)*, Leusden, The Netherlands, 1999, *.
- van Leeuwen, D. A.; Wijngaarden, S. (2000): "Automatic speech recognition of non-native speakers using consonant-vowel-consonant (CVC) words." *Proceedings International Conference on Spoken Language Processing (ICSLP 2000)*, Peking, *.
- Viergege, W. H.; Rietveld A. C. M.; Jansen, C.I.E. (1984): "A distinctive feature based system for the evaluation of segmental transcription in Dutch". *Proceedings of the 10th International Congress of Phonetic Sciences*. Dordrecht, 654-659.
- Wall, L. (2001): *Programming Perl*. Cambridge u.a.: O'Reilly & Associates.
- Wardhaugh, R. (1970): "The contrastive analysis hypothesis". *TESOL Quarterly* 4, 123-130.
- Warnant, L. (1987): *Dictionnaire de la prononciation française dans sa norme actuelle*. 4^{ième} édition. Paris: Duculot.
- Weinreich, U. (1953): *Languages in Contact: Findings and Problems*. Den Haag, Paris: Mouton.
- Weinreich, U. (1957): "On the description of phonic interference". *Word* 13 (1957), 1-11.
- Werker, J. F.; Tees, R. C. (1984): "Cross-language speech perception: Evidence for perceptual reorganisation during the first year of life". *Infant Behavioral Development* 7, 49-63.
- Wells, J. C. (1999): "Overcoming phonetic interference". *Speech, Hearing and Language: Work in progress*, Vol. 11, 117-128.
- Wells, J. C. (2001a): *Longman Pronunciation Dictionary*. Harlow: Longman.
- Wells, J. C. (2001b): "Orthographic diacritics and multilingual computing". *Language Problems and Language Planning* 24.3, *.
- Wells, J. C. (2004): *SAMPA. Computer Readable Phonetic Alphabet*.
<http://www.phon.ucl.ac.uk/home/sampa/home.htm>
 [Seiten werden in unregelmäßigen Abständen aktualisiert]
- Wesenick, M. B. (1996): "Automatic generation of German pronunciation variants". *Proceedings International Conference on Spoken Language Processing (ICSLP 1996)*, Philadelphia, 125-128.
- Wester, M.; Kessens, J. M. (1998): "Improving recognition performance by modelling pronunciation variation." *Proceedings Workshop Modeling Pronunciation Variation for Automatic Speech Recognition*, Kerkrade, 145-150.

- Wester, M.; Kessens, J. M.; Strik, H. (2000): "Pronunciation variation in ASR: Which variation to model?" *Proceedings International Conference on Spoken Language Processing (ICSLP 2000)*, Peking, *.
- Wester, M.; Fosler-Lussier, E. (2000): "A comparison of data-derived and knowledge-based modeling of pronunciation variation". *Proceedings International Conference on Spoken Language Processing (ICSLP 2000)*, Peking, *.
- Wiese, R. (2003): "The unity and variation of (German) /r/". *Zeitschrift für Dialektologie und Linguistik* 70 (2003), 25-43.
- Witt, S.; Young, S. (1999): "Off-line acoustic modelling of non-native accents". *Proceedings Eurospeech 1999*, Budapest, Vol. 3, 1367-1370.
- Yamada, R. A. (1993): "Effect of extended training on /r/ and /l/ identification by native speakers of Japanese". *Journal of the Acoustical Society of America* 4 (1993), 2391.
- Zue, V.; Seneff, S.; Polifroni, J.; Meng, H.; Glass, J. (1996): "Multilingual human-computer interactions: From information access to language learning". *Proceedings International Conference on Spoken Language Processing (ICSLP 1996)*, Philadelphia, *.

11 Notationskonventionen

I. Phonetisches Alphabet

Für phonetische Transkriptionen in dieser Arbeit wird das IPA-Alphabet verwendet (International Phonetic Association 1999). Eine Übersicht des IPA-Alphabets ist u.a. zu finden unter der Internet-Adresse

<http://www.arts.gla.ac.uk/IPA/ipachart.html>

Der Druck der phonetischen Zeichen erfolgte mit der beim *Summer Institute of Linguistics* (*www.sil.org*) frei verfügbaren Schrifttype **SIL Doulos IPA 93**.

In allen in dieser Arbeit beschriebenen Datensammlungen, Transkriptionen und Computerprogrammen wird dagegen zur phonetischen Umschrift das SAMPA- und X-SAMPA-Alphabet verwendet (Wells 2004). Die Definition von SAMPA für viele (europäische) Sprachen sowie das X-SAMPA-Alphabet finden sich auf der regelmäßig aktualisierten Seite

<http://www.phon.ucl.ac.uk/home/sampa/home.htm>

II. Phonemische, phonetische und graphemische Umschrift

/R/	phonemische Umschrift
[nɔ̃ʁt]	phonetische Umschrift
<pf>	Graphem/Graphemsequenz
∅	„Nullphonem“, Nullsegment: Nicht artikuliertes Phonem (meist als Folge der Tilgung eines Segments durch eine phonologische Regel)
<i>Haus</i>	Lexem, Beispielwort, -phrase oder -satz
/Haus/	Morphem
[±nasal]	binäres phonologisches Merkmal, noch unspezifiziert
[-nasal]	binäres phonologisches Merkmal, negativ spezifiziert
[+obstruent]	binäres phonologisches Merkmal, positiv spezifiziert
$\left[\begin{array}{l} +\text{consonantal} \\ +\text{lateral} \\ +\text{alveolar} \end{array} \right]$	phonologisches Merkmalsbündel

III. Notation phonologischer Regeln und Prozesse

$X \rightarrow Y$	Kontextfreie phonologische Regel: <i>Ersetze X durch Y</i>
$X \rightarrow Y / Z$	Kontextabhängige phonologische Regel: <i>Ersetze X durch Y im Kontext Z</i>
—	Platzhalter für das Eingabesymbol in einer Kontextbedingung
$X \rightarrow Y / _ Z$	Kontext Z steht rechts vom Eingabesymbol X
$X \rightarrow Y / Z _$	Kontext Z steht links vom Eingabesymbol
#	In Kontextbedingungen: Morphemgrenze
##	In Kontextbedingungen: Wortgrenze
\$	In Kontextbedingungen: Silbengrenze
$\left\{ \begin{array}{c} C \\ \# \end{array} \right\}$	Alternative Phonemsegmente in Kontextbedingung
$\left\{ \begin{array}{c} C _ \# \\ V _ \end{array} \right\}$	Vollständige alternative Kontextbedingungen

IV. Sonstige Konventionen

$[ɔ] \sim [o]$	Phonetische Alternation zwischen zwei Lauten
EN \rightarrow FR	Sprachrichtung. Notation MUTTERSPRACHE \rightarrow ZIELSPRACHE z.B. <i>Englisch</i> \rightarrow <i>Deutsch</i>
$a \Rightarrow b$	unidirektionale Implikation: <i>a</i> impliziert <i>b</i>
$a \Leftrightarrow b$	bidirektionale Implikation: <i>a</i> impliziert <i>b</i> und <i>b</i> impliziert <i>a</i>