Lexikalische Modellierung fremdsprachlicher Akzente – Methoden, Möglichkeiten und Grenzen eines regelbasierten Ansatzes

Stefan Schaden

mail@stefanschaden.de

Abstract: Dieser Beitrag bietet eine zusammenfassende Übersicht eines mehrjährigen Forschungsprojekts, dessen Ziel darin lag, nicht-muttersprachliche Aussprachevarianten auf der Ebene der phonetischen Transkription mittels Regeln zu modellieren. Auf diese Weise wurde ein Ansatz zur automatisierten Erstellung phonetischer Lexika geschaffen, die zur Adaption sprachtechnologischer Systeme an nicht-muttersprachliche Nutzer eingesetzt werden können.

1 Einleitung

1.1 Hintergrund und Zielsetzung

Fremdsprachliche Akzente stellen bei der Entwicklung sprachtechnologischer Systeme oftmals eine besondere Herausforderung dar. Abweichungen von einer idealtypischen „Normaussprache“ sind hier oft besonders ausgeprägt, während gleichzeitig die auftretenden Aussprachevarianten in hohem Maße von Sprecher zu Sprecher variieren können. So sind bei der Modellierung von akzentbehafteter Aussprache (beispielsweise für Spracherkennungssysteme, die auch von Nicht-Muttersprachlern verschiedener Herkunft genutzt werden sollen) zahlreiche, oft nur schwer vorhersehbare phonetische Varianten zu berücksichtigen.

Das Ziel des vorgestellten Projekts bestand in der Entwicklung einer Methode, mit der prototypische akzentgefärbte Aussprachevarianten für verschiedene Sprachenkombinationen auf der Ebene der phonetischen Transkription mittels eines Regelsystems automatisch generiert werden können. Auf diese Weise wurden Verfahren der lexikalischen Adaption, die bisher u.a. zur Modellierung intra-lingualer Variation angewandt worden sind (für der Bereich der Spracherkennung siehe z.B. [7]), auf den Bereich sprachenübergreifender Variation ausgeweitet. Ein solches Regelsystem kann dazu eingesetzt werden, existierende phonetische Lexika um potentielle akzentgefärbte Varianten zu erweitern und so als Grundlage einer Adaption sprachtechnologischer Systeme an nicht-muttersprachliche Nutzer verwendet werden. Der Schwerpunkt der Arbeiten lag jedoch nicht auf der Optimierung eines spezifischen Systems, sondern in der Erarbeitung einer allgemeinen Methode, die in verschiedenen sprachtechnologischen Anwendungszusammenhängen zum Einsatz kommen kann. Die Untersuchungen wurden an sechs europäischen Sprachen in verschiedenen Sprachrichtungen, d.h. Kombinationen aus L1 (Muttersprache) und L2 (Zielsprache) durchgeführt. Regelsätze wurden bisher für die folgenden neun Sprachrichtungen erstellt (Notation *Ausgangssprache* → *Zielsprache*):

1. Englisch → Deutsch
2. Deutsch → Englisch
3. Französisch → Deutsch
4. Deutsch → Französisch
5. Englisch → Französisch
6. Französisch → Englisch
7. Italienisch → Deutsch
8. Deutsch → Italienisch
9. Italienisch → Englisch

Jeder Regelsatz enthält zwischen 50 und 110 Einzelregeln; der formale Aufbau des Regelsystems ist dabei für alle Sprachen einheitlich (Details hierzu siehe [6]). Die Methode und ihre

programmtechnische Implementierung sind so konzipiert, dass sie prinzipiell auf neue Sprachenkombinationen übertragbar sind.

1.2 Datenbasis

Eine Modellierung von Aussprachevarianten jeglicher Art setzt voraus, dass zunächst die phonetischen Besonderheiten der zu modellierenden Varianten bekannt sind. Für akzentgefärbte Aussprache existiert jedoch aus naheliegenden Gründen keine phonetische Referenzform, die als „idealer Output“ eines Regelsystems herangezogen werden könnte. Daher ist zunächst zu untersuchen, ob für bestimmte Sprachenkombinationen typische Aussprachevarianten vorherrschen und wodurch diese gekennzeichnet sind (z.B. bei Aussprache des Deutschen durch französische Muttersprachler). Zu diesem Zweck wurde eine Sprachdatenbank mit ca. 160 Sprechern fünf verschiedener Muttersprachen (*Deutsch, Englisch, Französisch, Italienisch, Spanisch*) erstellt. Aufgabe der Sprecher war es, in standardisierten Aufnahmesitzungen Material aus verschiedenen Zielsprachen (a) zu lesen und (b) einem Muttersprachler der Zielsprache nachzusprechen [6].

Das Sprachmaterial bestand aus Ortsnamen, die den Schwerpunkt der Studie bildeten, sowie kurzen, syntaktisch einfach gebauten Sätzen. Die untersuchten Zielsprachen waren *Deutsch*, *Englisch*, *Französisch*, *Italienisch* und *Niederländisch*. Die Datenbank weist einen Gesamtumfang von ca. 65.000 Äußerungen (isolierte Wörter und Sätze) bzw. ca. 25 Stunden Audiomaterial auf. Für einen Teil des Materials (ca. 13.500 Äußerungen) wurden unter Verwendung des SAMPA- und X-SAMPA-Alphabets manuelle phonetische Transkriptionen erstellt. Diese erfüllten eine zweifache Funktion: Zunächst dienten sie der Analyse der auftretenden Aussprachevarianten auf symbolisch-phonetischer Ebene. Durch eine Klassifizierung und Gruppierung gleicher oder ähnlicher Varianten konnte das transkribierte Material genutzt werden, um Aussprachevarianten zu identifizieren, die sprecherübergreifend auftreten und somit Eingang in das Regelsystem finden. Darüber hinaus erfüllten die Transkriptionen in einem späteren Stadium eine wichtige Funktion bei der Evaluierung des Regelsystems (siehe unten Abschnitt 4).

2 Variation und Akzentstufen

Selbst wenn man sich vorübergehend auf die Untersuchung von Sprechern einer einzigen Herkunftssprache beschränkt, liegt ein wesentliches Problem bei der Systematisierung nichtmuttersprachlicher Aussprachevarianten in ihrer hohen interindividuellen Variabilität. Während beispielsweise dialektale Varianten oft durch relativ systematische Abweichungen zur hochsprachlichen Aussprachenorm gekennzeichnet sind, sind fremdsprachliche Akzente lautlich äußerst heterogen. Für eine symbolisch-phonetische Modellierung der Varianten bedeutet dies u.a., dass das Spektrum potentieller Varianten mit nur *einer* alternativen Aussprache pro zielsprachlichem Wort nicht hinreichend abgedeckt werden kann[1]. Andererseits kann es jedoch kaum als ein realistisches Ziel erscheinen, bei einer symbolischen Modellierung *jede* potentielle Aussprachevariante zu erfassen.

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 stark akzentgefärbten Aussprache erstreckt. Dieses Kontinuum wurde in diskrete *Akzentstufen* unterteilt, mit denen die auf verschiedenen Kompetenzstufen auftretenden Varianten beschrieben werden können. Jede dieser Akzentstufen ist durch prototypische Aussprachevarianten repräsentiert.

[1] Dieser Ansatz wurde beispielsweise in [3] verfolgt.

Dieser Ansatz ist angelehnt an die in [2] entworfene Konzeption von Lernersprachen als *approximative Systeme*. Dabei handelt es sich um vorübergehende, jedoch temporär stabile Sprachsysteme, die sich die Lernenden von der Zielsprache aufgebaut haben. Es wird dabei weiterhin angenommen, dass die approximativen Systeme von Lernenden, die auf gleicher Lernstufe stehen, Ähnlichkeiten aufweisen (prototypische approximative Systeme). Das folgende (fiktive) Beispiel zeigt eine solche Akzentstufung für den englischen Ortsnamen *Appleton Thorn*, gesprochen von Muttersprachlern des Deutschen. Ausgehend von einer akzentfreien Aussprache (Stufe 0) repräsentiert dabei jede der Stufen 1 bis 4 eine zunehmend stärkere Akzentfärbung (im Sinne von stärkeren und zusätzlichen Abweichungen von der L2-Zielform):

Akzentstufe	**Variante**
Stufe 0	[æpɫtən θɔːn]
Stufe 1	[ɛpltən θɔn̪]
Stufe 2	[ɛpltɔn θɔn̪]
Stufe 3	[ɛpltɔn sɔn̪]
Stufe 4	[aplətɔn tɔn̪]

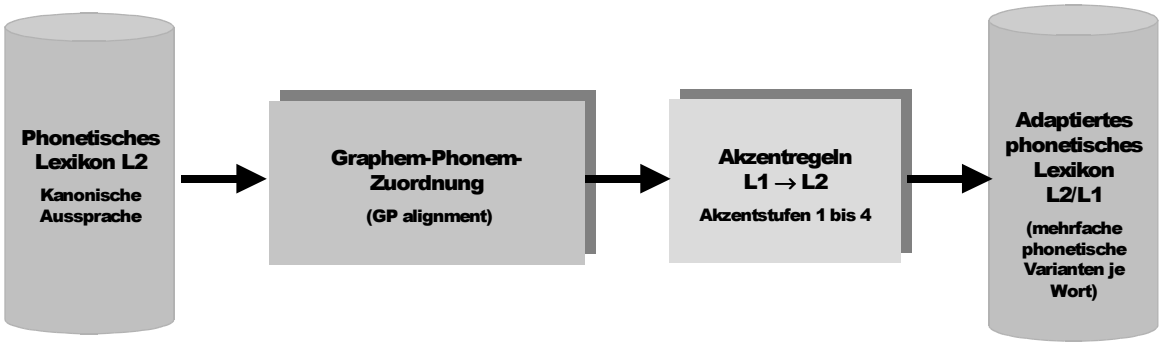
Das Modell der Akzentstufen diente im vorliegenden Projekt als Grundlage des Regelsystems zur Generierung von Varianten. Es ist so konzipiert, dass es je Eingabewort vier alternative Aussprachevarianten mit zunehmender Akzentstärke generiert. Diese Stufung ist inkrementell angelegt, d.h., höhere Akzentstufen schließen die auf den niedrigeren Akzentstufen modellierten Fehler stets ein. Dies ist nicht nur technisch – im Sinne eines modularen Aufbaus des Regelsystems – sinnvoll, sondern entspricht durchaus der sprachlichen Realität, da ein Sprecher mit einem stark ausgeprägten Akzent oftmals alle Fehlertypen zugleich produziert, während mit zunehmender L2-Kenntnis die Fehler der hohen Akzentstufen schrittweise eliminiert werden und nur die Fehler der niedrigen Stufen (möglicherweise dauerhaft) verbleiben. Die Anzahl von vier Stufen ist willkürlich gewählt und kann prinzipiell reduziert oder erhöht werden. Sie hat sich aber als sinnvoll erwiesen, um einerseits wichtige Fehlertypen mit erkennbaren phonetischen Unterschieden voneinander abzugrenzen, dabei aber die Variantenanzahl überschaubar zu halten. Zu erwähnen ist an dieser Stelle noch, dass nicht zwingend für jeden Lexikoneintrag vier voneinander verschiedene Varianten generiert werden. So kann es in Einzelfällen dazu kommen, dass Regeln einer bestimmten Akzentstufe nicht auf ein Wort anwendbar sind und daher die auf dieser Stufe erzeugte phonetische Variante identisch mit derjenigen der vorhergehenden Stufe ist.

3 Regelsystem

3.1 Allgemeiner Aufbau des Systems

In [4] wurde ein Prototyp des hier vorgestellten Regelsystems skizziert, der auf der Idee beruht, nicht-muttersprachliche Aussprachevarianten mittels eines erweiterten Graphem-nach-Phonem-Umsetzers aus einer orthographischen Eingabe zu generieren. Dieses Konzept wurde jedoch später abgelöst durch ein System, welches eine bereits existierende kanonische Transkription nachträglich modifiziert und mit spezifischen Merkmalen akzentbehafteter Aussprache versieht. Die Varianten werden somit nicht mehr aus einer *orthographischen*, sondern einer *phonetischen* Eingabe generiert. Dieser Ansatz bietet u.a. den Vorteil der Anwendbarkeit auf bereits existierende phonetische Lexika. Dennoch hat sich erwiesen, dass die orthographi-

sche Information für die Erzeugung der Varianten weiterhin benötigt wird (siehe unten 3.2). Um einen Zugriff auf die orthographische Ebene zu ermöglichen, wurde den eigentlichen Akzent-Regeln ein Modul vorangestellt, welches zunächst die orthographische und phonetische Sequenz, aus denen sich ein Lexikoneintrag zusammensetzt, zueinander synchronisiert (Graphem-Phonem-Alignment). Die folgende Abbildung zeigt den Gesamtaufbau des Systems:



3.2 Regeltypen

Die Basis des vorgestellten Systems bilden phonetische Ersetzungsregeln der allgemeinen Form X→Y, bei denen X einen zielsprachlichen Laut (oder eine Lautverbindung) und Y einen Ersatzlaut repräsentiert, der für X eintritt. Diese Regeln können kontextunabhängig formuliert werden (eine Regel gilt uneingeschränkt für jedes Vorkommen von X) oder mit Kontextbedingungen versehen werden (eine Regel gilt nur dann, wenn X in einem bestimmten phonetischen Umfeld auftritt). Kontexte können nicht nur als phonetische Segmente, sondern auch in Form von Phonemklassen (Vokale, Konsonanten, Frikative, Plosive etc.) angegeben werden, um eine kompaktere Regelformulierung zu erzielen. Dies ermöglicht es auch, systematische phonologische Regeln einer Sprache (etwa die deutsche „Auslautverhärtung“) und deren Transfer auf die Zielsprache auf ökonomische Weise zu formulieren.

Ein großer Teil der formulierten Regeln enthält Kontextbedingungen des oben genannten Typs. Darüber hinaus existiert ein weiterer Regeltyp, bei dem nicht die phonetische Umgebung des zu ersetzenden Lauts, sondern vielmehr dessen *orthographische Repräsentation* als Kontextbedingung angegeben ist. Zur Formulierung solcher Regeln kann die oben erwähnte Zuordnung von Graphem- und Phonem-Ebene (GP-Alignment) herangezogen werden, die es ermöglicht, Lautsubstitutionen an bestimmte graphemische Kontexte zu binden. Solche *graphemabhängigen Lautersetzungen* wurden eingeführt, um der Tatsache Rechnung zu tragen, dass eine Reihe von Ersatzlauten nur dann eintritt, wenn der ersetzte Laut durch bestimmte Grapheme repräsentiert wird. Beispielsweise ist ein Ersatz des deutschen [v] durch [w] bei englischen Muttersprachlern nur dann zu beobachten, wenn [v] orthographisch durch <w> repräsentiert ist. Darüber hinaus kann in bestimmten Fällen der in Frage kommende Ersatzlaut nicht bestimmt werden, wenn die orthographische Form des betreffenden Wortes unbekannt ist. So ist bei der Zielsprache Englisch häufig zu beobachten, dass reduzierte Vokale wie in *Aston* [æstən] von nicht-muttersprachlichen Sprechern als Vollvokale realisiert werden, woraus Varianten wie [ɛstɔn] resultieren. Eine einfache, nur auf der phonetischen Ebene operierende Ersetzungsregel [ə] → [ɔ] könnte den Sachverhalt, dass der Ersatzlaut [ɔ] hier durch das orthographische <o> motiviert ist, nicht adäquat erfassen und würde darüber hinaus zahlreiche abwegige Formen wie beispielsweise [pɛkɔm] für engl. *Peckham* [pɛkəm] generieren.

Bei den zuletzt beschriebenen Aussprachevarianten handelt es sich um partielle Übertragungen der muttersprachlichen Graphem-Phonem-Zuordnungen auf die Zielsprache. Sie treten

vorwiegend dann auf, wenn das fremdsprachliche Material gelesen wird[2]. Die Sprachdaten zeigen, dass ein solcher Transfer der GTP-Beziehungen in den wenigsten Fällen *vollständig* auf zielsprachliche Wörter angewendet wird (z.B. engl. *Shrewsbury*, das nach deutschen Regeln als [sʁɛfsbuːri] realisiert würde). Weitaus häufiger tritt der Fall auf, dass zielsprachliche Graphemsequenzen nur teilweise nach muttersprachlichen Regeln realisiert werden, während die verbleibenden Graphemfolgen annähernd den zielsprachlichen GTP-Regeln folgen. Dies erschwert prinzipiell die Modellierung von Fehlern, da die resultierenden Aussprachevarianten sowohl Elemente der muttersprachlichen als auch der zielsprachlichen Graphem-Phonem-Beziehungen enthalten. Der Regeltyp der graphemabhängigen Lautersetzungen erwies sich jedoch in vielen Fällen als geeignet, um solche Mischformen zu generieren. Falls erforderlich, kann er auch zur Generierung von Varianten genutzt werden, die fast vollständig den GTP-Regeln der Muttersprache des Sprechers entsprechen.

4 Evaluierung

4.1 Verfahren

Da diese Studie, wie bereits eingangs erwähnt, nicht unmittelbar auf eine Systemoptimierung abzielte, wurde auch die Evaluierung des Ansatzes von einem konkreten Anwendungszusammenhang abgekoppelt. Sie erfolgte auf der symbolisch-phonetischen Ebene, ohne Aussagen darüber zu treffen, ob und in welcher Weise ein bestimmtes sprachtechnologisches System vom Einsatz der Regeln profitieren kann. Ziel der Evaluierung war es vielmehr, den Grad der Approximation der automatisch generierten Varianten an tatsächlich auftretende Aussprachevarianten zunächst allein auf der Ebene der phonetischen Transkription zu ermitteln. Im Mittelpunkt stand dabei die Frage, ob die Regeln geeignet sind, auch Aussprachevarianten *unbekannter* Sprecher – also solcher Sprecher, deren Aussprache nicht selbst in die Erstellung der Regeln eingegangen ist – plausibel zu modellieren[3].

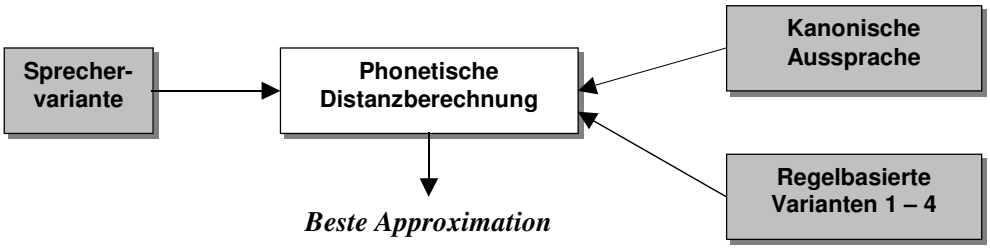
Den Kern der Evaluierung bildet ein Modul zur Berechnung der Distanz zwischen zwei Phonemsequenzen. Es basiert auf der *Edit-Distanz* (auch *Levenshtein-Distanz),* welche die Ähnlichkeit zwischen zwei Zeichenketten A und B auf der Basis der Anzahl von Substitutionen, Einfügungen und Löschungen ermittelt, die erforderlich sind, um A in B zu konvertieren (vgl. [1]). Da jedoch die Edit-Distanz in ihrer unmodifizierten Form phonetische Ähnlichkeiten zwischen den ersetzten Segmenten nicht berücksichtigt (es wird für jede Substitution ein konstanter Wert angenommen), wurde der Algorithmus um eine phonetische Gewichtung erweitert, welche die Ähnlichkeit der Segmente in die Distanzberechnung einbezieht. Eine solche Modifikation ist für einen aussagekräftigen Vergleich phonetischer Sequenzen notwendig, da beispielsweise eine Substitution des Vokals [eː] durch [ɛː] als weniger gravierend einzustufen ist als die Substitution desselben Vokals durch [iː] oder [aː].

Durch einen solchen Algorithmus kann nun ermittelt werden, ob die automatisch generierten Aussprachevarianten eine im Vergleich zur kanonischen Transkription höhere Approximation an die tatsächlich in einer Sprechergruppe auftretenden Varianten erzielen können. Die folgende Abbildung illustriert das Verfahren:

---

[2] Allerdings werden solche Fehler oftmals auch vorübergehend oder dauerhaft in die gesprochene Sprache übernommen, so dass eine Klassifizierung als reiner „Lesefehler“ nicht sinnvoll erscheint.

[3] Eine weitere zentrale Frage, die hier zugunsten einer gestrafften Darstellung nicht erörtert wird, ist die nach der *vokabularübergreifenden* Gültigkeit der Regeln, d.h. ihrer Anwendbarkeit auf bisher unbekanntes Vokabular. Siehe jedoch hierzu die Bemerkungen am Ende des Abschnitts 4.2.



Wie oben dargelegt, werden vier verschiedene Aussprachevarianten je Eingabewort generiert. Liegen nun phonetische Transkriptionen der Aussprachevarianten von Sprechern vor, so können diese einerseits mit der kanonischen Transkription und andererseits mit den automatisch generierten Transkriptionen verglichen werden. Als optimale Approximation gilt dabei jeweils diejenige Variante, die die geringste phonetische Distanz (bzw. höchste Ähnlichkeit) zur Aussprache der Sprecher aufweist. Diese wird entweder (i) durch die kanonische Form oder (ii) durch eine der vier regelbasierten Varianten erreicht. Das folgende Beispiel illustriert dies anhand des englischen Ortsnamens *Ashford*, bei dem die phonetischen Ähnlichkeiten zwischen einer Sprechervariante, der kanonischen Aussprache und den regelbasierten Varianten der Akzentstufen 1 bis 4 berechnet wurden[4]:

Sprechervariante:	[ɛʃfɔt]	**Ähnlichkeit**[5] (0 <= N <= 1)	
kanonisch:	[æʃfəd]	0,47	
Stufe 1:	[ɛʃfət]	0,85	
Stufe 2:	[ɛʃfɔ̯ʁ̯t]	0,90	⇐ beste Approximation
Stufe 3:	[aʃfɔ̯ʁ̯t]	0,74	
Stufe 4:	[aʃfɔ̯ʁ̯t]	0,74	

4.2 Ergebnisse

In der vollständigen Evaluierungsprozedur wurde die phonetische Approximation stets für (a) eine spezifische Sprechergruppe und (b) ein spezifisches Test-Vokabular ermittelt. Die dabei akkumulierten Werte geben an, mit welcher relativen Häufigkeit die kanonische Transkription bzw. eine der vier regelbasierten Varianten die beste Approximation erzielt haben. Da eine detaillierte Darstellung der Ergebnisse jedoch schon aufgrund der Vielzahl der untersuchten Sprachrichtungen den Umfang dieses Beitrags sprengen würde, sind im Folgenden lediglich einige wichtige Tendenzen am Beispiel der Zielsprachen Deutsch und Englisch zusammengefasst (eine ausführliche Darstellung findet sich in [6]).

Aus der Sprachdatensammlung wurden zunächst zwei Sprechergruppen verschiedener Muttersprachen als Testsets ausgewählt, deren Aussprachevarianten nicht zuvor in die Erstellung der Regeln eingeflossen waren. Während Gruppe A hinsichtlich ihres durchschnittlichen Alters (~25 Jahre) und ihrer relativ guten Kenntnisse der Zielsprachen vergleichbar mit den Sprechern war, die auch zur Erstellung der Regeln herangezogen wurden, handelte es sich bei der zweiten Gruppe B um Sprecher einer anderen Altersgruppe (45–65 Jahre) mit nur geringen durchschnittlichen Kenntnissen der Zielsprachen.

---

[4] Das Beispiel illustriert auch den in Abschnitt 2 erläuterten Umstand, dass die vier Akzentstufen nicht immer voneinander verschieden sein müssen. Im vorliegenden Fall sind die Stufen 3 und 4 identisch.

[5] Der Distanzwert wurde hier in einen *Ähnlichkeitswert* transformiert. Dabei bezeichnet der Wert 1.0 eine maximale Ähnlichkeit (= Identität) der Phonemsequenzen. Niedigere Werte kennzeichnen eine geringere Ähnlichkeit; der (in der Praxis nicht erreichbare) Wert 0 würde eine maximale Differenz bedeuten.

Für Gruppe A ist zu beobachten, dass unter den insgesamt fünf Varianten (kanonische Transkription sowie automatisch generierte Akzentstufen 1 bis 4) die beste Approximation bei einigen Sprachrichtungen (etwa Englisch → Deutsch) durch Akzentstufe 1, bei anderen dagegen (z.B. Französisch → Englisch) durch die kanonische Form erzielt wird. Dabei liegen die Werte sowohl für die kanonische Form als auch für Akzentstufe 1 jeweils im Bereich von ca. 30–40%. Es kann daher für Sprechergruppe A kein eindeutiger, für alle Sprachrichtungen gültiger Trend benannt werden. Akzentstufe 2 erzielt bei einigen Sprachrichtungen (z.B. Italienisch → Englisch) noch einen relativ hohen Anteil von ca. 25%, während die höheren Akzentstufen 3 und 4 in dieser Sprechergruppe keine nennenswerte Rolle mehr spielen. Da es sich aber bei Gruppe A um Sprecher mit relativ guten Kenntnissen der Zielsprachen handelt, steht dies durchaus im Einklang mit dem Annahmen des Modells.

In Gruppe B erhöht sich dagegen der Gesamtanteil der automatisch generierten Varianten an der besten Approximation und verschiebt sich zugleich zugunsten höherer Akzentstufen. Dabei sind jedoch auch hier z.T. auffallende Unterschiede zwischen den untersuchten Sprachrichtungen zu verzeichnen. So ergab sich für die Sprachrichtung Italienisch → Deutsch ein deutlicher Vorteil der Akzentstufen 2 und 3 gegenüber der kanonischen Transkription, während dies für die Richtung Italienisch → Englisch nicht zutraf. Diese Diskrepanz kann unterschiedlich interpretiert werden. Sie kann einerseits ein Indiz dafür sein, dass die Qualität der sprachenspezifischen Regelsätze sich noch voneinander unterscheidet und daher einzelne Regelsätze weiter zu optimieren sind. Es ist jedoch auch denkbar, dass die in den Sprachdaten vorliegenden Varianten bei einzelnen Sprachrichtungen besonders heterogen sind und nur wenig sprecherübergreifende Regularität aufweisen, so dass diese Sprachrichtungen durch einen regelbasierten Ansatz generell schwieriger zu erfassen sind. Weiterhin muss auch in Betracht gezogen werden, dass allgemeine L2-Kenntnisse und Akkuratheit der L2-Aussprache nicht immer unmittelbar miteinander korrelieren und möglicherweise bei einzelnen Sprachrichtungen trotz geringer formaler L2-Kenntnisse der Sprecher eine durchaus adäquate zielsprachliche Aussprache erreicht wird.

Zusätzlich zu der hier skizzierten Evaluierung wurden die Regeln auch erprobt, indem sie auf phonetische Lexika von je 1000 neuen Ortsnamen der Sprachen Englisch, Deutsch und Französisch angewendet wurden [5]. Obwohl für diese Korpora keine formale Evaluierung durchgeführt wurde, vermitteln sie einen Eindruck davon, welche Ergebnisse bei der Anwendung der Regeln auf neues, umfangreicheres Vokabular zu erwarten sind[6]. Derartige Testkorpora eignen sich u.a. dazu, unplausible regelbasierte Aussprachevarianten zu identifizieren und die Regeln entsprechend zu modifizieren. Hierin ist ein wichtiger Ansatzpunkt für eine weitere, ergänzende Evaluierungsstrategie zu sehen.

5 Ausblick

Die vorgestellte Studie zielte darauf ab, fremdsprachliche Akzente auf der Ebene des phonetischen Lexikons zu modellieren. Einsatzmöglichkeiten eines solchen Verfahrens liegen auf dem Gebiet der automatischen Spracherkennung, der Sprachsynthese und in angrenzenden Gebieten, von denen einige sich möglicherweise erst durch zukünftige Einsatzbereiche von Sprachtechnologien erschließen werden. Ob und in welchem Maße dabei einzelne Anwendungen von der dargestellten Methode profitieren können, ist letztlich von der konkreten Applikation abhängig. Beispielsweise wird im Fall der automatischen Spracherkennung ein Teil der auftretenden Aussprachevariation bereits durch akustische Modelle abgedeckt, so dass sich hier eine Erweiterung des Lexikons um Varianten mit nur geringer Abweichung von der ka-

[6] Diese Lexika können auf Anfrage zur Verfügung gestellt werden.

nonischen Form (etwa die beschriebene Akzentstufe 1) unter Umständen als redundant erweisen kann. Dagegen kann die Erkennung von Sprache mit stark ausgeprägtem Akzent möglicherweise von einer lexikalischen Modellierung profitieren. So bietet die vorgestellte Methode der graphemabhängigen Lautersetzungen (siehe 3.2) die Möglichkeit, auch solche Aussprachevarianten zu modellieren, bei denen es durch den Einfluss der Orthographie zu gravierenden Phonemsprüngen kommt, die durch akustische Modellierung nur schwierig zu erfassen sind. Bei einer möglichen Anwendung in der Sprachsynthese wiederum werden zwar unter Umständen nicht sämtliche Varianten zugleich benötigt; dennoch kann das Modell der Akzentstufen hier von Nutzen sein, da mit ihm die Stärke des zu erzeugenden Akzents vom Systementwickler vorab bestimmt werden kann.

Grenzen der vorgestellten Methode werden zweifellos durch die hohe sprecherübergreifende Variabilität akzentbehafteter Aussprache gesetzt. Wenn auch eine Gruppierung nach Sprachrichtungen sowie die Systematisierung phonetischer Varianten durch das Modell der Akzentstufen eine Eingrenzung der Vielfalt ermöglicht, so verbleibt im Bereich akzentbehafteter Aussprache dennoch ein erhebliches Variationsspektrum, das sich einer genauen Vorhersage oft entzieht. Sprecherspezifische Merkmale wie Alter, Bildung, L2-Kenntnisse, Beginn und Art des L2-Erwerbs sowie situationsspezifische Rahmenbedingungen und psychologische Dispositionen wie Gesprächstyp, Motivation, Verhältnis zum Gesprächspartner oder Grad der Aufmerksamkeit des Sprechers sind nur einige der zahlreichen Variablen, die letztlich auf die lautliche Beschaffenheit der L2-Produktion einwirken können. Es ist eine offene Frage, in welchem Maße es möglich sein wird, dieses komplexe Ursachengefüge so weit zu systematisieren, dass sprachtechnologische Anwendungen davon profitieren können.

6 Danksagung

Diese Arbeit entstand mit Unterstützung der Deutschen Forschungsgemeinschaft (DFG) am Institut für Kommunikationsakustik der Ruhr-Universität Bochum (Projektleitung: Prof. Dr. Ute Jekosch).

Literatur

[1] 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.

[2] Nemser, W. (1971): “Approximative systems of foreign language learners”. *International Review of Applied Linguistics (IRAL)* 9, 115–123.

[3] Onomastica Consortium (1995): “The Onomastica Interlanguage Pronunciation Lexicon.” *Proc. Eurospeech 1995*, Madrid, Vol. 1, 829–832.

[4] Schaden, S. (2002): „Regelbasierte Generierung fremdsprachlich akzentgefärbter Aussprachevarianten“. In: Hoffmann, R. (Hrsg.): *Elektronische Sprachsignalverarbeitung. Tagungsband der 13. Konferenz ESSV 2002*. Dresden: w.e.b. Universitätsverlag, 289–293.

[5] Schaden, S. (2004): “CrossTowns: Automatically generated phonetic lexicons of cross-lingual pronunciation variants of European city names.” *Proc. Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lissabon, 1395–1398.

[6] Schaden, S. (2007): *Fremdsprachlicher Akzent als Problem der Sprachtechnologie. Grundlagen, Entwurf und Implementierung eines Regelsystems zur Modellierung akzentgefärbter Aussprache*. Aachen: Shaker Media.

[7] Strik, H. (2001): “Pronunciation adaptation at the lexical level”. *Proc. ISCA Workshop ‘Adaptation Methods for Speech Recognition’*, Sophia Antipolis, 123–131.