

# Ein erweiterter Graphem-nach-Phonem-Umsetzer zur Modellierung nicht-muttersprachlicher Aussprachevarianten

Stefan Schaden

Institut für Kommunikationsakustik (IKA), Ruhr-Universität Bochum, D-44780 Bochum  
schaden@ika.ruhr-uni-bochum.de

## ZUSAMMENFASSUNG

Vorgestellt wird der Prototyp eines regelbasierten Graphem-nach-Phonem-Umsetzers, der charakteristische Aussprachefehler von Nicht-Muttersprachlern auf der Ebene der phonetischen Transkription modelliert. Dies eröffnet die Möglichkeit, an bestimmte Sprechergruppen angepasste Aussprachelexika für Spracherkennungs- oder Sprachsynthesysteme automatisch zu generieren.

## 1 Einleitung

Abweichungen von der Norm-Aussprache, die wir als fremdsprachlichen ‚Akzent‘ wahrnehmen, treten auf, wenn Sprecher einer Ausgangssprache (L1), die nur über unvollständige Kenntnisse bzw. Fertigkeiten der Phonetik einer Zielsprache (L2) verfügen, sprachliches Material der Zielsprache artikulieren. Diese nur partielle Kenntnis manifestiert sich in charakteristischen lautlichen Erscheinungen wie z.B. der Substitution von zielsprachlichen Lauten durch muttersprachliche Äquivalente, dem Transfer phonologischer Regelmäßigkeiten der Muttersprache auf die Zielsprache, oder – im Falle gelesener Sprache – einer Übertragung der muttersprachlichen Schriftaussprache (Graphem-Phonem-Beziehungen) auf die Zielsprache.

Sprachtechnologische Relevanz gewinnt dieses Problem vor allem im Zusammenhang mit der automatischen Spracherkennung: So ist bei nicht-muttersprachlichen Sprechern oftmals ein deutliches Absinken der Erkennungsraten zu beobachten. Da eine Adaption der akustischen Modelle an diese Sprechergruppen mit beträchtlichem Aufwand verbunden ist – Voraussetzung für ein Neutraining der Modelle ist es, für jede mögliche Sprachrichtung zunächst Sprachdaten in großem Umfang zu sammeln – besteht eine mögliche Alternative in einer Adaption auf der Ebene des Lexikons. Hierbei werden dem Lexikon Aussprachevarianten hinzugefügt, die typischerweise in einer bestimmten Sprechergruppe auftreten. Goronzy et al. (2001) konnten auf diese Weise für die Zielsprache Englisch die Erkennungsleistung bei italienischen und deutschen Muttersprachlern verbessern. Da jedoch bei umfangreichen Lexika eine manuelle Erstellung von Varianten sehr arbeitsintensiv ist, ist eine automatische Generierung der Varianten sinnvoll. Dies kann entweder (a) durch Umsetzung des orthographischen Eingabetextes in eine gezielt fehlerhafte phonetische Transkription oder (b) durch Umwandlung eines bereits vorliegenden Lexikons für die Standard-Aussprache in ein Varianten-Lexikon bzw. ‚Akzent-Lexikon‘ geschehen.

Auf der Basis der Textvorverarbeitungs-Komponenten des am IKA entwickelten Sprachausgabesystems SyRUB (Böhm 1993) wurde der Prototyp eines solchen erweiterten Graphem-nach-Phonem-Umsetzers erstellt. Es wurden Regelsätze der Ausgangs- und Zielsprache miteinander in einer Weise kombiniert, die es ermöglicht, verschiedene Übergangsstufen zwischen einer stark muttersprachlich geprägten, fehlerhaften Lautung und einer korrekten zielsprachlichen Lautung zu generieren. Die Regeln wurden an einer am IKA erstellten Sprachdatenbank mit Aufzeichnungen nicht-muttersprachlicher Sprecher verschiedener Herkunftssprachen gewonnen. Als Sprachmaterial dienten Ortsnamen sowie kurze Sätze aus den untersuchten Sprachen. Im Fokus der Arbeiten stehen derzeit die Sprachen Deutsch, Englisch und Französisch in verschiedenen L1/L2-Kombinationen.

## 2 Regeltypen

### 2.1 Standard-Graphem-nach-Phonem-Regeln

Grundlage des Graphem-nach-Phonem-Umsetzers bilden Regelsätze, mit denen zunächst die kanonische (d.h. Standard-) Aussprache der jeweiligen Zielsprache generiert werden kann. Entsprechende Regelsätze liegen derzeit für die Sprachen Deutsch sowie – in weniger ausgearbeiteter Form – für Englisch, Franzö-

sisch, Italienisch und Niederländisch vor. Eine möglichst fehlerfreie Generierung der kanonischen Aussprache durch die Standard-GTP-Regeln ist eine entscheidende Voraussetzung für die Zuverlässigkeit der nachfolgend beschriebenen Akzentregeln, da letztere die korrekte kanonische Form als Eingabe erwarten; Abweichungen davon haben zur Folge, daß die Akzentregeln nicht zur Anwendung kommen können. Da die bisher vorliegenden Standard-GTP-Regeln lediglich für das Deutsche eine zuverlässige Umsetzung gewährleisten, liegt der aktuelle Schwerpunkt der Arbeiten auf dem Deutschen als Zielsprache. Dabei wurden Akzentregeln bisher für die Ausgangssprachen Englisch und Französisch erstellt.

### 2.2 Phonem-Mapping

Ein wesentliches Merkmal fremdsprachlich akzentgefärbter Aussprachevarianten liegt darin, daß Sprecher dazu neigen, zielsprachliche Phoneme bzw. Allophone, die im Lautinventar ihrer Muttersprache nicht enthalten sind, durch muttersprachliche Äquivalente, d.h. ähnliche Laute aus L1 zu ersetzen.

Akzentvarianten, die auf phonemischen oder allophonischen Ersetzungen beruhen, werden im vorgestellten System durch die Anwendung von sprachübergreifenden Phonem-Abbildungsregeln (*mapping rules*) modelliert. Als Eingabe für diese Regelsätze dient die kanonische Aussprache von L2, welche entweder in Form eines statischen Referenzlexikons vorliegt oder automatisch nach den oben beschriebenen Standard-Regeln von L2 aus der orthographischen Eingabe generiert wird. Ausgabe der Abbildungsregeln ist eine in einzelnen Phonemen modifizierte Standard-Aussprache von L2. Hierbei sind verschiedene Regeltypen zu unterscheiden:

(a) *Kontextunabhängiges Phonem-Mapping*. Zum Teil ist es sinnvoll, Regeln für Lautersetzungen so zu formulieren, daß sie unabhängig von ihrem lautlichen Kontext angewandt werden. Dies gilt etwa beim Phonem /r/, dessen deutsche Realisierungsvarianten als uvularer Frikativ [R]<sup>1</sup> oder alveolarer Vibrant [r] von englischen Muttersprachlern durch den Approximanten [r̥] ersetzt werden. Dieser Lautersatz ist – soweit es die bisherigen Untersuchungen zeigen – unabhängig vom lautlichen Kontext und kann somit als kontextunabhängige Ersetzungsregel [R] → [r̥] formuliert werden.

(b) *Kontextabhängiges Phonem-Mapping*. In anderen Fällen hingegen ist die Wahl des Ersatzlautes abhängig vom lautlichen Kontext des ersetzten Lautes. Hier ist es erforderlich, Kontextbedingungen für die Anwendung einer Regel zu formulieren. So wird dt. [a:] von englischen Sprechern in einigen Kontexten durch den Diphthong [eɪ], in anderen Kontexten dagegen durch [A:] substituiert. Eine entsprechende Regel:

*Ersetze [a:] durch [eɪ], wenn ein Konsonant und mindestens ein Vokal folgen; ersetze [a:] durch [A:] in allen anderen Kontexten*

kann formal wie folgt wiedergegeben werden:

[a:] → [eɪ] / \_\_ CV  
[a:] → [A:] elsewhere

Dabei gilt, daß die jeweils spezifischere Regel stets vor der allgemeineren Regel mit der Bedingung *elsewhere* angewandt wird.

### 2.3 Graphemabhängiger Phonemersatz

In dem Fall, daß das fremdsprachliche Material gelesen wird, sind die vom Sprecher verwendeten Ersatzphoneme oftmals abhängig

<sup>1</sup> Im folgenden wird zur phonemischen/phonetischen Umschrift das SAMPA-Alphabet sowie dessen Erweiterung X-SAMPA verwendet (Wells 2000).

von der orthographischen Repräsentation des zu ersetzenden Lautes. So tritt z.B. ein Ersatz des labiodentalen Frikativs [v] durch den Approximanten [w] bei englischen Sprechern mit der Zielsprache Deutsch nur dann auf, wenn [v] orthographisch durch <w> repräsentiert ist, wie im folgenden Beispiel:

*Wernitz* → [w3:nIts] Anwendung der Regel [v] → [w]  
*Ravensberg* → [rA:y@nsb3:k] keine Anwendung der Regel

Zur Modellierung solcher Ersetzungen ist es erforderlich, daß die orthographische Eingabeinformation während des Umsetzungsprozesses nicht verloren geht, d.h., daß die Regeln für den Phonem-Ersatz auf Informationen der Graphem-Ebene zugreifen können. Im Bochumer Sprachausgabesystem SyRUB bleibt diese Information zu jedem Zeitpunkt der Umsetzung erhalten, so daß es möglich ist, Phonemsubstitutionen an bestimmte graphemische Kontexte zu binden. Die oben genannte Regel für den Ersatz von [v] durch [w] läßt sich dann wie folgt schreiben:

PHONEM-EBENE: [v] → [w]  
 |  
 GRAPHEM-EBENE: <w>

Die Verbindungslinie zwischen dem Eingabesymbol [v] und der Graphem-Ebene kennzeichnet die Bedingung, daß [v] orthographisch durch <w> repräsentiert sein muß.

#### 2.4 Transfer phonologischer Regeln

Ein Teil der bei Nicht-Muttersprachlern auftretenden Aussprachefehler kann als eine Übertragung phonologischer Regeln der Muttersprache auf die Zielsprache erklärt werden (Kenstowicz 1994: 59). Solche phonologischen Regeln betreffen oftmals nicht nur einzelne Phoneme, sondern Phonemklassen, welche in bestimmten Kontexten gleichartige systematische Veränderungen erfahren. Eine phonologische Regel des Deutschen besagt beispielsweise, daß stimmhafte Obstruenten in morphemfinaler Position stimmlos werden („Auslautverhärtung“). Tatsächlich liegt hierin ein charakteristisches Merkmal eines ‚deutschen Akzents‘, etwa bei der Aussprache des Englischen:

engl. *field* → [fi:ld] dt.-engl. [fi:lt]

Um eine Übertragung solcher systematisch auftretender Regeln auf die Zielsprache ohne zusätzliche Regelsätze modellieren zu können, ist die Graphem-nach-Phonem-Umsetzung für jede Sprache zweistufig angelegt. Die orthographische Eingabe wird zunächst in eine phonologische Basisrepräsentation umgewandelt. Sprachenspezifische phonologische Regeln („postlexikalische Regeln“) wandeln diese Basisrepräsentation schließlich in eine allophonische Repräsentation um. Durch diese Aufspaltung in zwei separate Regelmodule ist es möglich, sprachenspezifische phonologische Regeln aus L1 auch auf L2 anzuwenden, um damit einen fremdsprachlichen Akzent zu modellieren. Ein Beispiel soll dies verdeutlichen:

Bei einigen Sprechern des Französischen ist die Tendenz zu beobachten, die Nasalisierung von Vokalen in bestimmten Kontexten auf das Deutsche zu übertragen, etwa in *langsam* [la~za:m] oder *Frankfurt* [fra~kfURt]. Sie folgen damit einer phonologischen Regel des Französischen, die vereinfacht lautet: Ein Vokal wird nasalisiert, wenn er von einem nasalen Konsonanten und einer Wortgrenze oder einem weiteren Konsonanten gefolgt wird:

V [-nasal] → V [+nasal] / \_\_ C [+nasal] { ##, C }

Diese Regel ist Bestandteil der vorhandenen Standard-Regelsätze für Französisch. Wendet man sie auf das Deutsche an, erhält man – sofern der Kontext für die Anwendung der Regel in deutschen Wörtern erfüllt ist – eine modifizierte Transkription, die ein charakteristisches Merkmal eines französischen Akzentes enthält. Die Erstellung zusätzlicher Akzentregeln entfällt hier also, da die entsprechenden Varianten durch die Kombination existierender Regelmodule aus L1 und L2 generiert werden können.

#### 3 Kombination der Regelmodule

Die oben beschriebenen Regeltypen – Phonem-Mapping, phonologische Regeln und graphemabhängiger Phonemersatz – sind im hier vorgestellten System als separate Module implementiert, welche einzeln oder in Kombination aktiviert werden können. Dies ermöglicht es, für ein Eingabewort (bzw. Eingabelexikon) mehrere alternative Aussprachevarianten zu generieren. Obwohl eine ex-

akte Prognose einer von einem Sprecher produzierten Aussprachevariante kein realistisches Ziel ist, kann durch das Bereitstellen mehrerer Alternativen die Wahrscheinlichkeit einer Annäherung an die tatsächlich auftretende Aussprache erhöht werden.

Idealerweise sollten durch eine sequentielle Anwendung der einzelnen Regelmodule verschiedene Akzentstärken – von leichten Abweichungen von der Standard-Aussprache bis hin zu einer stark fehlerhaften Aussprache – modelliert werden können. Derzeit wird untersucht, mit welcher Anordnung bzw. Kombination der einzelnen Module dies erreicht werden kann. Dabei hat sich bisher gezeigt, daß leichte Abweichungen von der kanonischen Aussprache durch phonologische Regeln modelliert werden können, während das Phonem-Mapping und insbesondere der graphemabhängige Phonemersatz sich zur Modellierung stark muttersprachlich gefärbter Aussprachevarianten eignen. Darüber hinaus besteht die Möglichkeit, Wortschatz aus L2 vollständig nach den unmodifizierten Regeln von L1 umsetzen zu lassen, um so eine ‚naive‘ Schriftausssprache zu modellieren (Schaden 2001).

#### 4 Adaption an spezifische Phoneminventare

Der hier vorgestellte GTP-Umsetzer zielt darauf ab, eine phonetisch relativ detailreiche und möglichst realistische Repräsentation charakteristischer Aussprachevarianten nicht-muttersprachlicher Sprecher zu generieren. Als Lautinventar wird dabei das vollständige von SAMPA (und, sofern benötigt, X-SAMPA) abgedeckte Phonem- bzw. Allophoninventar der behandelten Sprachen genutzt. Dieses Inventar wird in den meisten Fällen nicht deckungsgleich mit systemspezifischen Lautinventaren von Spracherkennern oder Sprachsynthesystemen (akustische Modelle bzw. Sprachbausteine) sein. Es ist im Einzelfall zu entscheiden, welche phonetische Information relevant ist oder als redundant erachtet werden kann. Gegebenenfalls ist es erforderlich, die Ausgabe des Umsetzers mittels eines weiteren Regelmoduls auf spezifische Lautinventare zu reduzieren. Da der vorgestellte GTP-Umsetzer unabhängig von einem spezifischen System ist, sind solche zusätzlichen Regeln nicht enthalten.

#### 5 Zukünftige Arbeiten

Der beschriebene erweiterte GTP-Umsetzer befindet sich derzeit in einem prototypischen Stadium. Zukünftige Arbeiten umfassen eine Erweiterung der vorhandenen Regeln, eine Ausdehnung auf Varianten italienischer und niederländischer Sprecher sowie Untersuchungen zur Variation *innerhalb* der verschiedenen Sprechergruppen. Darüber hinaus werden derzeit Versuche zur Spracherkennung bei Nicht-Muttersprachlern durchgeführt. Hierbei soll vor allem ermittelt werden, welche Modifikationen des Lexikons sich positiv auf die Erkennungsleistung auswirken. Die Ergebnisse werden in die Weiterentwicklung des hier vorgestellten Systems einfließen.

#### 6 Literatur

- Böhm, A. (1993): *Maschinelle Sprachausgabe deutschen und englischen Textes*. Aachen: Shaker.
- Goronzy, S./Sahakyan, M./Wokurek, W. (2001): "Is non-native pronunciation modelling necessary?" Proceedings Eurospeech 2001, Aalborg, Denmark, 309-312.
- Kenstowicz, M. (1994): *Phonology in Generative Grammar*. Cambridge/MA, Oxford: Blackwell.
- Schaden, S. (2001): „Spracherkennung bei akzentgefärbten Aussprachevarianten.“ Fortschrittsberichte DAGA 2001, Hamburg.
- Wells, J.C. (2000): *SAMPA. Computer Readable Phonetic Alphabet*. <http://www.phon.ucl.ac.uk/home/sampa/home.htm>

Diese Arbeit entstand mit der Unterstützung der Deutschen Forschungsgemeinschaft (DFG) unter der Betreuung von PD Dr. Ute Jekosch am Institut für Kommunikationsakustik der Ruhr-Universität Bochum.