

Fay, Johanna; Berkling, Kay; Stüker, Sebastian

Automatische Analyse von Rechtschreibfähigkeit auf Basis von Speech-Processing-Technologien

Didaktik Deutsch : Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur 17 (2012) 33, S. 14-36



Empfohlene Zitierung/ Suggested Citation:

Fay, Johanna; Berkling, Kay; Stüker, Sebastian: Automatische Analyse von Rechtschreibfähigkeit auf Basis von Speech-Processing-Technologien - In: Didaktik Deutsch : Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur 17 (2012) 33, S. 14-36 - URN: urn:nbn:de:0111-pedocs-211893 - <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-211893>

in Kooperation mit / in cooperation with:



<http://www.paedagogik.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft



Didaktik Deutsch

2012

33

Mit Forschungsbeiträgen von

Johanna Fay · Kay Berking · Sebastian Stöcker
Christopher Sappok

Debatte
Forschungsbeiträge
Berichte und Ankündigungen
Rezensionen und Neuerscheinungen

Bibliographischer Hinweis sowie Verlagsrechte bei den online-Versionen der DD-Beiträge:



**Halbjahresschrift für die Didaktik
der deutschen Sprache und
Literatur**

<http://www.didaktik-deutsch.de>
17. Jahrgang 2012 – ISSN 1431-4355
Schneider Verlag Hohengehren
GmbH

*Johanna Fay, Kay Berkling &
Sebastian Stüker*

**AUTOMATISCHE ANALYSE VON
RECHTSCHREIBFÄHIGKEIT AUF
BASIS VON SPEECH-
PROCESSING-TECHNOLOGIEN**

In: Didaktik Deutsch. Jg. 17. H. 33. S. 14-36.

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. – Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden.

Inhalt

DEBATTE

Albert Bremerich-Vos

Ein kurzer Kommentar zum Beitrag „Grammatische Terminologie
in der Schule“ 5

Reinold Funke

Mehr als eine Terminologieliste 8

Eduard Haueis

Ein Auftakt 11

FORSCHUNGSBEITRÄGE

Johanna Fay, Kay Berkling & Sebastian Stüker

Automatische Analyse von Rechtschreibfähigkeit auf Basis von
Speech-Processing-Technologien..... 14

Christopher Sappok

Wann wird eine Kommposition als solche wahrgenommen?
Eine Untersuchung in Jgst. 6 mit einem zu kommatierenden Fremdtext..... 38

BERICHTE & ANKÜNDIGUNGEN

Maik Philipp

Wirksame Schreibförderung. Metaanalytische Befunde im Überblick 59

Christoph Bräuer & Iris Winkler

Aktuelle Forschung zu Deutschlehrkräften. Ein Überblick 74

REZENSIONEN UND NEUERSCHEINUNGEN

Ulf Abraham

Literaturdidaktik heute – zwei Neuerscheinungen 92

Neuerscheinungen 98

Abstract

The interdisciplinary research presented in this paper introduces a prototype for an automatic mechanism of analyzing and classifying spelling errors in freely written text by learners of the German writing system. Know-how from didactics and computer linguistics and automatic speech processing is combined to close the gap between economic and detailed analysis of spelling by automating the process. Large amounts of data can now be processed and analyzed without additional effort. This innovative approach is based on the connection between writing and pronunciation in addition to the usual study of grapheme-sequences. Know-how from automatic speech recognition and synthesis is leveraged to derive an alignment between grapheme and phoneme for both incorrect and correct spelling. Using this alignment, a detailed spectrum of error-types is detected and classified. This paper will detail the system setup and then proceed to apply it to data consisting of 120 texts collected from elementary school kids in grades 1 through 4. A comparison of hand-labeled and automatic procedure in terms of accuracy is carried out, showing that the approach is viable. Finally, the necessary steps to obtain a fully interconnected version are discussed.

Zusammenfassung

Der vorliegende Beitrag stellt ein interdisziplinäres Forschungsprojekt zur Entwicklung eines Instruments zur automatisierten Rechtschreibanalyse in frei verfassten Lernertexten vor. Mit diesem Instrument kann dem bekannten Dilemma der ‚ökonomischen vs. differenzierten Rechtschreibdiagnostik‘ begegnet werden, da binnen kürzester Zeit große Datenmengen orthographisch detailliert analysiert werden können. Der innovative Ansatz basiert auf der Analyse der Lernerschriften unter Einbezug ihrer automatisch generierten Aussprache. Erkenntnisse der automatischen Spracherkennung und -synthese erlauben eine automatisierte Gegenüberstellung aus fehlerhaftem Text und einer auf Basis von assoziierten Wahrscheinlichkeiten ermittelten korrekten Version des Textes. Rechtschreibfehler und Richtigschreibungen können dann automatisch annotiert und klassifiziert werden. Das Instrument wird hier zunächst in seiner Anlage erklärt, dann werden die Ergebnisse aus der Anwendung auf 120 Lernertexte aus Kl. 1 bis 4 vorgestellt. Der Vergleich von automatischer und manueller Analyse zeigt die Machbarkeit dieses Ansatzes sowie die Wege, die begangen werden müssen, um zu einem vollständig autonom agierenden Verfahren zu gelangen.

Johanna Fay, Kay Berkling & Sebastian Stüker

AUTOMATISCHE ANALYSE VON RECHTSCHREIBFÄHIGKEIT AUF BASIS VON SPEECH-PROCESSING-TECHNOLOGIEN

1 Einleitung

In der Diagnostik von Rechtschreibfähigkeit trifft man auf ein zunächst unlösbar scheinendes Dilemma: Verfahren, die eine Schreibproduktanalyse zur Grundlage haben, sind – verkürzt gesagt – entweder besonders ökonomisch in Durchführung und Auswertung oder aber besonders differenziert in der orthographischen Analyse. Beides in einem gibt es bislang nicht.

Ökonomische Verfahren liegen in der Diagnostik als standardisierte Rechtschreibtests wie bspw. die Hamburger Schreibprobe, HSP (May u. a. 2007), der Diagnostische Rechtschreibtest, DRT (Müller 2004), oder der Deutsche Rechtschreibtest, DERET (Stock/ Schneider 2008), vor. Sie erfüllen nach Autorenangaben die einschlägigen Gütekriterien psychologischer Tests (vgl. Lienert/ Raatz 1998, S. 7f.) und werden gerade aufgrund ihrer ausgeprägten Durchführungs- und Auswertungsökonomie in der schulischen Praxis vielfach verwendet (vgl. Siekmann/ Thomé 2012, S. 164). Aus orthographiedidaktischer Sicht spricht jedoch einiges gegen eine alleinige Diagnostik mittels derlei standardisierter Testverfahren. Zum einen lässt die Fehleranalyse hinsichtlich ihrer qualitativen Differenziertheit in Teilen zu wünschen übrig. Zugunsten von Ökonomie und Interpretationsobjektivität werden rechtschriftliche Phänomene unter wenigen groben Kategorien zusammengefasst (bspw. HSP: drei „Strategien“: alphabetische, orthographische, morphematische Strategie, vgl. May u. a. 2007, S. 24 f.). Dadurch wird „[insbesondere in der ‚orthographischen Strategie‘] [...] einiges zusammengeworfen [...], was graphematisch zu differenzieren wäre“ (Eisenberg/ Fuhrhop 2007, S. 35). Aus förderdiagnostischer Sicht können demzufolge wichtige Informationen über individuelle Stärken und Schwächen des Lerners verloren gehen. Zum anderen muss die Inhaltsvalidität standardisierter Tests kritisch betrachtet werden, da sie rechtschriftliche Leistungen hervorbringen, die sich mit denen in frei verfassten Schreibungen nicht unbedingt decken. Die Schreibdimension der Tests (vorgegebene Items in Einzelwortschreibungen, Diktaten oder Lückentexten) führt zu einer sehr viel höheren „Dichte der Fehlerfallen“ als in frei verfassten Schreibungen. Ein wichtiger Grund – so der Rückschluss von Fay (2010a, b) – weshalb die Leistungen im Test tendenziell schwächer sind als in der authentischen Schreibdimension der frei verfassten Schreibungen.

Begegnen kann man beiden Kritikpunkten durch die Anwendung orthographisch ausgefeilter Fehlerschlüssel auf frei verfasste Lernerschreibungen, womit wir bei der zweiten Konstituente des Dilemmas wären: der differenzierten Rechtschreibdiagnostik. Derlei Instrumente liegen bspw. mit der Aachener förderdiagnostischen Recht-

schreibfehleranalyse, AFRA (Herné/ Naumann 2005), der Oldenburger Fehleranalyse, OLFA (Thomé/ Thomé 2004) und dem Raster von Fay (2010a) vor. Durch hohe Anzahl und Genauigkeit der Rechtschreibkategorien und durch die Tatsache, dass diese Instrumente auf jegliche frei verfassten Schreibprodukte angewendet werden können, wird die Rechtschreibleistung sehr gründlich und „in der Ernstsituation“ des Schreibens (Brügelmann; u. a. 1992, S. 7) erfasst. Nachteil dabei ist wiederum, dass diese Form der Rechtschreibdiagnostik einer Überprüfung bzgl. der Gütekriterien, insbesondere der Ökonomie, Normierung und Vergleichbarkeit, kaum standhalten kann. Vor allem im Hinblick auf eine flächendeckende Anwendung des Verfahrens auf große Gruppen führt die manuelle Analyse frei verfasster Schreibungen zu einem prohibitiv hohen Arbeitsaufwand. So ist eine intensive Nutzung von Fehler-schlüsseln auf frei verfasste Schreibungen im Schulalltag kaum zu finden.

Aus diesem Dilemma – ‚ökonomische vs. differenzierte Rechtschreibdiagnostik‘¹ – entsteht ein *Desiderat: Benötigt wird ein Instrument zur automatisierten Analyse von Rechtschreibfähigkeit in frei verfassten Texten*. Durch die Implementierung eines solchen Instruments wird die Messung der Rechtschreibfähigkeit auf eine neue Stufe gehoben: In der Förderung von Rechtschreibkompetenz ist es damit möglich, auf Basis jeglicher Schreibungen individuelle und orthographisch differenzierte ‚Fehler- und Könnensprofile‘ zu ermitteln, die Ausgangspunkt für zielgerichtete Reflexionen und Übungen sind. Für die didaktische Rechtschreibforschung bedeutet dies ebenso einen großen Fortschritt, da äußerst umfangreiche Stichproben ‚mit einem Mausklick‘ analysiert werden können. So können beispielsweise die gängigen Schreibentwicklungsmodelle der letzten 35 Jahre (Eichler 1976; Dehn 1978, Scheerer-Neumann 1989, Spitta 1986 u. v. a. m.) erstmals anhand umfassender Datensätze überprüft und ggf. modifiziert werden.

Ziel der hier vorgestellten Studie ist daher die Entwicklung eines Instruments zur Automatisierung orthographischer Analyseprozesse freier Texte. Unsere Forschungsgruppe geht aus einem Zusammenschluss von Sprachdidaktik einerseits und automatischer Verarbeitung natürlicher Sprache aus der Informatik andererseits hervor. Die Innovation besteht u. E. darin, dass wir die Lernerschreibungen unter Einbezug ihrer automatisch generierten Aussprache analysieren – als Zusatz zur üblichen Betrachtung der falschen Graphemfolge.

2 Stand der Forschung

In der sprachdidaktischen Forschungslandschaft zur Rechtschreibdiagnostik sind in den letzten Jahren einige computergestützte Verfahren entwickelt worden. Computergestützt bedeutet jedoch nicht in jedem Fall „automatisch“. Im Folgenden unterscheiden wir daher semi-automatische Verfahren (Kap. 2.1), die manuelle Analysen nur in einem engen Korsett vorgefertigter Items ersetzen können, von ersten

1 Ganz ähnlich auch die Beschreibung zum „Objektivitäts-Validitäts-Dilemma“ (u. a. bei Bremerich-Vos 2009, S. 23).

Schritten zu einer vollautomatischen Analyse (Kap. 2.2), die Item-unabhängig, also bei jedem Wortmaterial, funktionieren soll. In Kap. 3 zeigen wir dann, inwiefern der hier verfolgte Ansatz einige der bisherigen Hürden vollautomatischer Analyse überwinden kann.

2.1 Semi-automatische Verfahren

Zu den semi-automatischen Verfahren zählen wir jene Tests, die vorgegebene Items im Einzelwort- und Satzformat anbieten. Frei verfasste Lernerschriften können damit nicht analysiert werden. Es handelt sich dabei um Online-Versionen von bestehenden Paper-and-Pencil-Tests, um genuin Internet-basierte Diagnose- und Förderangebote und um eine Vielzahl an Angeboten von Schulbuchverlagen. Zu nennen sind bspw. „HSP plus“ (www.hsp-plus.de), „gutschrift“ (www.gutschriftinstitut.de), „Münsteraner Rechtschreibanalyse“ (www.lernserver.de) und das Angebot des Schroedel-Verlages (www.grundschuldiagnose.de) und des Cornelsen-Verlages (foedern@cornelsen unter www.cornelsen.de).²

Das technische Vorgehen bei der Auswertung der Rechtschreibleistung in den genannten Online-Tests wird nicht offengelegt. Es ist allerdings zu vermuten, dass jede einzelne Schreibung automatisch mit einer hinterlegten Liste möglicher Falschreibungen verglichen wird, deren Varianten jeweils bestimmten Fehlerkategorien zugeteilt sind. So wird es wahrscheinlich zu dem Item „Spinnennetz“ (HSP 3) Listeneinträge geben wie bspw. *Schpinnennetz (Fehler in der alphabetischen Strategie), *Spinennetz (Fehler in der orthographischen Strategie) und *Spinnenetz (Fehler in der morphologischen Strategie).

Bei der Analyse frei verfasster Schreibungen kann im Gegensatz dazu nicht auf vorgefertigte Fehlerlisten zurückgegriffen werden. Dies macht die besondere technische Herausforderung einer vollautomatischen Analyse aus.

2.2 Erste Schritte zu einer vollautomatischen Analyse

Weil das Wortmaterial in der Analyse freier Schreibungen zuvor nicht bekannt sein kann, ist der entscheidende Weg zu einer vollautomatischen Analyse eine wortunabhängige, abstrakte Betrachtung von Schreibungen. Für die weiteren Ausführungen möchten wir folgende Termini einführen:

Achieved-Text: die Lernerschreibung (inkorrekt oder korrekt)

Target-Text: die intendierte Schreibung, die korrekte ‚Zielform‘

Für eine vollautomatische Rechtschreibanalyse ergibt sich der Bedarf an folgenden beiden Software-Modulen:

Modul 1: Eine automatische Rekonstruktion des Target- aus dem Achieved-Text.

² Alle Internetquellen abgerufen am 26.06.2012.

Modul 2: Eine automatische Annotation und Klassifikation der Rechtschreibfehler über einen Vergleich von Target- und Achieved-Text.

Die bislang einzige Arbeit, die auf eine automatische Analyse rechtschriftlicher Fähigkeiten in frei verfassten deutschen Schreibungen abzielt, liegt mit der Dissertation von Tobias Thelen vor (vgl. Thelen 2010), obgleich das Fehlen einer solchen Software bereits mehrfach thematisiert worden ist (vgl. z. B. Berndt 2007). Thelen gelingt in seiner Arbeit eine relativ zuverlässige, jedoch teils manuell ergänzte Umsetzung des zweiten Moduls, der Klassifikation der Fehler. Dabei wird das Target-Wort linguistisch analysiert und mit verschiedenen phonologischen und morphologischen Zusatzinformationen versehen. Automatisch kann dann im Abgleich mit dem Achieved-Wort angezeigt werden, an welcher Stelle ein Fehler vorliegt (bspw. Fehler in der Verschriftung des Silbenkernes). Eine Abstufung im Grad der Korrektheit differenziert das Ergebnis („repräsentiert, phonetisch plausibel und orthographisch korrekt“, Thelen 2010, S. 40). Nicht hinreichend zufriedenstellend ist hingegen in Thelens Verfahren das Ergebnis des ersten Moduls, der automatischen Rekonstruktion des Target- aus dem Achieved-Text. Hierbei betrachtet er die isolierte Wortform und entwickelt Verfahren auf Basis von Vollformlexika, manuell erstellten Regeln und automatisiert gewonnenen Mustermengen, die jedoch nur ein F-Maß³ von 0,77 erreichen können (vgl. ebd., S. 136). Daraus schließt Thelen: „Ein vollständig autonom agierendes Analyseverfahren für beliebige freie Texte ist nach derzeitiger linguistischer Modellierung der deutschen Orthographie und den derzeitigen Möglichkeiten computerlinguistischer Verfahren unmöglich“ (ebd., S. 137).

Studien aus der Disziplin der automatischen Sprachverarbeitung, die sich unabhängig von der sprachdidaktischen Forschung der Rekonstruktion des Target- aus dem Achieved-Text widmen, kommen hier einen Schritt weiter: Im Gegensatz zu Thelen, der eine isolierte Korrektur einzelner Wörter vornimmt, wird eine kontextabhängige Fehlerkorrektur vorgenommen, bei der die textliche Umgebung des einzelnen Wortes in die Analyse einbezogen wird (Kukich 1992, Verberne 2002). Dadurch können insbesondere auch sog. Realwortfehler erkannt und korrigiert werden (z. B. <Ich liege in *meinen Bett> statt <Ich liege in meinem Bett>). Kontextabhängige Verfahren ermöglichen das Erkennen und die Korrektur solcher Falschschreibungen, indem jeweils ganze Sequenzen von Wörtern betrachtet und deren Auftretenswahrscheinlichkeit mittels sog. N-Gramm-Sprachmodellen⁴ bewertet werden (vgl. Jelinek 1991; Kneser/ Ney 1995).

Alle kontextabhängigen Verfahren und die meisten isolierten Verfahren haben gemeinsam, dass bei der Distanzbestimmung zwischen Target- und Achieved-Text nur die graphemische Form der Wörter betrachtet wird. Lediglich in zwei Studien wird die Aussprache der Wörter mit in Betracht gezogen, um die Schreibung zu korrigie-

3 Das F-Maß ist ein einheitliches Gütemaß, das sowohl die Genauigkeit als auch die Trefferquote einer Klassifikation angibt (vgl. Evert/ Frötschl/ Lindstrot 2010, S. 154 f.).

4 Ein N-Gramm-Sprachmodell zerlegt einen Text in seine sprachlichen Elemente, wobei N deren jeweils gemeinsam betrachtete Anzahl darstellt (vgl. ebd., S. 125).

ren. Diese beiden Arbeiten haben jedoch nur mit isolierten Korrekturverfahren gearbeitet (van Berkel/ Smedt 1988, Toutanova/ Moore 2002).

Wir entwickeln in der vorliegenden Studie ein kontextabhängiges Verfahren zur Rechtschreibkorrektur, das auch die Aussprache des Target- und Achieved-Textes berücksichtigt.

3 Automatische Analyse auf Basis von Speech-Processing-Technologien

Ausgangspunkt für die Analyse von Rechtschreibfähigkeit auf Basis von Speech-Processing-Technologien sind zwei Überlegungen.

Die erste Überlegung betrifft das genuine „Geschäft“ der automatischen Verarbeitung natürlicher Sprache als Forschungsdisziplin. Wie in kaum einer anderen sprachbetrachtenden Wissenschaftsrichtung ist man es in der automatischen Verarbeitung natürlicher Sprache gewöhnt, mit Inkorrektem umzugehen. Gesprochene Sprache ist – betrachtet man sie als informationstechnologischen Gegenstand – per se „unsauber“, d. h. sie weicht durch dialektal geprägte Varietäten und durch artikulatorische Phänomene wie Koartikulationseffekte, Elisionen, Substitutionen usw. von einer theoretischen standardsprachlichen Norm ab. Dennoch gelingt es mit Methoden der automatischen Spracherkennung, das Gesprochene in einen orthographisch korrekten Text zu überführen. Dieses in der Forschungsdisziplin verankerte Potenzial wollen wir für den Umgang mit Inkorrektem in Lernerschriften nutzen.

Die zweite Überlegung betrifft eine etablierte Annahme über das Schreibenlernen. Lerner orientieren sich beim schriftlichen Konstruieren von Sprache zuallererst an ihrer Aussprache. In diesem Sinne gibt es in allen Schreibentwicklungsmodellen eine „alphabetische Phase“ (Thomé 2003; andere Autoren benutzen andere Termini), die beschreibt, wie Schreiblerner durch das Begreifen der basalen Phonem-Graphem-Korrespondenz zu phonographischen Schreibungen wie bspw. <*schprechn> statt <sprechen> gelangen. Deswegen ist es in der Rekonstruktion des Target- aus dem Achieved-Text (Modul 1) sinnvoll, die Aussprache hinzuzuziehen (Bsp.: <*schprechn> → ʃprɛçn → entspricht der Aussprache des Target-Wortes <sprechen>).

Um den Aufbau unseres Systems im Folgenden nachvollziehen zu können, erläutern wir zunächst zwei wichtige Elemente aus der automatischen Verarbeitung natürlicher Sprache (Kap. 3.1). Dann wird das Vorgehen in den beiden Modulen (s. o.) genauer erklärt. Exemplarisch spielen wir die Analyseschritte an folgendem Achieved-Text durch:

*<Es gap einmal ein Könik der hate Soldarte> (entnommen aus Fay 2010a).

Leser, die darüber hinaus an technischen Details interessiert sind, möchten wir auf Stüker/ Fay/ Berkling 2011, Berkling/ Fay/ Stüker 2011 und Fay/ Berkling (i. V.) verweisen.

3.1 Bausteine des Systems

3.1.1 Sprachsynthese

Unter „Sprachsynthese“ versteht man jenes Teilgebiet der automatischen Verarbeitung natürlicher Sprache, das sich mit der Übersetzung geschriebener Sprache in gesprochene Sprache befasst. Sog. „text-to-speech“-Systeme sind jedem bekannt, der z. B. mal eine SMS auf einem Festnetzanschluss erhalten und von einer künstlichen Stimme „vorgelesen“ bekommen hat. Eingabe ist dabei ein Satz als eine Folge von geschriebenen Wörtern, als Ausgabe erfolgt eine akustische Realisation mit einer vorausgewählten Stimme und Sprache. In einem ersten Schritt werden dabei mit Hilfe eines robusten *Syntax-Taggers*⁵ die Satzstrukturen analysiert und Wortarten werden bestimmt. So kann z. B. die Satzintonation entsprechend der Satzart (Frage-, Aussage- oder Nebensatz) angepasst werden. In einem zweiten Schritt werden die Wörter des Satzes in ihre Phonemfolge mit Markierung von Betonung und Silbenstruktur überführt. Die Umwandlung erfolgt dabei entweder über ein *Aussprachewörterbuch* oder, für den Fall, dass für das Wort ein Eintrag im Wörterbuch fehlt, durch *Ausspracheregeln*. Für unser Verfahren haben wir die Sprachsynthese-Software MARY verwendet, die eine XML⁶-basierte Datenrepräsentation verwendet und die Ergebnisse der einzelnen Zwischenschritte der Synthese so verfügbar macht (Schröder/ Trouvain 2003). Für den obigen, falsch geschriebenen Beispielsatz wirft MARY aus:

```
" ? E s" | " g a: p" | " ? aI n - m a: l" | " ? aI n" | " k 2: - n I k" | " d E 6" |
" h a: - t @" | " z O l - d a 6 - t @"7
```

Das System kann also auch falsch Geschriebenes verarbeiten. Die Phonemfolge „klingt“ dann entsprechend der eingegebenen Schreibung entweder richtig (z. B. bei / ? aI n - m a: l/) oder aber verfälscht (z. B. bei / h a: - t @/). Bei phonographischen Falschschreibungen gibt es keinen Unterschied im Output (z. B. bei / g a: p/). Auf Basis der Phonemfolge und der Satzanalyse wird dann mittels *akustischer Modelle* die eigentliche Aussprache in Form eines Sprachsignals generiert. Dieser letzte Schritt fällt für unsere Bedürfnisse weg; wir benötigen nur die Phonemfolge und die Silbenstruktur und -betonung (vgl. Abbildung 1).

5 „Tagging“ meint hier die Annotation von sprachlichen Elementen.

6 „XML“ steht für „Extensible Markup Language“ und ist eine Auszeichnungssprache zur Darstellung hierarchisch strukturierter Daten in Form von Textdaten (vgl. Bray u. a. 2000).

7 Die Ausgabe der Phoneme erscheint nicht in dem bekannten Internationalen Phonetischen Alphabet (IPA-Lautschrift), sondern in der computerlesbaren Form SAMPA (Wells 1989), die mit den normalen Tastaturzeichen auskommt. Das @-Zeichen steht bspw. für den Schwa-Laut.

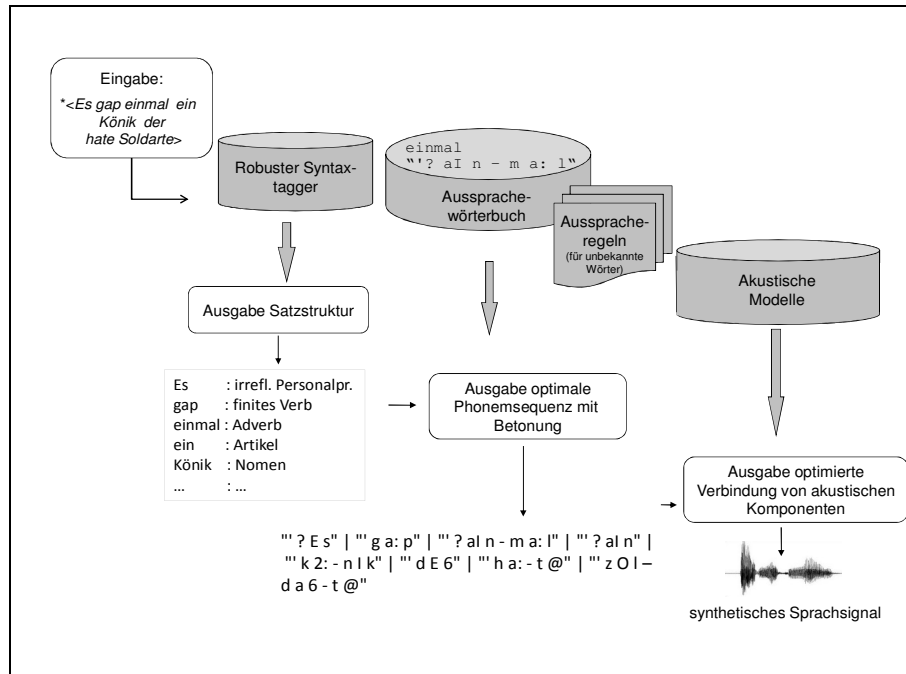


Abbildung 1: Sprachsynthesemodell

3.1.2 Spracherkennung

Unter „Spracherkennung“ versteht man jenes Teilgebiet der automatischen Verarbeitung natürlicher Sprache, das sich mit der automatischen Umwandlung gesprochener Sprache in eine orthographisch korrekte Niederschrift befasst. Solche Programme sind im Arbeitsalltag vor allem im Kontext des direkten In-den-Computer-Diktierens populär geworden. Eingabe ist dabei eine gesprochene Äußerung; Ausgabe ist die hypothetisierte, textuell-graphematische Form der Aufnahme. In unserem Fall: *<Es gab einmal einen König der hatte Soldaten>*⁸. Da gesprochene Sprache, wie eingangs erwähnt, ohnehin „unsauber“ und hoch variabel ist, müssen Spracherkennungssysteme auch mit „falsch klingenden“ Phonemfolgen wie bei / h a : - t @/ und mit grammatischen Fehlern wie bei / ? a l n/ relativ gut umgehen können. Das gelingt mittels einer statistischen Analyse auf verschiedenen Sprachebenen. Abbildung 2 zeigt schematisch den Aufbau eines solchen statistischen Spracherkennungssystems.

⁸ Leider beherrscht das System keine zuverlässige Interpunktion.

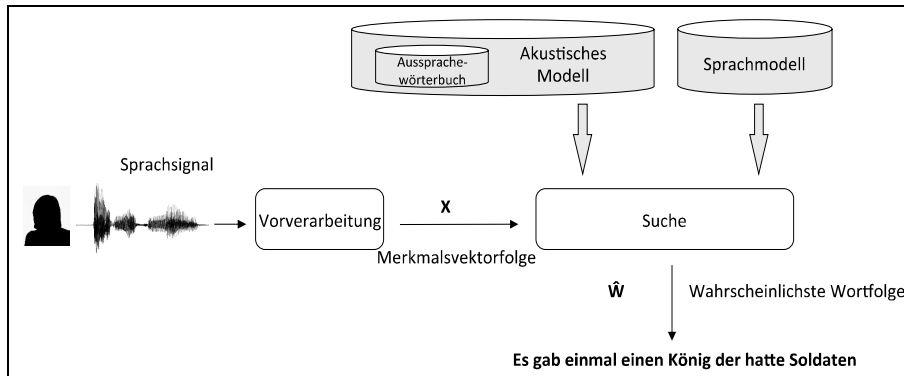


Abbildung 2: Aufbau eines statistischen Spracherkennungssystems

Solche Systeme suchen für ein Sprachsignal nach der wahrscheinlichsten Wortfolge W . Dazu wird die Aufnahme zuerst *vorverarbeitet*, um aus ihr eine Folge von Merkmalsvektoren X zu gewinnen, die für diese Wahrscheinlichkeitssuche, eben der „Spracherkennung“, geeignet ist. Grob geht es dabei um das Produkt zweier Analysen: Zum einen berechnet man die Wahrscheinlichkeit, dass eine bestimmte Wortsequenz W die aufgezeichnete Aufnahme X hervorgebracht hat. Diese Wahrscheinlichkeit wird vom *akustischen Modell* berechnet, das mit Hilfe eines Aussprachewörterbuchs „weiß“, in welche Phonemfolgen ein Wort zerlegt werden kann. Zum anderen wird die Wahrscheinlichkeit berechnet, dass die so ermittelte Wortsequenz in der Sprache überhaupt vorkommt, unabhängig von der gesprochenen Äußerung. Dies wird durch das *Sprachmodell* berechnet, das nicht bloß das einzelne Wort, sondern seinen Kontext in den Blick nimmt. An dieser Stelle findet das System z. B. heraus, dass die Lautfolge / ? aI n/ im Kontext des obigen Satzes <einen> heißen muss, da die statistische Wahrscheinlichkeit dieser Wortfolge höher ist (Deshalb heißt es „kontextabhängiges Verfahren“). Zur Berechnung des akustischen Modells verwenden man in der Regel sog. „Hidden Markov Modelle“ (HMMs)⁹, zur Berechnung des Sprachmodells N-Gramm-Modelle (s. o.).

⁹ Stark verkürzt dargestellt sind Hidden Markov Modelle stochastische Modelle, die z. B. in der Computerlinguistik verwendet werden, um die Wahrscheinlichkeit zu berechnen, mit der bestimmte sprachliche Einheiten aufeinander folgen. In unserem Fall gehen u. a. die Achieved-Graphemkette, die generierte Phonemkette und die Phonem-Graphem-Korrespondenz in die Wahrscheinlichkeitsrechnung mit ein (vgl. ausführlich z. B. Evert/Frötschl/ Lindstrot 2010, S. 130 ff.).

3.2 Modul 1: Automatische Rekonstruktion des Target- aus dem Achieved-Text

Die Rekonstruktion des Target-Textes aus dem Achieved-Text wird in unserem Projekt unter Hinzunahme von Techniken aus der Sprachsynthese als Spracherkennungsproblem formuliert.

Abbildung 3 zeigt das System in der Übersicht.

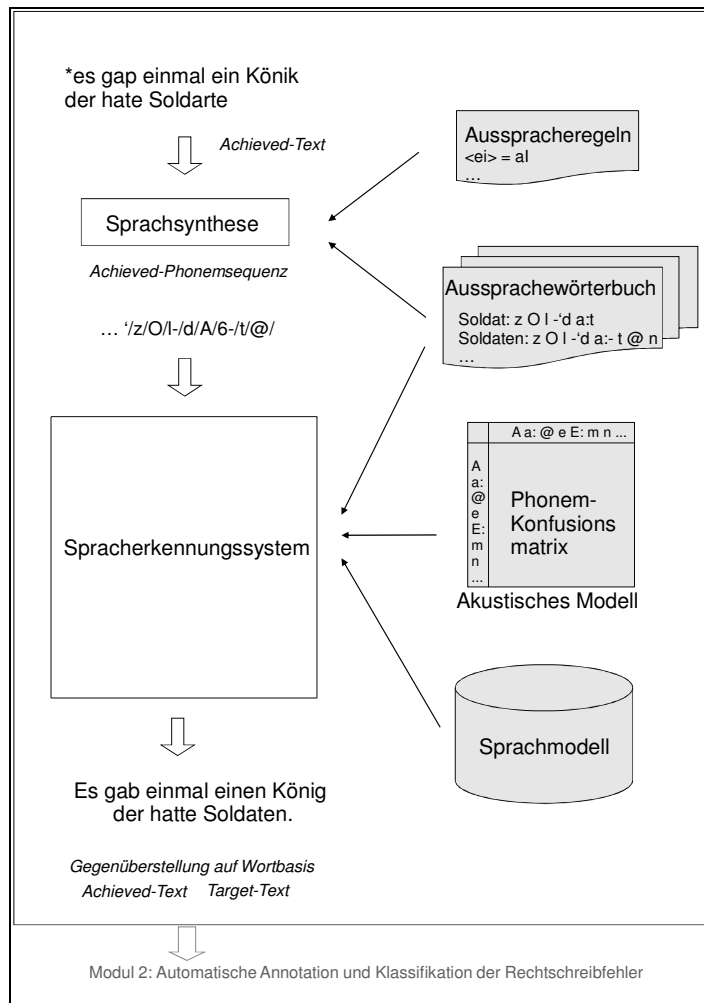


Abbildung 3: Modul 1

Anstatt aus einer Aufnahme die Wortfolge zu rekonstruieren (s. Kap. 3.1.2), wird jetzt aus dem Achieved-Text die korrekt geschriebene Wortfolge gewonnen. Mit

Hilfe der Sprachsynthese (MARY) wird der Achieved-Text erst in eine Phonemfolge umgewandelt – für bekannte Wörter mittels des Aussprachewörterbuchs, für unbekannte, bzw. falsch geschriebene Wörter mit Hilfe von Ausspracheregeln. Diese Phonemfolge dient jetzt als Merkmalsvektor X für das Erkennungssystem.

Das akustische Modell benötigt für unseren Fall Wahrscheinlichkeitsverteilungen, die beschreiben, mit welcher Wahrscheinlichkeit ein Phonem des Target-Textes mit einem Phonem im Achieved-Text assoziiert ist (vgl. Stüker/ Kilgour/ Niehues 2010). Mangels Trainingsmaterials wurden diese Verteilungen nicht in einem System-Trainingsprozess aufgebaut, sondern manuell basierend auf Phonemähnlichkeiten in einer sog. „Phonem-Konfusionsmatrix“ definiert. In einer solchen Matrix werden phonetische Merkmale wie Stimmhaftigkeit und Artikulationsart und -ort mit Werten belegt, mithilfe derer die phonetische Nähe zweier Phoneme berechnet werden kann. So haben bspw. /b/ und /m/ mehr phonetische Gemeinsamkeiten als /t/ und /b/ (vgl. ausführlich Berkling 2001). Mit der Phonem-Konfusionsmatrix kann daher eingeschätzt werden, wie hoch die Gefahr ist, dass ein Phonem im Analyseprozess mit einem anderen verwechselt wird. Zusätzlich benötigen wir diese Daten für die Phonemgegenüberstellung in Modul 2 (vgl. Kap. 3.3, ad. 2). Für das Sprachmodell haben wir ein allgemeines, deutsches N-Gramm-Sprachmodell verwendet, das auf großen Mengen deutscher Texte trainiert wurde, aber nicht domänenspezifisch ist. Die programmatische Umsetzung des Erkennungsprozesses erfolgte mit Hilfe des Janus Recognition Toolkits (JRTk⁺ Soltau/ Metze/ Fügen/ Waibel 2001).

3.3 Modul 2: Automatische Annotation und Klassifikation der Rechtschreibfehler

Liegt nach dem ersten Schritt der Target-Text vor, werden im zweiten Schritt in einer Wort-Wort-Gegenüberstellung von Target- und Achieved-Text Fehler lokalisiert und kategorisiert. Dies geschieht in vier Schritten, in denen die Gegenüberstellung von der bloßen Wortbasis auf eine Graphem- und Phonembasis erweitert wird. Mithilfe dieses Gesamtbildes werden dann je nach Fehlertyp verschiedene Ebenen zur Lokalisierung und Kategorisierung hinzugezogen. So spielen z. B. die Betrachtung der Aussprache, Silbengrenzen und Betonung in einigen Fällen eine entscheidende Rolle; in jedem Fall dienen sie dazu, die regelbasierte Graphemsegmentierung zu unterstützen. Modul 2 vollzieht sich in folgenden vier Schritten (vgl. auch Abbildung 4):

1. Generierung der Aussprache von Target- und Achieved-Text
2. Optimale Gegenüberstellung auf der Phonemebene
3. Graphemsegmentierung und -gegenüberstellung
4. Automatische Fehlerkategorisierung

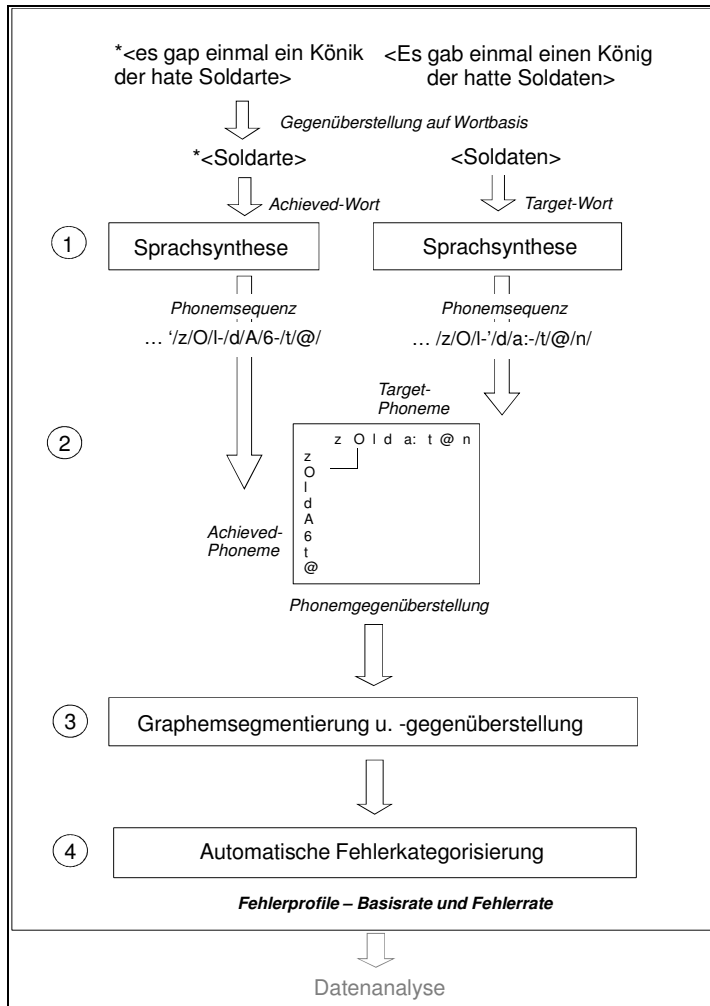


Abbildung 4: Modul 2

Ad 1: Generierung der Aussprache von Target- und Achieved-Text

Wie bereits in Modul 1 verwenden wir die Sprachsynthese-Software MARY, um Target- und Achieved-Text in eine Phonemsequenz umzuwandeln (vgl. Kap. 3.2).

Ad 2: Optimale Gegenüberstellung auf der Phonemebene

Um die graphematischen Abweichungen der Lernerschriften zur korrekten Fassung genau lokalisieren zu können, müssen Wort-, Graphem- und Phonemsequenz präzise („optimal“) gegenübergestellt werden. Je nach Abweichung des Achieved-Textes stellt dies eine weitere Herausforderung für unser System dar. So wird bspw.

<König> und *<Könich> gleich „ausgesprochen“ (/ k 2: - n I C/), wohingegen <der> und *<dea> in ganz unterschiedliche Phonemsequenzen übersetzt werden (/ d E 6/ und / d e: - j a:/). Um nun z. B. Tilgungen, Einfügungen und Substitutionen von Phonemen erkennen zu können, müssen zunächst die Phonemsequenzen von Target- und Achieved-Wort einander zugeordnet werden. Hierzu benötigen wir wieder die Informationen aus der „Phonem-Konfusionsmatrix“ (vgl. Kap. 3.2), die die phonetischen Ähnlichkeitswerte liefert.

Tabelle 1 zeigt eine solche optimale Gegenüberstellung, in der 1.) eine Substitution /a/ - /a:/, 2.) eine Einfügung /6/ und 3.) eine Tilgung des /n/ auftritt.

<Soldaten>	→	z	O	l	´d	a: ^{1.)}	-	t	@	<u>n</u>
*<Soldarte>	→	z	O	l	d	a	6 ^{2.)}	t	@	- ^{3.)}

Tabelle 1: Optimale Gegenüberstellung zweier Phonemfolgen

Ad. 3: Graphemsegmentierung und -gegenüberstellung

Durch die optimale Gegenüberstellung der Phoneme können nunmehr auch die Grapheme einander gegenübergestellt werden, bevor die verschiedenen Ebenen gemeinsam im letzten Schritt der Fehleranalyse übergeben werden. Einige Schwierigkeiten der automatischen Graphemsegmentierung, die sich aufgrund des heterographemen Charakters des deutschen Schriftsystems auftun, können im Zusammenhang mit der Aussprache minimiert werden. Folgende Fälle, bei denen unterschiedliche Graphemtrennungen möglich sind, möchten wir hier beispielhaft nennen.

- Gleiche Buchstabenfolge, aber unterschiedliche Phoneme:
 <s> und <p> in <spielen> und in <Wespe>
- Multifunktion des <h> sorgt für unterschiedliche Segmentierung, die automatisch differenziert werden muss:
 <h> als Silbengrenze in <gehen>, als Dehnungs-h in <nehmen>, als optische Verlängerung in <ihr> oder als einfaches Basisgraphem in <Haus>.
- Morphemgrenzen und Konsonantenbuchstabenverdopplung:
 Kanne>, aber <n> und <n> in <annehmen>

Entscheidend ist bei diesem Schritt, ob das System auf automatisch erkennbare Trennhilfen zurückgreifen kann oder nicht. Silbengrenzen werden in der Regel sicher angezeigt, so dass viele der oben genannten Fälle korrekt analysiert werden (z. B. <sp>-Trennung in / v E s - p @/, nicht aber in / S p i: - l @ n/). Problematisch ist bislang die Sprachsynthese von morphematisch komplexen Wörtern, da Morphemgrenzen noch nicht in der Phonemsequenz markiert sind. So kommt es zu Fehlern wie z. B. falsche Analysen des Morphemanschlusses in {an} {nehmen}:

*/? a - n e: - m @ n/.

Unsere Ergebnisse zeigen deutlich, wie abhängig die ganze Fehleranalyse von einer korrekten Sprachsynthese ist und wie sich insbesondere eine fehlende automatische Erkennung von Morphemgrenzen in diversen Folgefehlern im Analysesystem fort-pflanzt (vgl. Kap. 4.2).

Das Ergebnis der Graphemsegmentierung ist hier beispielhaft an dem Wort <Soldaten> dargestellt (vgl. Tabelle 2).

Achieved-	Phoneme	z	O	l		d	a	6		t	@	
	Grapheme	S	o	l		d	a	r		t	e	
Target-	Grapheme	S	o	l		d	a			t	e	n
	Phoneme	z	O	l	-	'd	a:			t	@	n

Tabelle 2: Graphemsegmentierung und -zuordnung mithilfe der Aussprache am Bsp. <Soldaten>

Ad. 4: Automatische Fehlerkategorisierung

Die auf die beschriebene Weise entstandene (wahrscheinlichste) differenzierte Gegenüberstellung von Achieved- und Target-Text ermöglicht nun eine automatische Benennung der Normabweichungen, d. h. eine Fehlerkategorisierung. Grundlage bietet dafür das Fehlerkategoriensystem von Fay (2010a). In diesem Raster ist das Spezifische des Analysematerials ‚Text‘ berücksichtigt, indem „induktiv Analysekategorien entwickelt worden [sind], die über die orthographische Einzelwortschreibung hinausgehen. Dazu zählen z. B. verschiedene morphosyntaktische Kategorien wie die <das/dass>-Schreibung sowie Fehler in den Morphemanschlüssen über die Wortgrenze hinaus.“ (ebd., S. 57). Das Raster gliedert sich in vier schriftsystematische Ebenen, denen insgesamt 36 Fehlerkategorien zugeordnet sind. In der Software haben wir bislang 24 Kategorien implementiert, die nach und nach systematisch evaluiert werden.

Die Programmierung der Fehleranalyse entspricht nun einem komplexen Bedingungsgefüge (wenn – dann), in dem die Algorithmen jeder Kategorie verzeichnet sind. Dabei muss unbedingt berücksichtigt werden, dass die deutsche Orthographie durch ihren historisch gewachsenen Charakter auch Phänomene aufweist, die sich einer gesetzmäßigen Systematik entziehen. Einige Ausnahmen lassen sich als solche zusätzlich programmieren, andere wird man in der Analyse nicht fassen können.

4 Erste Systemanalysen

Die in Kapitel 3 beschriebenen Module der Software wurden in ersten explorativen Experimenten evaluiert, indem sie (bislang getrennt voneinander) auf das Textkorpus von Fay (2010a) angewendet wurden (vgl. dazu auch Stüker/ Fay/ Berkling 2011; Berkling/ Fay/ Stüker 2011). Bei dem Textkorpus handelt es sich um 120 frei verfasste Lernertexte der Klassen 1 bis 4, wobei jeweils 30 Texte pro Klassenstufe vertreten sind. Das Korpus umfasst 6.312 Wörter (tokens). Die Texte sind zustande gekommen, indem die Versuchsleiterin den Kindern die Geschichte „Eine Schlacht, die könnt ihr haben“ (Röckener 1996) vorgelesen und sie danach gebeten hat, „ihre eigene Geschichte dazu zu schreiben“ (Fay 2010a, S. 102). Die Texte lie-

gen in digitalisierter Form („MS Excel“) als Schülerschreibung (Achieved-Text) und als manuell korrigierte Fassung (Target-Text) vor.¹⁰

4.1 Ergebnisse zu Modul 1: Automatische Rekonstruktion des Target- aus dem Achieved-Text

Die Rekonstruktion der Schreibungen wurde überprüft, indem sie mit den Ergebnissen des bestehenden Programms „Hunspell“ verglichen wurde (<http://hunspell.sourceforge.net/>, Stand: 26.06.2012). Hunspell ist ein mittlerweile etabliertes Rechtschreibkorrekturprogramm auf Basis isolierter Wortfehleranalyse. Es steht als Open-Source-Format zur Verfügung, so dass es problemlos integrierbar ist. Hunspell folgt einem interaktiven Ansatz, bei dem der Nutzer aus mehreren Korrekturvorschlägen das gewünschte Wort wählt. Derlei Formate sind bspw. aus dem Textverarbeitungsprogramm „MS Word“ annähernd jedem PC-Nutzer bekannt. Für die hier beschriebene Evaluation wurde der jeweils erste Korrekturvorschlag eines Wortes gewählt. Die Rechtschreibkorrektur von Hunspell und dem neu entwickelten Verfahren wird in dem Maß „word-error-rate“ (WER), also dem prozentualen Verhältnis der falschen Wörter zur Gesamtwortzahl, angegeben und verglichen (vgl. Tabelle 3).

Text	WER
Achieved- vs. Target-Text (manuell)	20,1%
Hunspell vs. Target-Text	15,6%
„Modul 1“ vs. Target-Text	9,7%

Tabelle 3: Erste Ergebnisse der Evaluation von Modul 1

Die Lernerschreibungen zeigen im Vergleich zu der manuellen Korrektur einen WER von 20,1 %. Die Hunspell-Korrektur kann die WER auf 15,6 % reduzieren. Verwendet man nun unser Modul 1 unter Einbezug des beschriebenen Sprachmodells so ergibt sich eine WER 9,7 % – eine deutliche Verbesserung gegenüber dem etablierten System Hunspell. Dies liegt auch daran, dass traditionelle Rechtschreibkorrekturprogramme viel Gewicht auf die Korrektur von Tastatur-Tippfehlern legen – einem Fehlertyp, der bei der Analyse ursprünglich handschriftlich gefertigter Texte keine Rolle spielen sollte (vgl. Cölfen/ Haller 2011).

Ein Teil zukünftiger Arbeit wird sein, die WER weiter zu minimieren, um so viele Fehler wie möglich für die Analyse in Schritt 2 zu erfassen (vgl. dazu Kap. 5).

¹⁰ Streng genommen handelt es sich bei einigen wenigen Texten bloß um eine Variante des Target, und zwar immer dann, wenn der Achieved-Text mehrere Interpretationsmöglichkeiten offen lässt. „Diffuse Schreibungen“ (Fay 2010, S. 68) werden aber ohnehin aus der weiteren Analyse rausgenommen und separat geführt, so dass der Interpretationsspielraum bei der Korrektur des restlichen Achieved-Textes als nicht all zu groß betrachtet werden darf.

4.2 Ergebnisse zu Modul 2: Automatische Annotation und Klassifikation der Rechtschreibfehler

Im Folgenden zeigen wir exemplarisch den Stand der automatischen Annotation und Klassifikation von Rechtschreibfehlern. Es handelt sich um die Kategorie <ie>-Schreibung („Fehler bei der Schreibung von /i:/ in der betonten Silbe als <ie>, z. B. *spilen“; Fay 2010a, S. 185).

In dieser Kategorie ist eine systematische Modulevaluation mittels Validitätsprüfung so weit vorangeschritten, dass wir die dabei zutage gekommenen „Werkstattprobleme“ gewinnbringend präsentieren und reflektieren können.¹¹ Unser System ist regelbasiert und daher naturgemäß fehleranfällig. Mit der Validitätsprüfung können wir beispielhaft zeigen, inwiefern die Ergebnisse der maschinellen Analyse mit denen der manuellen Analyse (in Fay 2010a) übereinstimmen, und etwaige Abweichungen auf beiden Seiten erklären.

Bei der Beurteilung der Ergebnisse führen wir eine sog. „Wahrheitsmatrix“ an, die die Klassifikationen in vier Fälle einteilen:

1. Richtig positiv (rp): Die Basisrate/der Fehler liegt vor; der Klassifikator hat dies richtig angezeigt
2. Falsch negativ (fn): Die Basisrate/der Fehler liegt vor; der Klassifikator hat dies nicht angezeigt
3. Falsch positiv (fp): Die Basisrate/der Fehler liegt nicht vor; aber der Klassifikator hat dies trotzdem angezeigt
4. Richtig negativ (rn): Die Basisrate/der Fehler liegt nicht vor; der Klassifikator hat dies nicht angezeigt (vgl. Makhoul u. a. 1999).

Mit dieser differenzierten Betrachtung der Klassifikationsergebnisse lassen sich ferner Kenngrößen zur Beurteilung der Klassifikatoren „AUTO“ (maschinelle Analyse) vs. „HAND“ (manuelle Analyse) berechnen, allen voran die „Korrekt-klassifikationsrate“. Sie gibt den Anteil der korrekt annotierten Objekte an.

4.2.1 Ergebnisse zur Analyse der Basisrate

Die „Basisrate“ drückt die Menge der theoretischen Auftretenshäufigkeiten von Rechtschreibfehlern aus und wird daher auch „Fehlerfalle“ oder „Fehlerverlockung“ genannt (Herné/ Naumann 2005, S. 17). Es geht also in unserem Fall um die Frage, wie viele <ie>-Schreibungen in den Target-Wörtern auftreten. Die Analyse der Basisrate ist eine notwendige Voraussetzung, wenn man Rechtschreibfähigkeit

¹¹ Außerdem liegen die Ergebnisse der Kategorie „Reduktionssilbe <e>“ vor („Fehlerhafte Verschriftung der offenen Reduktionssilbe <e> [meistens auf Grund von /ə/-Elision oder der Verschriftung von Buchstabennamen], z. B. *Heft_“; Fay 2010a, S. 186). Diese sind in Fay/ Berkling (i. V.) nachzulesen.

differenziert hinsichtlich Stärken und Schwächen in den Blick nehmen möchte (vgl. ausführlich Fay 2010a, S. 80f.).

Sowohl die maschinelle als auch die manuelle Analyse kommen hier zu ähnlichen, aber eben nicht gleichen Ergebnissen (vgl. Tabelle 4).

		<ie> liegt vor (266)	<ie> liegt nicht vor (6046)
annotiert <ie>	Klassifikator AUTO (258)	(r _p) 258	(f _p) 0
	Klassifikator HAND (252)	(r _p) 252	(f _p) 0
annotiert nicht <ie>	Klassifikator AUTO (6054)	(f _n) 8	(r _n) 6046
	Klassifikator HAND (6060)	(f _n) 14	(r _n) 6046

Tabelle 4: Wahrheitsmatrix zur Basisrate der Kategorie <ie>-Schreibung

Die Ergebnisse unterscheiden sich minimal zugunsten der maschinellen Analyse und sind insgesamt mit einer Korrekturklassifikationsrate von AUTO = 0,99873 und HAND = 0,99778 als sehr zuverlässig einzuschätzen (auch wenn bedacht werden muss, dass die „wahre“ Klassifikation (obere Zeile in der Wahrheitsmatrix) analyseimmanent durch eine Kombination der maschinellen und manuellen Ergebnisse erhoben worden ist). Ein Blick in die Daten erklärt die Unterschiede.

HAND: In drei Fällen wurde die Basisrate übersehen (<befiehlt, Beispiel, sieben>) – ein echter f_n-Fehler. Die übrigen elf Fälle kommen zustanden durch eine systematische Ausklammerung sog. „Sparschreibungen“ aus der Analyse, da von diesen in Fay (2010a) eine relativ zügige Aneignung über einen Sichtwortschatz angenommen worden ist (<diese/r/n/s> und <niemand>; Letzteres erweist sich nun sehr wohl als Hürde im Rechtschreiberwerb, s. z. B. <*nimand>).

AUTO: Acht Mal tritt das Wort <verschieben> auf und wird falsch annotiert. Die Analyseprozesse der Software zeigen, dass ein Fehler in der Sprachsynthese vorliegt. Das System liest fälschlicherweise */f E 6 s - C i: - ? e: - b @ n/, erkennt also nicht die richtigen Morphem- und Silbengrenzen, was dazu führt, dass die Bedingungen für die Verschriftung eines <ie> nicht erfüllt sind.

4.2.2 Ergebnisse zur Analyse der Fehler

Bei der Analyse der Fehler geht es nun um die tatsächlich falsch verschrifteten <ie>-Stellen in den Achieved-Wörtern, ganz gleich auf welche Weise diese falsch geschrieben worden sind. Die Wahrheitsmatrix in Tabelle 5 zeigt die Ergebnisse von AUTO und HAND.

		<ie>-Fehler liegt vor (70)	<ie>-Fehler liegt nicht vor (6242)
annotiert <ie>-Fehler	Klassifikator AUTO (66)	(r_p) 66	(f_p) 0
	Klassifikator HAND (68)	(r_p) 66	(f_p) 2
annotiert nicht <ie>-Fehler	Klassifikator AUTO (6246)	(f_n) 4	(r_n) 6242
	Klassifikator HAND (6244)	(f_n) 4	(r_n) 6240

Tabelle 5: Wahrheitsmatrix zu Fehlern der Kategorie <ie>-Schreibung

Auch hier sind beide Ergebnisse mit einer Korrekturklassifikationsrate von AUTO = 0,99937 und HAND = 0,99905 sehr genau und weisen einen minimalen Unterschied zugunsten der AUTO-Analyse auf. Die Unterschiede erklären sich im Einzelnen wie folgt:

HAND: Die vier f_n -Fehler treten in zwei Fällen in Wörtern auf, die auch schon in der Analyse der Basisraten nicht annotiert worden sind (<*befehlt, *nimand (2x), *rufte>¹²). Zwei Wörter wurden hingegen fälschlicherweise als <ie>-Fehler annotiert (f_p), <Apfeltiere, Hies> (soll heißen: <Apfeltiere, hieß>).

AUTO: Die vier f_n -Fehler sind systematisch über die falsche Basisratenanalyse des Wortes <verschieben> zu erklären (s. o.). Daher sind alle vier Falschschreibungen davon nicht annotiert, <*versiben, *verschiben, *ferschiBen, *verSchiBen>. f_p -Fehler macht AUTO hier nicht.

5 Diskussion und Ausblick

Die ersten Ergebnisse aus der Anwendung unseres Instrumentes und deren Evaluation zeigen, dass sich der unkonventionelle Umweg über die maschinell generierte gesprochene Sprache für die Diagnostik von Rechtschreibleistung lohnt. Sowohl die automatische Generierung des Target-Textes als auch die Fehleranalyse selbst profitieren erheblich von diesem Vorgehen. Automatische und manuelle Analyse liegen in der Beispielskategorie <ie>-Schreibung im Ergebnis nah beieinander; fehlerfrei sind sie beide nicht. Während die Verfehlungen per Hand in der Regel zufällig und planlos passieren, macht unsere Software systematische Fehler, deren Lösungen weitestgehend bekannt sind (s. weiter unten). Der entscheidende Unterschied zwischen den beiden Klassifikatoren ist jedoch der zeitliche Aufwand: Die manuelle Analyse der 120 Lernertexte nahm mehrere Monate in Anspruch, die automatische Analyse einer Fehlerkategorie dauert hingegen keine zwei Minuten – ein Unterschied, der die noch bestehenden Einbußen in der vollständigen Korrektheit erträg-

12 Bei <*rufte> ließe sich darüber streiten, ob es sich hier wirklich um einen <ie>-Fehler oder nicht viel eher um einen Tempus-Fehler handele. Wir entscheiden in solchen Fällen nach der Logik des „objektiven Fehlers“ (Meyer-Schepers/ Löffler 1994, S. 7) und der „Mehrfachsignierung“ (Herné/ Naumann 2005, S. 19), bemühen also ggf. mehrere Fehlerkategorien und verzichten auf so viele Ursachenvermutungen wie möglich.

lich erscheinen lässt. Zusätzlich kommt durch die Arbeit mit der Software eine neue Flexibilität im Umgang mit den Daten hinzu. Mit wenigen Programmierschritten können zum einen die Schreibdaten der Lerner gefiltert werden, z. B. hinsichtlich Lernergruppen (sortiert nach Klassenstufen, DaZ vs. DaM, Geschlecht usw.) oder hinsichtlich des Wortmaterials (nur Substantive betrachten, alle Artikel auslassen usw.). Zum anderen ist durch eine Zusammenstellung der verwendeten Fehlerkategorien via Mausclick ein schon fast spielerischer Vergleich von Bereichen rechtsschriftlicher Kompetenz möglich. Was also manuell sorgfältig geplant werden muss, um Arbeitsressourcen gut einzuteilen, kann maschinell durchaus einfach ausprobiert werden.

Für „ein vollständig autonom agierendes Analyseverfahren“ (Thelen 2010, S. 137) liefern wir daher eine Basis, die in folgender Hinsicht ausgebaut werden muss:

Für die Eingabe in Modul 1 müssen die Lernertexte in digitaler Form vorliegen; ein Zustand, der bisher nicht dem Gros der Praxis schulischer Textproduktion entspricht. Hierfür wäre die Überführung handschriftlicher Texte in das digitale Medium via *automatischer Handschrifterkennung* hilfreich. Aktuelle Handschrifterkennungssysteme arbeiten allerdings häufig auch wort- oder zumindest silbenbasiert mit einem Hintergrundvokabular, so dass eine direkte Anwendung auf Fehlschreibungen nicht möglich ist (vgl. Plamondon u. a. 2000). Hier bedarf es vielmehr weiterer Anpassungen der Technik.

Für eine *Weiterentwicklung des Modul 1* gilt es in Zukunft, die Beschränkungen des Systems aufzuheben. Das bedeutet, alle Parameter der zu Grunde liegenden Wahrscheinlichkeitsverteilungen zu lernen, die Neusegmentierung der Phonemfolge in Wörter zuzulassen sowie die Emissionswahrscheinlichkeitsberechnung¹³ der HMMs zu verbessern. Ferner muss das verwendete Sprachmodell speziell für die gestellte Aufgabe (Lernertexte) entwickelt und durch eine detaillierte Leistungsanalyse die Qualität der Rekonstruktion des korrekten Textes verbessert werden.

Modul 1 und Modul 2 müssen *miteinander verknüpft* werden, indem mehr Informationen von Modul 1 an Modul 2 gegeben werden und nicht nur die als korrekt erkannte Wortfolge, so z. B. alternative Rekonstruktionshypothesen.

In der Fehleranalyse des Modul 2 ist eine entscheidende Hürde, die *Leistung des Bausteins „Sprachsynthese“* zu erhöhen. Solange es keine automatische Erkennung von *Morphemgrenzen* gibt, wird auch die so wichtige Silbenanalyse fehleranfällig bleiben, was wiederum zu diversen Folgefehlern führt (s. Kap. 4.2). In diesem Zusammenhang sind Forschungsarbeiten von Kurimo u. a. (2010) verheißungsvoll.

In der Implementierung der Fehlerkategorien steht zunächst eine gründliche Evaluation auf einer breiten Datenbasis an. Danach können auch *grammatikalische Fehlerkategorien* (u. a. Flexion nach Kasus und Tempus) angegangen werden. Hier sind die Erfolgsaussichten insofern günstig, als dass morphosyntaktische Informationen

13 Die Emissionswahrscheinlichkeiten berechnen, mit welcher Wahrscheinlichkeit ein bestimmtes Phonem einen bestimmten, kleinen Ausschnitt aus dem Audiosignal produziert hat.

wie Wortart, Kasus und Tempus in vielen Fällen mit der automatischen Sprachsynthese mitgeliefert werden.

Zu guter Letzt muss unser System eine anwenderfreundliche Oberfläche bekommen, so dass es tatsächlich in der Schule und der didaktischen Rechtschreibforschung eingesetzt werden kann.

Literatur

- Berkling, Kay (2001): SCoPE, syllable core and periphery evaluation: automatic syllabification and foreign accent identification. In: *Speech Communication* volume 35, numbers 1-2, S. 125-138.
- Berkling, Kay/ Fay, Johanna/ Stüker, Sebastian (2011): Speech Technology-based Framework for Quantitative Analysis of German Spelling Errors in Freely Composed Children's Texts. In: *Proceedings of the 2011 workshop of the ISCA (International Speech Communication Association). Special Interest Group (SIG) on Speech and Language Technology*, S. 24-26. August 2011, Venedig, Italien.
- Berndt, Elin Birgit (2007): Warum Fachdidaktiker den Dialog mit der Informatik suchen sollten - am Beispiel der Rechtschreibkorrekturprogramme. In: Josting, Petra/ Jonas, Hartmut/ Albrecht, Christian (Hrsg.): *Themen-Schwerpunkt Intermediale und interdisziplinäre Lernansätze im Deutschunterricht*. München, S. 136-152.
- Bray, Tim/ Paoli, Jean/ Sperberg-McQueen, Michael/ Maler, Eve/ Yergeau, François (2000): *Extensible Markup Language (HML) 1.0 (Second Edition)*. Technische Spezifikation, W3C. [online]. Verfügbar unter: <http://www.w3.org/TR/REC-xml>, [26.06.2012].
- Bremerich-Vos, Albert (2009): Die Bildungsstandards Deutsch. In: Ders. u. a. (Hrsg.): *Bildungsstandards für die Grundschule: Deutsch konkret*. Berlin, S. 14-42.
- Brügelmann, Hans/ Spitta, Gudrun/ Lange, Inge (1992): Zum ‚Laufstalleffekt‘ im Rechtschreibunterricht. Neue Ergebnisse aus dem ‚Schreibvergleich Ost-West‘ 1990 und 1991. In: Balhorn, Heiko (Hrsg.): *Fibel ade? Lesen und Schreiben in der Grundschule*. Hamburg, S. 3-13.
- Cölfen, Hermann/ Haller, Johann (2011): Sprachbezogene Computeranwendung. In: Knapp, Karlfried (Hrsg.): *Angewandte Linguistik. Ein Lehrbuch mit CD-ROM*. Tübingen, S. 177-195.
- Dehn, Mechthild (1978): Strategien beim Erwerb der Schriftsprache. In: *Grundschule/10*, Heft 7, S. 308-310.
- Eichler, Wolfgang (1976): Zur linguistischen Fehleranalyse von Spontanschreibungen bei Vor- und Grundschulkindern. In: Hofer, Adolf (Hrsg.): *Lesenlernen: Theorie und Unterricht*. Düsseldorf, S. 246-265.
- Eisenberg, Peter/ Fuhrhop, Nanna (2007): Schulorthographie und Graphematik. In: *Zeitschrift für Sprachwissenschaft*, Heft 26 Jubiläumshft, S. 15-41.
- Evert, Stefan/ Frötschl, Bernhard/ Lindstrot, Wolf (2010): Statistische Grundlagen. In: Carstensen, Kai-Uwe u. a. (Hrsg.): *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg, S. 114-158.
- Fay, Johanna (2010a): Die Entwicklung der Rechtschreibkompetenz beim Textschreiben. Eine empirische Untersuchung von Klasse 1 bis 4. Frankfurt/ Main.

- Fay, Johanna (2010b): Kompetenzfacetten in der Rechtschreibdiagnostik. Rechtschreibleistung im Test und im freien Text. In: Didaktik Deutsch, Heft 29. Baltmannsweiler, S. 15-36.
- Fay, Johanna (2012): Prozessorientierte Rechtschreibdiagnostik – Wie kommen richtige und falsche Schreibungen zustande? In: Grundschulunterricht Deutsch. Heft 3/2012, S. 33-37.
- Fay, Johanna/ Berkling, Kay (i. V.): Computergestützte Rechtschreibdiagnostik. In: Fay, Johanna (Hrsg.): (Schrift-)Sprachdiagnostik heute. Theoretisch fundiert, interdisziplinär, prozessorientiert und praxistauglich. Baltmannsweiler.
- Herné, Karl-Ludwig/ Naumann, Carl Ludwig (2005): Aachener Förderdiagnostische Rechtschreibfehler-Analyse. Systematische Einführung in die Praxis der Fehleranalyse mit Auswertungshilfen zu insgesamt 31 standardisierten Testverfahren als Kopiervorlagen. Aachen.
- Huang, Xuedong/ Acero, Alex/ Hon, Hsiao-Wuen (2001): Spoken Language Processing: A Guide to Theory, Algorithm and System Development. New Jersey.
- Hunspell-Programm [online] Download verfügbar unter <http://sourceforge.net/projects/hunspell> [26.06.2012]
- Jelinek, Fred (1990): Self-organized language modeling for speech recognition. In: Waibel, Alex u. a. (Hrsg.): Readings in Speech Recognition. San Mateo, S. 450-506.
- Kneser, Reinhard/ Ney, Hermann (1995): Improved backing-off for n-gram language modeling. In: Proceedings the 1995 IEEE International Conference on Acoustics, Speech and Signal Processing. Detroit, S. 181-184.
- Kukich, Karen (1992): Techniques for automatically correcting words in text. In: ACM computing surveys (CSUR) volume 24, issue 4. New York, S. 377-439.
- Kurimo, Mikko u. a. (2010): Morpho Challenge competition 2005-2010: evaluations and results. In: *Proceedings of the 11th Meeting of the ACL Special Interest Group on Computational Morphology and Phonology (SIGMORPHON '10)*. Association for Computational Linguistics, Stroudsburg, PA, USA, S. 87-95.
- Lienert, Gustav/ Raatz, Ulrich (1998): Testaufbau und Testanalyse. Weinheim.
- Makhoul, John/ Kubala, Francis/ Schwartz, Richard/ Weischedel, Ralph (1999): Performance measures for information extraction. In: Proceedings of the DARPA Broadcast News Workshop. Herndon 1999.
- Mariani, Joseph (2009): Spoken language processing. London/ Hoboken.
- May, Peter/ Vieluf, Ulrich/ Malitzky, Volkmar (2007): Diagnose orthographischer Kompetenz. Handbuch, Manual. Hamburger Schreib-Probe S. 1-9. Hamburg.
- Meyer-Schepers, Ursula/ Löffler, Ilona (1994): Fehlertypologien und Rechtschreibfehler. In: Praxis Deutsch, Heft 124, S. 6-13.
- Müller, Rudolf (2004): Diagnostischer Rechtschreibtest für 3. Klassen. DRT 3. Göttingen.
- Plamondon, Réjean/ Srihari, Sargur N. (2000): On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. In: IEEE Transactions on Pattern Analysis and Machine Intelligence volume 22, number 1, S. 63-84.
- Röckener, Andreas/ Wiencirz, Gerlinde (1996): Eine Schlacht? Die könnt ihr haben! München.
- Scheerer-Neumann, Gerheid (1989): Rechtschreibschwäche im Kontext der Entwicklung. In: Naegele, Ingrid u. a. (Hrsg.): LRS in den Klassen 1-10. Handbuch der Lese- und Rechtschreibschwierigkeiten. Weinheim/ Basel, S. 25-35.

- Schröder, Marc/ Trouvain, Jürgen (2003): The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. In: *International Journal of Speech Technology* volume 6, number 4, S. 365-377.
- Siekmann, Katja/ Thomé, Günther (2012): *Der orthographische Fehler. Grundzüge der orthographischen Fehlerforschung und aktuelle Entwicklungen*. Oldenburg.
- Soltau, Hagen/ Metze, Florian/ Fügen, Christian/ Waibel, Alex (2001): A one pass decoder based on polymorphic linguistic context assignment. In: *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '01)*. Madonna di Campiglio, S. 214-217.
- Spitta, Gudrun (1986): Kinder entdecken die Schriftsprache. In: Naegele, Ingrid u. a. (Hrsg.): *'Schreiben ist wichtig!'* Frankfurt/ Main, S. 67-83.
- Stock, Claudia/ Schneider, Wolfgang (2008): *Deutscher Rechtschreibtest für das erste und zweite [dritte und vierte] Schuljahr*. Göttingen.
- Stüker, Sebastian/ Kilgour, Kevin/ Niehues, Jan (2010): Quero speech-to-text and text translation evaluation systems. In: *Proceedings of High Performance Computing in Science & Engineering 2010. The 13th Results and Review Workshop of the HLRS*. Karlsruhe.
- Stüker, Sebastian/ Fay, Johanna/ Berkling, Kay (2011): Towards Context-dependent Phonetic Spelling Error Correction in Children's Freely Composed Text for Diagnostic and Pedagogical Purposes. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)*, 28.-31- August 2011, Florenz, Italien.
- Thelen, Tobias: (2010): *Automatische Analyse orthographischer Leistungen von Schreibanfängern*. Dissertation. Universität Osnabrück.
- Thomé, Günter (2003): Entwicklung der basalen Rechtschreibkenntnisse. In: Bredel, Ursula u. a. (Hrsg.): *Didaktik der deutschen Sprache*. Paderborn, S. 369-380.
- Thomé, Günther/ Thomé, Dorothea (2004): *OLFA. Oldenburger Fehleranalyse. Instrument und Handbuch zur Ermittlung der orthographischen Kompetenz ab Klasse 3*. Oldenburg.
- Toutanova, Kristina/ Moore, Robert C. (2002): "Pronunciation modeling for improved spelling correction". In: *40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, S. 144-151.
- van Berkel, Brigitte/ de Smedt, Koenraad (1988): "Triphone analysis: a combined method for the correction of orthographical and typographical errors". In: *ANLC '88 Proceedings of the second conference on Applied natural language processing*. Stroudsburg, S. 77-83.
- Verberne, Suzan (2002): *Context-sensitive spell checking based on word trigram probabilities*. Master thesis Taal, Spraak & Informatica. University of Nijmegen.
- Wells, John C. (1989): Computer-coded phonemic notation of individual languages of the European Community. In: *Journal of the International Phonetic Association* volume 19, number 1, S. 31-54.

Internetquellen (abgerufen am 26.06.2012):

Hamburger Schreibprobe: www.hsp-plus.de

gutschrift: www.gutschrift-institut.de

Münsteraner Rechtschreibanalyse: www.lernserver.de

Schroedel-Verlag: www.grundschuldiagnose.de

fördern@cornelsen: www.cornelsen.de

Anschrift der Verfasserinnen und des Verfassers:

Juniorprof. Dr. Johanna Fay, Pädagogische Hochschule Karlsruhe; Institut für deutsche Sprache und Literatur und ihre Didaktik; Bismarckstraße 10; 76133 Karlsruhe

fay@ph-karlsruhe.de

Prof. Kay Margarethe Berkling, PhD, Duale Hochschule Baden-Württemberg Karlsruhe; Fakultät Technik; Erzbergerstraße 121; 76133 Karlsruhe

Berkling@dhbw-karlsruhe.de

Dr. Sebastian Stüker, Karlsruher Institut für Technologie (KIT); Institut für Anthropomatik; Nachwuchsgruppe 3-01 'Multilinguale Spracherkennung'; Adenauer-ring 2; 76135 Karlsruhe

sebastian.stueker@kit.edu