

# Ein Tool zur nutzerfreundlichen Anwendung automatischer Rechtschreibfehleranalyse in frei geschriebenen Kindertexten

Ronja Laarmann-Quante, Stefanie Dipper,  
Anna Ehlert, Simon Masloch, Katrin Ortmann, Doreen Scholz, Maurice Vogel  
{laarmann-quante|dipper}@linguistics.rub.de

Ruhr-Universität Bochum

## Ziel und Motivation

**Individuelle Förderung von Schülern:**  
Gerade im Schriftspracherwerb bringen Schüler/innen sehr unterschiedliche Voraussetzungen mit, z.B. Kinder mit Migrationshintergrund, die gleichzeitig mit der deutschen Schriftsprache auch das gesprochene Deutsch erst noch erlernen.

**Qualitative Rechtschreibfehleranalyse:**  
Wie Siekmann (2015) zeigte, ist die individuelle Fehleranalyse in frei geschriebenen Texten von Grundschulkindern eine gute Basis für die Förderung des Schriftspracherwerbs. Solch individuelle Analysen sind jedoch aufwendig und setzen ein hohes Fachwissen im Bereich Orthographie von Seiten der Lehrer/innen voraus.

**Automatisierung:** Das hier vorgestellte Tool nimmt mittels computerlinguistischer Verfahren eine automatische Rechtschreibfehleranalyse vor und kann so Lehrende beispielsweise beim Erkennen besonderer Problembereiche einzelner Schüler/innen unterstützen. Anders als existierende aktuelle Systeme zur Kategorisierung von Rechtschreibfehlern (Berkling & Lavalley, 2015; Ebner et al., 2017) hat es den Anspruch, alle Rechtschreibfehler in beliebigem Wortmaterial von Kindertexten zu kategorisieren und zudem Aussagen darüber treffen zu können, wo ein bestimmter Fehler hätte gemacht werden können, jedoch nicht aufgetreten ist.

## Automatische Rechtschreibfehlerkategorisierung

**Herausforderung:** Korrekte Lokalisierung eines Fehlers bzw. mehrerer Fehler

Original	v	e	r	m	i	e	s	s	t
Zielwort	v	e	r	m	i	s	s	t	
Fehler						repl_C			

Original	v	e	r	m	i	e	s	s	t
Zielwort	v	e	r	m		i	s	s	t
Fehler						rem_Vlong_short	Cdouble_beforeC		

basierend auf kleinstem String-Abstand  
→ 1 Fehler (<e> statt <s>)



korrekte Alignment  
→ 2 Fehler (<i> statt <i> und <s> statt <ss>)



**Lösung:** Annotation linguistischer Informationen mithilfe des BAS Web-Service G2P<sup>1</sup> (Reichel, 2012; Reichel & Kisler, 2014)

fröhlich	f r ' 2: . l I C	fröhlich	ADJ SFX
	<b>Phoneme + Silben</b>		<b>Morpheme Morphemtypen</b>

Mithilfe dieser Informationen: regelbasierte Erkennung, welche systematischen Fehler und Fehlerkombinationen in einem Zielwort gemacht werden könnten (z. Zt. 44 Kategorien)

→ Erzeugung von Fehlerkandidaten

Grapheme	v	e	r	m	i	ss	t
Phoneme	f	ɛ	ɣ	m	ɪ	s	t
repl_unmarked_marked	f	e	r	m	i	ss	t
voc_r	v	e	a	m	i	ss	t
rem_Vlong_short	v	e	r	m	ie	ss	t
Cdouble_beforeC	v	e	r	m	i	s	t
rem_Vlong_short + Cdouble_beforeC	v	e	r	m	ie	s	t

- Abgleich des Originalworts mit allen Fehlerkandidaten des Zielworts
- Bei **vollständiger Übereinstimmung**:
  - Ablesen der Fehler vom Fehlerkandidaten
- **Ansonsten**:
  - es müssen unsystematische Fehler vorliegen (z.B. \*<vemiest>, fehlendes <r>)
  - Fehlerkandidat mit geringstem Stringabstand zur Originalschreibung als Ausgangspunkt: Erfassung weiterer Fehler über Edit-Operationen

## Weitere Annotationen

Basierend u.a. auf der Fehlerkategorie und den linguistischen Annotationen des Zielworts werden durch Regeln weitere Merkmale ermittelt, die für die Beurteilung einer Fehlschreibung und angemessenes Feedback von Bedeutung sein können (siehe auch Laarmann-Quante et al., 2016):

- **foreign\_target**: Ist das Zielwort ein Fremdwort bzw. folgt es nicht dem Aufbau des dt. Kernwortschatzes?
- **exist\_orig**: Ergibt eine Falschschreibung ein existierendes Wort (basierend auf dem Kinderbuchkorpus *childLex* (Schroeder et al., 2015))?
- **syllable\_plausible\_orig**: Sind die Silben des Originalworts plausible Silben des Deutschen?
- **phon\_orig\_ok** (pro Fehler): Bleibt die Aussprache des Wortes trotz des Fehlers erhalten?
- **morph\_const** (pro Fehler): Ist die richtige Schreibung von einem verwandten Wort ableitbar?
- **Sprachausgabe** der Original- und Zielschreibung mittels MARY TTS (Schröder & Trouvain, 2003)

### EVALUATION

866 Ziel-Tokens, 3 Annotatoren

Level	Anz.	autom./Mensch		Mensch	
		Proz.	k	Proz.	k
error_cat (alle)	261	78,67%	0,77	77,39%	0,82
phon_orig_ok	227	80,35%	0,63	78,03%	0,72
morph_const	227	79,54%	0,57	70,99%	0,55

## Anwendungsprototyp

Konsolenanwendung mit HTML-Ausgabe

**Alle Fehler**  
Anzahl Wörter: 48    Anzahl fehlerhafter Wörter: 16    Fehlerrate: 33,33%    Anzahl Fehler: 21

Klicke auf einen Fehler, um nähere Informationen angezeigt zu bekommen

Der Lars und Lea gehen mit Dodo zu draußen aber  
Der Lars und Lea gehen mit Dodo zu draußen aber

Dodo reht zu schnel . Der Dodo ist gerant der ist  
Dodo reht zu schnell . Der Dodo ist gerannt der ist

auf die Wiese gerant . Der Dodo sieht ein Frsch .  
auf die Wiese gerannt . Der Dodo sieht ein Frösch .

der Lars sagt . Dodo will auf die Mau springen .  
der Lars sagt . Dodo will auf die Mauer springen .

der hat gesprngt . Und da ist auf die Was .  
der hat gesprngt . Und dann ist auf die Wasser

springen ?!  
springen ?!

Oder wähle eine bestimmte Fehlerkategorie aus:

Fehlerkategorie	falsch alle
repl_unmarked_marked	3
voc_r	3
Cdouble_final	2
hyp_final_device	2
hyp_voice	2
rem_Vlong_short	2
repl_C	1
Cdouble_interV	1
Vlong_h	1
Vlong_l	1
final_device	1
rem_Cdouble_long	1
low_up	1
repl_V	1
repl_C	1
rem	0
repl_unmarked_unmarked	0
rem_Vlong_short	0
Vlong_h	0
Vlong_single	0

**Fehler und korrekte Verschriftungen der Kategorie voc\_r**

Alle Fehler anzeigen

Der Lars und Lea gehen mit Dodo zu draußen aber  
Der Lars und Lea gehen mit Dodo zu draußen aber

Dodo reht zu schnel . Der Dodo ist gerant der ist  
Dodo reht zu schnell . Der Dodo ist gerannt der ist

auf die Wiese gerant . Der Dodo sieht ein Frsch .  
auf die Wiese gerannt . Der Dodo sieht ein Frösch .

der Lars sagt . Dodo will auf die Mau springen .  
der Lars sagt . Dodo will auf die Mauer springen .

der hat gesprngt . Und da ist auf die Was .  
der hat gesprngt . Und dann ist auf die Wasser

springen ?!  
springen ?!

Fehler und korrekte Verschriftungen einer Kategorie

Hauptansicht mit Markierung aller gemachten Fehler sowie Aufschlüsselung nach Fehlerkategorien; Angabe, wie viele Fehler einer Kategorie hätten gemacht werden können und wie viele tatsächlich gemacht worden sind

Dies ist ein Fehler der Kategorie **final\_device**

falsch: **sagt**  
richtig wäre: **sagt**

Die Schreibung des Lerners ist phonetisch plausibel  
Die korrekte Schreibung kann von einer verwandten Form hergeleitet werden

Zusätzliche Hinweise zu einem Fehler mit Sprachausgabe der Aussprache

## Ausblick

- Arbeit an praxistauglichen Fehlerkategorien, Zusatzinformationen, visueller Darstellung etc. (Zusammenarbeit mit Didaktikern wünschenswert)
- Applikation, die direkt von Kindern genutzt werden kann; dazu: automatische Zielworterkennung + kindgerechtes Feedback
- Automatische Annotation der Fehler im Litkey-Korpus (~ 1900 Texte von Grundschulkindern der Klassen 2-4; Laarmann-Quante et al., 2017)

More Info: <https://www.linguistics.rub.de/litkey/Scientific/Corpusanalysis>

Laarmann-Quante, R. (2016). Automating multi-level annotations of orthographic properties of German words and children's spelling errors. In *Language Teaching, Learning and Technology* (pp. 14-22). ISCA + Laarmann-Quante, R., Keichel, L., Dipper, S., & Betken, C. (2016). Annotating spelling errors in German texts produced by primary school children. In A. Friedrich & K. Tomaneck (Eds.), *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)* (pp. 32-42). Laarmann-Quante, R. (2017). Towards a tool for automatic spelling error analysis and feedback generation for freely written German texts produced by primary school children. In *Proceedings of the Seventh ISCA Workshop on Speech and Language Technology in Education (SLTE)* (pp. 44-49).



<sup>1</sup><https://gitlab.phonetik.uni-muenchen.de/BASWebServices/Interface/Grapheme2Phoneme>  
Berkling, K., & Lavalley, R. (2015). WISE: A web interface for spelling error recognition for German. A description of the underlying algorithm. In *Proceedings of the 11th Conference of the German Society for Computational Linguistics and Language Technology (G2L)* (pp. 87-96). Duisburg/Essen, Germany + Ebner, M., Lötters, L., & Ebner, M. (2017). Taking writing spelling aids into a web-based platform for children. *Universal Access in the Information Society*, 1-10.  
Reichel, U. D. (2012). *Phoneme and Syllable: Tools for string alignment and text processing*. In *INTERPRECH*, Portland, Oregon + Reichel, U. D., & Kisler, T. (2014). Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In *International Journal of Computational Linguistics: Studies in Computational Linguistics* 77 (pp. 42-66).  
Schroeder, M., & Trouvain, J. (2003). The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology*, 6(4), 365-377.  
Schröder, S., Wotzner, K.-M., Heister, J., Gayken, A., & Ringel, R. (2015). *childLex: A lexical database of German read by children*. *Behavior Research Methods*, 47(1), 280-294.  
Siekmann, K. (2015). *Erdebeobachtete Förderung der Orthographieerwerbs auf der Grundlage individueller Fehleranalysen*. In K. Liebers, R. Landwehr, A. Marquardt, & K. Schuster (Eds.), *Wortbuch Grundschulunterricht: Band 15 Lernprozessbegleitung und adaptives Lernen in der Grundschule: Forschungsbegleitende Beiträge* (pp. 199-204). Wiesbaden: Springer VS.