

Detaillierte Rechtschreibdiagnostik effizient gestalten: Möglichkeiten der automatischen Rechtschreibfehleranalyse

Ronja Laarmann-Quante & Stefanie Dipper
{laarmann-quante|dipper}@linguistics.rub.de

Ruhr-Universität Bochum

Motivation

- Wie Siekmann (2015) zeigt, ist die **individuelle Auseinandersetzung** mit den Rechtschreibfehlern eines Kindes in frei geschriebenen Texten eine gute Basis für weitere Förderung
- Da dies manuell sehr **zeitaufwendig** ist, ist eine individuelle Diagnose von Fehlerschwerpunkten im Schulalltag von Lehrenden jedoch kaum zu leisten
- Computerlinguistische Anwendungen erlauben heute schon eine **automatische Analyse** von Rechtschreibfehlern, die bisher aber nicht alle Fehlertypen erfassen (Fay, Berking & Stüker, 2012; Berking & Lavalley, 2015)
- Der hier vorgestellte Anwendungsprototyp geht über bisherige Tools hinaus, indem basierend auf einer gegebenen Falschschreibung und der korrespondierenden Richtigschreibung
 - alle Fehler nach **verschiedenen Schemata** (Litkey, OLFA) kategorisiert werden
 - zu jeder Fehlschreibung **zusätzliche Informationen** bereitgestellt werden können

Automatisch generierte Informationen

Zum Zielwort: Phoneme, Silben, Morpheme sowie POS-Tag (Stanford Tagger; Toutanova, Klein, Manning & Singer, 2003; BAS Web-Service G2P⁺; Reichel, 2012; Reichel & Kisler, 2014)

Zu jedem Fehler:

- Fehlerkategorie aus feingliedrigem Litkey-Schema mit 80 Kategorien (Laarmann-Quante et al., 2016) oder OLFA-Kategorien (Thomé & Thomé, 2004)
- Sind die Silben des Originalworts graphotaktisch wohlgeformte Silben des Deutschen?
- Weicht das Zielwort strukturell vom deutschen Kernwortschatz ab?
- Ist die richtige Schreibung von einem verwandten Wort ableitbar?
- Bleibt die Aussprache des Wortes trotz des Fehlers erhalten?
- Ergibt eine Falschschreibung ein existierendes Wort (basierend auf dem Kinderbuchkorpus *childLex* (Schroeder et al., 2015))?

Anwendungsprototyp

1 Hauptansicht

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 **trau**en abba
Der Lars und Lea gehen mit Dodo zu draußen aber

Dodo **re**nt zu **schne**l . Der **do**do ist **gera**nd der ist
Dodo rennt zu schnell . Der Dodo ist gerannt der ist

auf die **W**ise **gera**nd . Der Dodo **si**ht ein **Fr**usch
auf die Wiese gerannt . Der Dodo sieht ein Frosch .

der Lars **sa**kt . Dodo will auf die **Mau**a springen
der Lars sagt . Dodo will auf die Mauer springen .

der hat **gesp**ingt . Und **da**n ist auf die **Wa**sa springen
der hat gesprängt . Und dann ist auf die Wasser springen

springen ?!
springen ?!

Oder wähle eine bestimmte Fehlerkategorie aus:

Fehlerkategorie	falsch	alle
Cdouble_beforeC	3	3
vocR	3	11
Cdouble_final	2	3
hvp_final_device	2	14
voice	2	40
Vlong_th_ieh	1	1
Cdouble_interV	1	1
Vlong_l_ie	1	4
final_device	1	8
ovr_Cdouble_afterVlong	1	19
low_up	2	-
repl_VV	1	-
ins_C	1	-
form	0	52
repl_marked_unmarked	0	38
ovr_Vlong_short	0	27

Litkey-Kategorien

Rote Markierung der Fehler direkt im Text; beim Drüberfahren mit der Maus wird die Fehlerkategorie aus dem gewählten Schema (Litkey oder OLFA) angezeigt

Die Tabelle zeigt an, wie oft ein Fehler einer Kategorie hätte gemacht werden können, (Spalte *alle*) und wie oft dabei tatsächlich ein Fehler gemacht worden ist (Spalte *falsch*).
≥ 50% falsch: rot, < 50% falsch: gelb, immer korrekt: grün)

OLFA-Kategorien

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 **trau**en abba
Der Lars und Lea gehen mit Dodo zu draußen aber

Dodo **re**nt zu **schne**l . Der **do**do ist **gera**nd der ist
Dodo rennt zu schnell . Der Dodo ist gerannt der ist

auf die **W**ise **gera**nd . Der Dodo **si**ht ein **Fr**usch
auf die Wiese gerannt . Der Dodo sieht ein Frosch .

der Lars **sa**kt . Dodo will auf die **Mau**a springen
der Lars sagt . Dodo will auf die Mauer springen .

der hat **gesp**ingt . Und **da**n ist auf die **Wa**sa springen
der hat gesprängt . Und dann ist auf die Wasser springen

springen ?!
springen ?!

Oder wähle eine bestimmte Fehlerkategorie aus:

Fehlerkategorie	falsch	alle
07-EinfDoppl	6	7
37-Sonst	4	86
20-bdg-ptk	2	11
13-s-b	1	1
19-ptk-bdg	1	3
09-EinfMarkL	1	4
33-KonsFalsch	1	44
11-DopplKons	1	76
01-KlGr	2	-
34-VokFalsch	1	-
29-KonsFehl	1	-
10-MarkLEinfl	0	22
18-ae-e	0	19
08-DoppVl	0	13

OLFA-Kategorien

2 Einzelkategorien

Beim Klick auf einen Fehler im Text oder eine Fehlerkategorie in der Tabelle wird im Text angezeigt, wo ein Fehler dieser Kategorie gemacht wurde (rot) und wo sonst noch ein Fehler dieser Kategorie hätte gemacht werden können aber nicht gemacht wurde (grün).

Fehler und korrekte Verschriftungen der Kategorie 07-EinfDoppl

Alle Fehler anzeigen

Der Lars und Lea gehen mit Dodo zu **trau**en abba
Der Lars und Lea gehen mit Dodo zu draußen aber

Dodo **re**nt zu **schne**l . Der dodo ist **gera**nd der ist
Dodo rennt zu schnell . Der Dodo ist gerannt der ist

auf die **W**ise **gera**nd . Der Dodo **si**ht ein **Fr**usch
auf die Wiese gerannt . Der Dodo sieht ein Frosch .

der Lars **sa**kt . Dodo **wi**ll auf die **Mau**a springen
der Lars sagt . Dodo will auf die Mauer springen .

der hat **gesp**ingt . Und **da**n ist auf die **Wa**sa springen
der hat gesprängt . Und dann ist auf die Wasser springen

?!
?!

Fehler und korrekte Verschriftungen der Kategorie 19-ptk-bdg

Alle Fehler anzeigen

Der Lars **un** Lea gehen mit Dodo zu trausen abba
Der Lars **un** Lea gehen mit Dodo zu draußen aber

Dodo **ren**t zu **schne**l . Der dodo ist **gera**nd der ist
Dodo rennt zu schnell . Der Dodo ist gerannt der ist

auf die **W**ise **gera**nd . Der Dodo **si**ht ein **Fr**usch
auf die Wiese gerannt . Der Dodo sieht ein Frosch .

der Lars **sa**kt . Dodo will auf die **Mau**a springen
der Lars sagt . Dodo will auf die Mauer springen .

der hat **gesp**ingt . Und **da**n ist auf die **Wa**sa springen
der hat gesprängt . Und dann ist auf die Wasser springen

?!
?!

OLFA-Kategorie „Einfachschiebung für Konfonantenverdoppelung“

OLFA-Kategorie „p t k für b d g im Silbenende“

3 Zusätzliche Informationen

Beim Klick auf einen Fehler im Text werden zusätzliche Informationen zu diesem Fehler angezeigt.

Dies ist ein Fehler der Kategorie 19-ptk-bdg

falsch: **sa**kt richtig wäre: **sa**gt

Die Schreibung des Lesers ist phonetisch plausibel

Die korrekte Schreibung kann von einer verwandten Form hergeleitet werden

Die Sprachausgabe der Original- und Zielschreibung erfolgt mittels MARY TTS (Schroder & Trouvain, 2003)

Ausblick & Desiderata

- Systematische Evaluation und weitere Verbesserung der automatischen Fehleranalyse
- Automatische Korrektur der falschgeschriebenen Wörter zur weiteren Reduktion des manuellen Aufwands
- Herausarbeitung des geeignetsten Feedbacks für die jeweilige Nutzergruppe (Lehrkräfte, Sprachförderkräfte, ggf. Lerner selbst)
- Zur Zeit: Ausarbeitung zur Erprobung im L2-Lerner Kontext

Mehr Informationen: <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 *Proceedings of the 2nd Language Teaching, Learning and Technology Workshop (LTL7)* (pp. 14-22). San Francisco, USA. • Laarmann-Quante, R., Krichel, U., Dipper, S., & Berken, C. (2016). Annotating spelling errors in German texts produced by primary school children. In A. Friedrich & K. Tomasek (Eds.), *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 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 (SLaTE)* (pp. 44-49).

<https://litr.phonetik.uni-muenchen.de/BASWebServices/Interface/Grapheme2Phoneme>
Referenzen: Berking, K., & Lavalley, R. (2015). WISE: A web-interface for spelling error recognition for German: A description of the underlying algorithm. In *Proceedings of the Int. Conference of the German Society for Computational Linguistics and Language Technology (GSCU)* (pp. 87-96). Duisburg/Essex, Germany. • Fay, J., Berking, K., & Stüker, S. (2012). Automatische Analyse von Rechtschreibfehlern auf Basis von Speech Processing-Technologien. *Diakata Deutsch*, 33, 15-36. • Reichel, U. D. (2012). Perma and Sallona: Tools for string alignment and text processing. In *INTERSPEECH*, Portland, Oregon. • Reichel, U. D., & Kisler, T. (2014). Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In K. Hoffmann (Ed.), *Elektronische Sprachverarbeitung: Studientexte zur Sprachkommunikation 71* (pp. 42-49). Tübingen. • 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. • Schroeder, S., Würzner, K. M., Henner, J., Grieken, A., & Kling, R. (2015). ChildLex: A lexical database of German used by children. *Behavior Research Methods*, 47(4), 1085-1094. • Siekmann, K. (2015). Evidenzbasierte Förderung des Orthographieerwerbs auf der Grundlage individueller Fehleranalysen. In K. Liebers, B. Landwehr, A. Marquardt, & K. Schlotter (Eds.), *Jahrbuch Grundschulforschung Band 18: Lernprozessbegleitung und adaptives Lernen in der Grundschule: Forschungsübergreifende Beiträge* (pp. 199-204). Weinheim: Springer VS. • Thomé, G., & Thomé, D. (2004). Oldenburg Fehleranalyse OLFA: Instrument und Handbuch zur Ermittlung der orthographischen Kompetenz aus freien Texten ab Klasse 3 und zur Qualitätssicherung von Fördermaßnahmen. Oldenburg: ISB Verlag. • Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In M. Hearst & M. Ostensdorf (Eds.), *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL 03* (pp. 173-180). Morristown, NJ, USA: Association for Computational Linguistics.

