

*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 Lernerschreibungen 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.