Towards automated literacy diagnostics: Insights from reading and spelling error analysis

Wieke Harmsen, Roeland van Hout, Catia Cucchiarini, Helmer Strik | Radboud University Nijmegen, The Netherlands

Reading and spelling are core skills that children learn in primary school. Learning these skills is a complicated process, as it requires the contribution and interaction of many different linguistic and cognitive components, such as phonological awareness, morphological awareness, working memory, visual competence and motor skills (e.g., Caravolas et al., 2012). This results in individual differences between children in reading and spelling development. It is especially important to identify children with reading and spelling problems in due time, so that personalized support can be offered before it is too late. Therefore, standardized oral word reading (decoding) and word dictation tests are administered in schools.

These tests give standardized global reading or spelling proficiency scores that are helpful in identifying children with reading or spelling problems. However, our claim is that more and more detailed diagnostic information can be extracted from these tests that provides insight into the problems these children experience in reading and writing. The present study develops an innovative approach that produces more specific information through automatic, detailed analyses of reading and spelling errors made by Dutch third and fourth graders with reading and spelling problems. The research question we address is: To what extent do analyses of recorded and transcribed oral word reading tests and digitized handwritten word dictations provide diagnostic information on specific reading and spelling errors? These analyses are performed through an algorithm that is able to detect and classify reading and spelling errors at Phoneme-Corresponding Unit (PCU) level (Harmsen et al., 2021, 2024). A PCU is one letter or a sequence of letters that corresponds to one phoneme. We present results to answer our research question and discuss insights on automatic reading and spelling diagnostics that can be obtained through this approach. We then present our conclusions and suggest avenues for future research.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. *Psychological Science*, *23*(6), 678-686. https://doi.org/10.1177/0956797611434536

Harmsen, W. N., Cucchiarini, C., & Strik, H. (2021). Automatic Detection and Annotation of Spelling Errors and Orthographic Properties in the Dutch BasiScript Corpus. *Computational Linguistics in the Netherlands Journal*, *11*, 281–306. https://www.clinjournal.org/clinj/article/view/140

Harmsen, W., Cucchiarini, C., van Hout, R., & Strik, H. (2024). A Joint Approach for Automatic Analysis of Reading and Writing Errors. In *Proceedings of the Second Workshop on Computation and Written Language (CAWL) at LREC-COLING 2024*, 8–17.