MultiplEYE Final Conference
August 20-21, 2026
Skopje, North Macedonia

Integrating Statistical and Computational Modeling of Eye Movement Control in Reading with a Linked Mixture Model

Reinhold Kliegl[*1] **and Maximilian M. Rabe**[2]

[1]Department of Sports and Health Sciences, University of Potsdam
[2]Department of Psychology, University of Potsdam

*contact: reinhold.kliegl@uni-potsdam.de

Background and Aims: Research on eye-tracking-while-reading corpora oscillates between two paradigms: statistical modeling of aggregated summary statistics (e.g., go-past times) and computational modeling of fixation sequences (e.g., SWIFT, E-Z Reader). Both approaches possess limitations; statistical models often rely on the questionable assumption that every fixation is intentional, while computational models frequently impose rigid processing assumptions and ignore the hierarchical nature of participant data. As large-scale, multilingual datasets like the MultiplEYE corpus emerge, there is a critical need for flexible modeling frameworks that can handle diverse linguistic structures without sacrificing statistical rigor.

Methodology: We present a highly extendable linked mixture model designed to bridge these traditions. Unlike traditional summary-statistic approaches, this model is fitted directly to raw fixation locations and durations. It accounts for stochasticity in target selection and saccadic error, similar to computational models, but avoids restrictive lexical processing assumptions. Crucially, the framework utilizes a hierarchical structure, allowing for highly customizable fixed and random effects. This makes it uniquely suited for large-scale corpora where individual differences and linguistic variability must be explicitly modeled.

Results: Applying the model to the Potsdam Sentence Corpus, we already recovered some well-established oculomotor patterns, including fixation durations, skipping probabilities, and saccade amplitudes with the first attempt; there are also some puzzling results currently under investigation. Importantly, the model setup enabled the explicit estimation of how categorical and continuous factors (such as reader age) influence and co-vary across individuals.

Discussion: The proposed linked mixture model offers a robust alternative for the analysis of reading data. In the context of the MultiplEYE initiative, this methodology provides a pathway to move beyond ”average” reading behaviors. By accounting for the interplay between saccadic execution and fixation durations within a unified statistical framework, this approach is particularly relevant for crosslinguistic comparisons where language-specific constraints may interact differently with the oculomotor system. We look forward to explore how this model can be applied to the newly released MultiplEYE corpus to enhance our understanding of reading behavior across the 25 documented languages.