Oculomotor Control in Reading: Findings from Turkish

Cengiz Acarturk, Aysegul Ozkan

Jagiellonian University, Krakow, Poland

{cengiz.acarturk, aysegul.ozkan}@uj.edu.pl

Eye movements have been investigated over the past decade to study reading and its associated perceptual, cognitive, and linguistic processes. However, challenges persist due to the simplicity of the metrics, the intricate nature of the processes under investigation, and technical challenges in the measurement techniques. The fundamental metrics revolve around patterns of fixations, focusing on aspects such as their duration and position on the text. These basic metrics are supplemented by derived ones that predominantly capture dynamic eye movement patterns over text, like regressions. To delve into the dynamics tied to these processes, one strategy involves analyzing word attributes that affect information absorption via foveal, parafoveal, and predictive mechanisms. Recent research on eye-tracking in reading largely deals with foveal information intake in major languages. Yet, the broad applicability of these findings is contingent upon patterns seen across a variety of languages. As such, it's vital to broaden the scope of the data by adding eye movement studies from languages that are less frequently studied.

This presentation introduces TURead, an eye movement dataset that captures both silent and spoken sentence reading in Turkish—a language with agglutinative properties and a shallow orthography that's been relatively unexplored in reading studies. The dataset encompasses 192 sentences, read by 215 participants both silently and aloud. The participants' eye movements were captured using a single-camera (right eye) integrated into the SR Research EyeLink 1000 eye tracking system with a tower mount setup, recording at a frequency of 1000 Hz. Audio recordings were captured for text and filler stimuli using a compatible sound card.

TURead provides empirical data that bridges morphology and oculomotor control. We employ a target-word approach, where we adjust target words based on their length and the integration of two prevalent suffixes in Turkish. This dataset incorporates established eye movement metrics, along with prelexical attributes like vowel harmony and bigram-trigram frequencies. Additionally, it considers word attributes such as length, predictability, frequency, eye-voice span measures, and Cloze test results for root word and suffix predictabilities. The data also encompasses scores from two distinct working memory tests. Besides the commonly examined eye movement metrics (such as first fixation duration, gaze duration, last saccade amplitude, next saccade amplitude, first fixation location, and launch site) for both oral and silent reading, TURead incorporates four specific measures exclusive to oral reading. These include fixation speech interval, eye-voice span based on character count, eye-voice span based on word count, and articulation duration.

Central to our findings is the replication of canonical effects highlighting the influence of frequency and predictability of fixated words on eye movement patterns in sentence reading. Notably, we found the FSI (Fixation Speech Interval) in Turkish exceeds that of English and German, aligning more closely with Finnish, another language with a direct orthographic structure. Such elevated FSI values in languages like Turkish and Finnish could signify the impact of a shallow orthography on preliminary phonological processing. Furthermore, we observed shorter EVS (Eye Voice Span) metrics in Turkish sentence compared to earlier research, implying a potential influence of the shallow orthography on the working memory buffer.

The dataset is publicly available at *TURead: An Eye Movement Dataset of Turkish Reading* in Open Science Framework OSF Repository (https://osf.io/w53cz/). More information can be obtained at Acarturk, C., Ozkan, A., et al. (2023). TURead: An eye movement dataset of Turkish reading. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02120-6.