Title: Reading development at the text level: an investigation of surprisal and embedding-based text similarity effects on eye-movements in Chinese early readers

Student name: Xi Fan

Supervisor: Ronan Reilly

Abstract: This study describes the use of semantic similarity measures based on distributed representations of words, sentences, and paragraphs (so-called “embeddings”) to assess the impact of supra-lexical factors on eye-movement data from early readers of Chinese. In addition, we used a corpus-based measure of surprisal to assess the impact of local word predictability. Eye movement data from 56 Chinese students were collected (a) in the students’ 4th grade and (b) one year later while they were in 5th grade. Results indicated that the surprisal and text similarity measures have a significant impact on the moment-to-moment processing of words in reading. The study is the first attempt, as far as we are aware, to track the developmental trajectory of these influences in developing readers across a range of reading abilities.