An eye-tracking database of natural reading in Italian children
Valentina N. Pescuma¹, Maria Ktori¹, Benedetta Cevoli¹,², Eleonora Lomi¹,³, Francesca Franzon¹, Davide Crepaldi¹
¹ International School for Advanced Studies (SISSA), Trieste, Italy. ² Royal Holloway University of London, UK. ³ University of Oxford, UK.

Background
• Reading proficiency may build up through the chunking of lower-level units (e.g., letters) into larger ones (e.g., words and morphemes)¹,².
• Morpho-orthographic chunking in adults may be interpreted similarly — morphology drives regularities in letter co-occurrence within words³,⁴,⁵, which the reading system may exploit to facilitate visual word identification.
• In this perspective, reading may be conceived as a form of statistical learning.

Aims and Hypotheses
• To identify statistical learning proxies in developing readers of Italian (3rd – 6th graders).
• Focus on nGram frequency. Other possible indexes (e.g., transitional probabilities, word predictability) are currently under investigation.
• Age expected to play a role in the development of sensitivity to statistical learning cues in reading.
• Data to be made available as one of the first eye-tracking databases in children.

Methods
Participants:
• 112 (63 M) native Italian speakers; age range: 8–12 years (mean=9.85, SD=1.13).

Procedure:
• Natural reading task on texts from kids’ books.
• Simple 2-AFC comprehension questions after every other excerpt displayed.
• Eye movements recorded through a tower-mount Eyelink 1000 Plus eye-tracker.
• Computerized cloze probability task, currently under analysis.

Additional assessment:
• Reading test (MT test – Speed proficiency and Accuracy⁶)
• Non-verbal intelligence test (Raven CPM-47⁷).

Results
Total Looking Time (TLT)
In line with previous data⁸,⁹, benchmark effects on total looking of word length (F[4, 59593] = 370.75, p<.001) and word frequency (F[4, 59593] = 105.12, p<.001).

Mean nGram Frequency
• Size gradient: no effect of 2Gram frequency (F[4, 16410]=0.85, p=.493);
• 3Grams slightly better (F[4, 16410]=2.36, p=.051); significant effect of 4Grams (F[4, 16410]=4.93, p=.001). Figures on the right.
• Size gradient particularly strong with reference to early processing measures (FoMFD).

First-of-Many Fixation Duration (FoMFD)
Effects of word frequency (F[4, 19012]=14.11, p<.001) on FoMFD. A trend for word length (F[4, 19012]=1.82, p=.121).

Mean 4Gram freq
FoMFD

Conclusions
• Sensitivity to letter co-occurrence statistics from a very young age.
• Higher sensitivity to larger clusters than to small ones.
• High grapheme–phoneme correspondence in Italian — does phonology contribute to the statistical regularities, that are “inherited” by orthography once the mapping is learnt? Cross-linguistic research is needed.
• What is coded is still not clear. Lexical mediation in the present analysis (interest areas are words) but this shouldn’t be taken for granted.