

MECO

**the Multilingual Eye Movement Corpus (and statistical learning,
and reading)**

Davide Crepaldi
SISSA and Università di Pavia
Italy

Prospettive interdisciplinari nella misura di competenze e capacità linguistiche in età scolare
June 24, 2024

The theoretical question

Reading and statistical learning

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals
- Measure Reading for the same individuals

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals
- Measure Reading for the same individuals
- Look for a correlation

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals
- Measure Reading for the same individuals
- Look for a correlation

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals
- Measure Reading for the same individuals
- Look for a correlation

Issues:

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals
- Measure Reading for the same individuals
- Look for a correlation

Issues:

- Tests unreliable at the individual level (reliability paradox)

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals
- Measure Reading for the same individuals
- Look for a correlation

Issues:

- Tests unreliable at the individual level (reliability paradox)
- Need a lot of power

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals
- Measure Reading for the same individuals
- Look for a correlation

Issues:

- Tests unreliable at the individual level (reliability paradox)
- Need a lot of power
- Construct validity

The theoretical question

Reading and statistical learning

- Measure Statistical Learning for some set of individuals
- Measure Reading for the same individuals
- Look for a correlation

Issues:

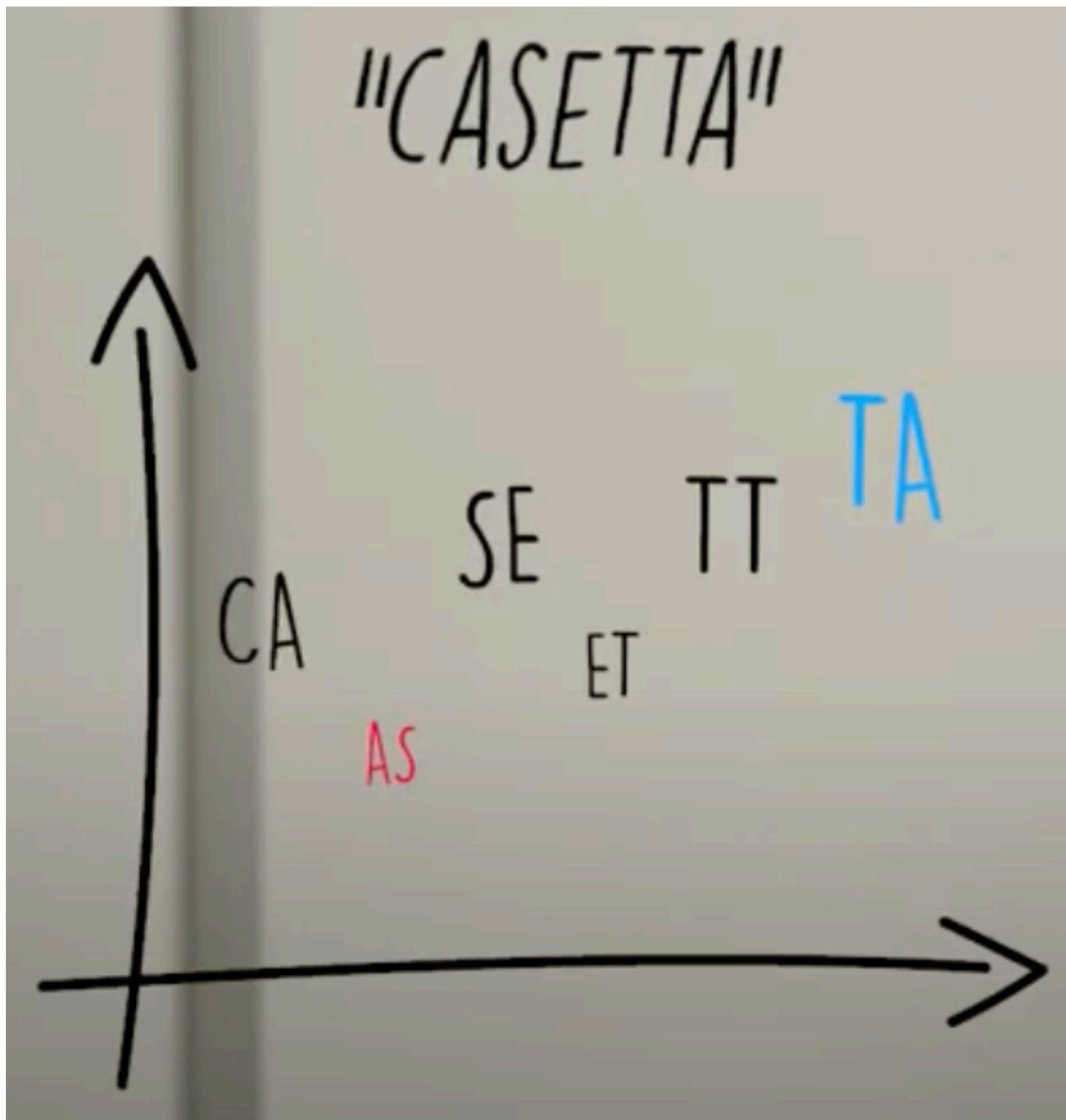
- Tests unreliable at the individual level (reliability paradox)
- Need a lot of power
- Construct validity
- Multi-componentiality: what do we measure for SL? And what for reading?

A different approach

Look for signs of SL natively, within reading itself

A different approach

Look for signs of SL natively, within reading itself



A different approach

Look for signs of SL natively, within reading itself



A different approach

Look for signs of SL natively, within reading itself



EyeReadIt

EyeReadIt

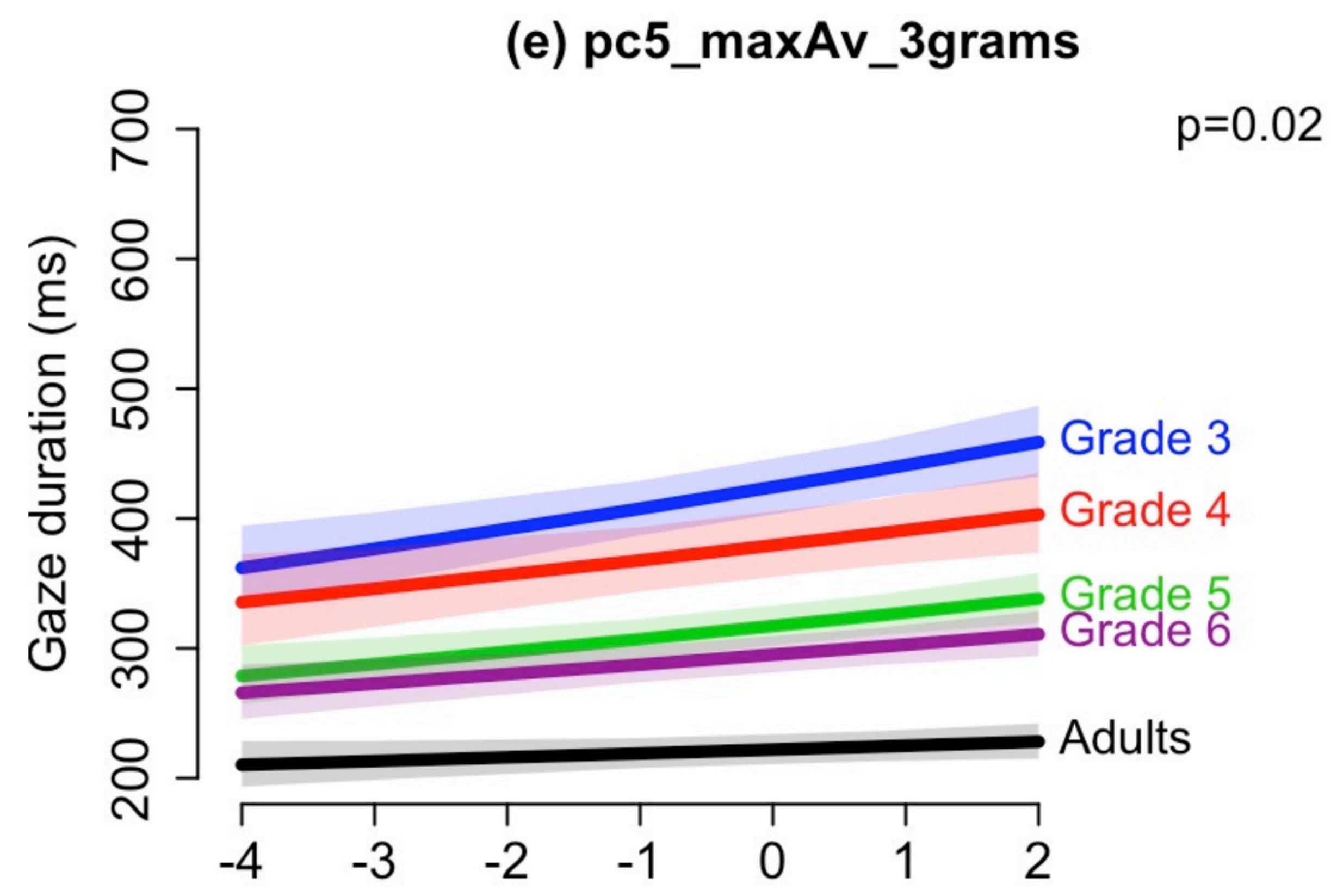
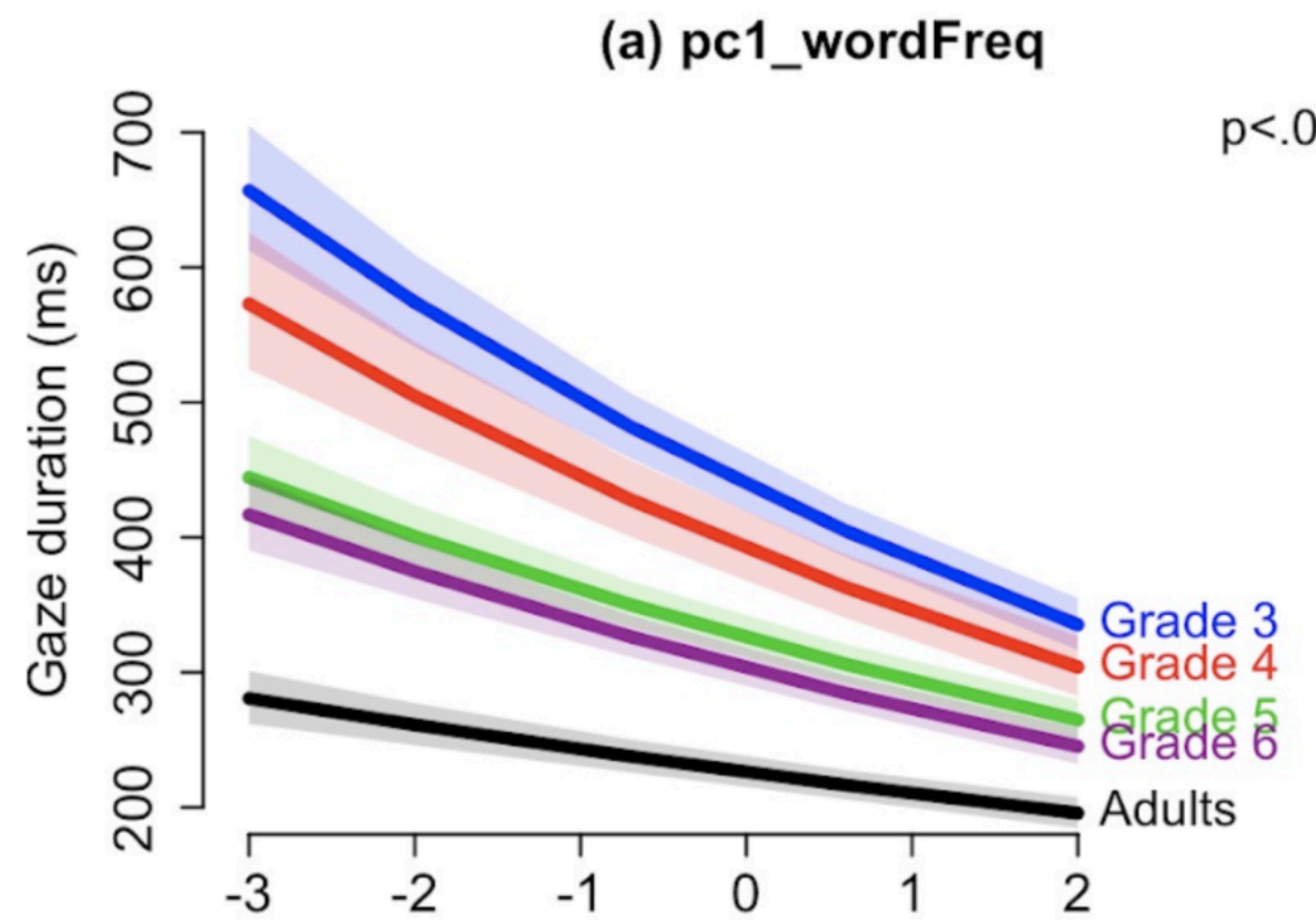
- Some 170 children, from 1st to 6th grade (though most are 3rd-5th graders)

EyeReadIt

- Some 170 children, from 1st to 6th grade (though most are 3rd-5th graders)
- 1566 words each

EyeReadIt

- Some 170 children, from 1st to 6th grade (though most are 3rd-5th graders)
- 1566 words each
- A total 117K interest area data, more than 204K fixations



MECO

MECO

- 600ish adult readers

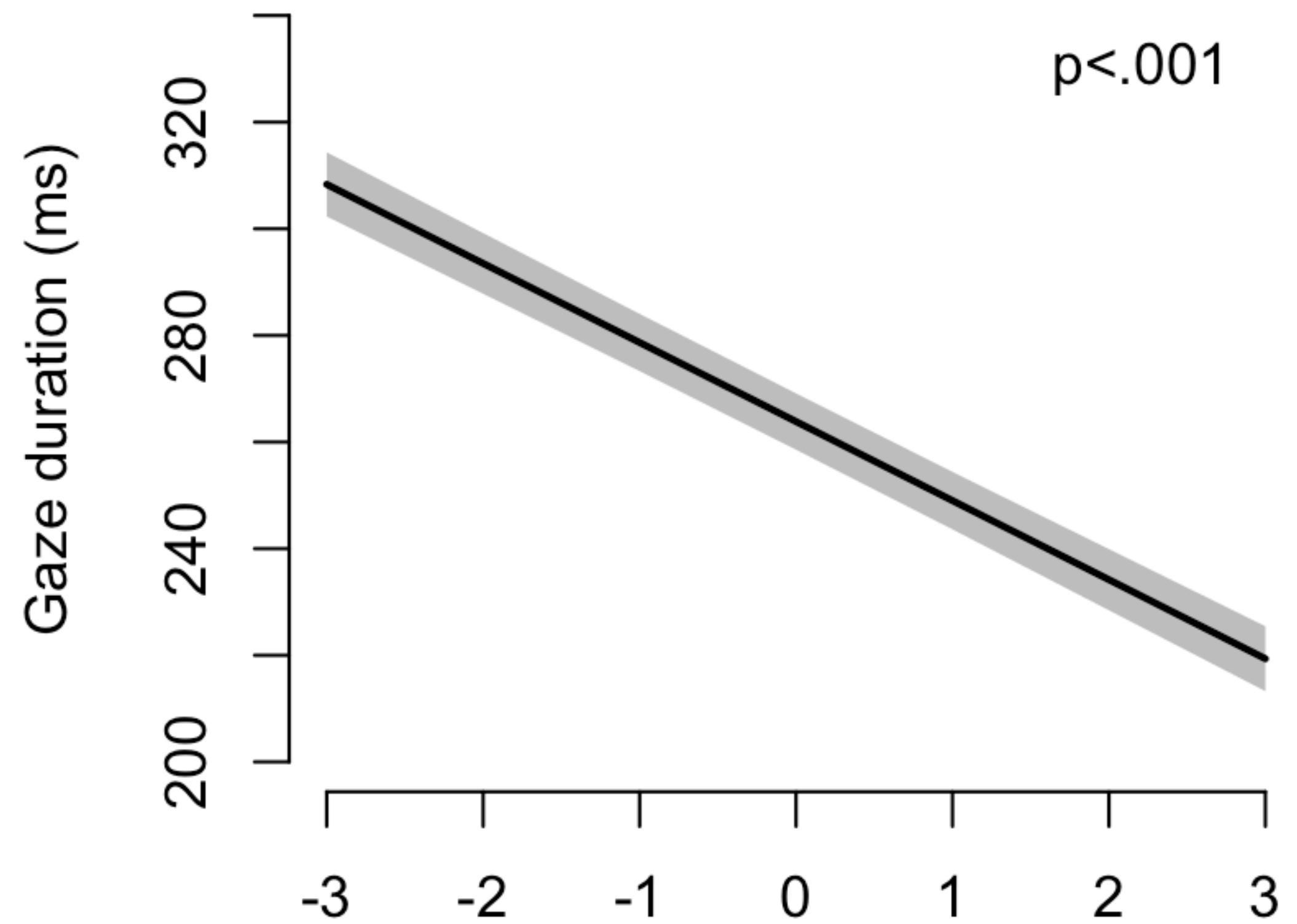
MECO

- 600ish adult readers
- 13 different L1 languages, English as L2

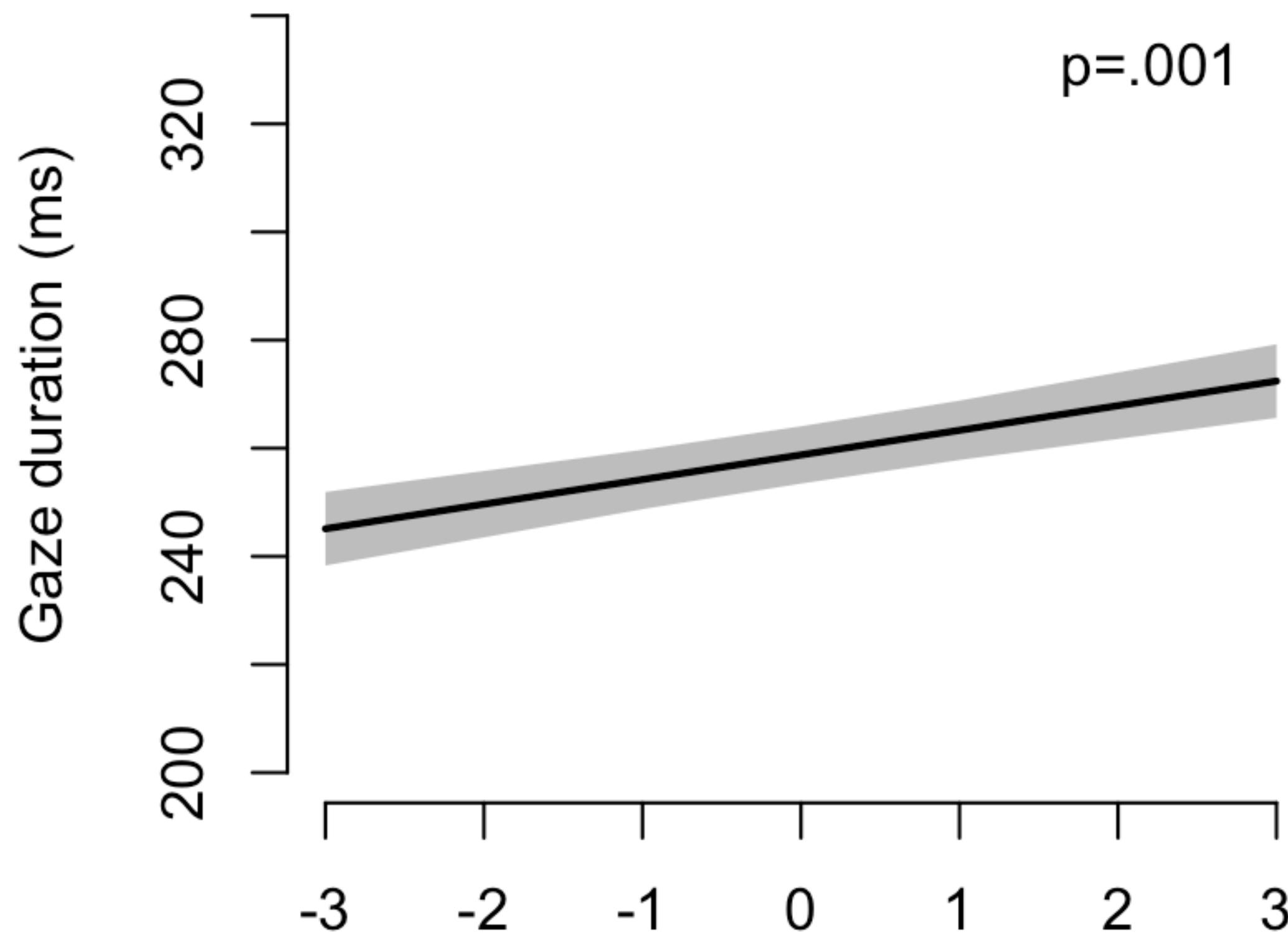
MECO

- 600ish adult readers
- 13 different L1 languages, English as L2
- 855K interest area data, nearly 1.2M fixations

(a) Word frequency



(a) Max 3gram frequency



What one can do with MECO (and similar stuff)

What one can do with MECO (and similar stuff)

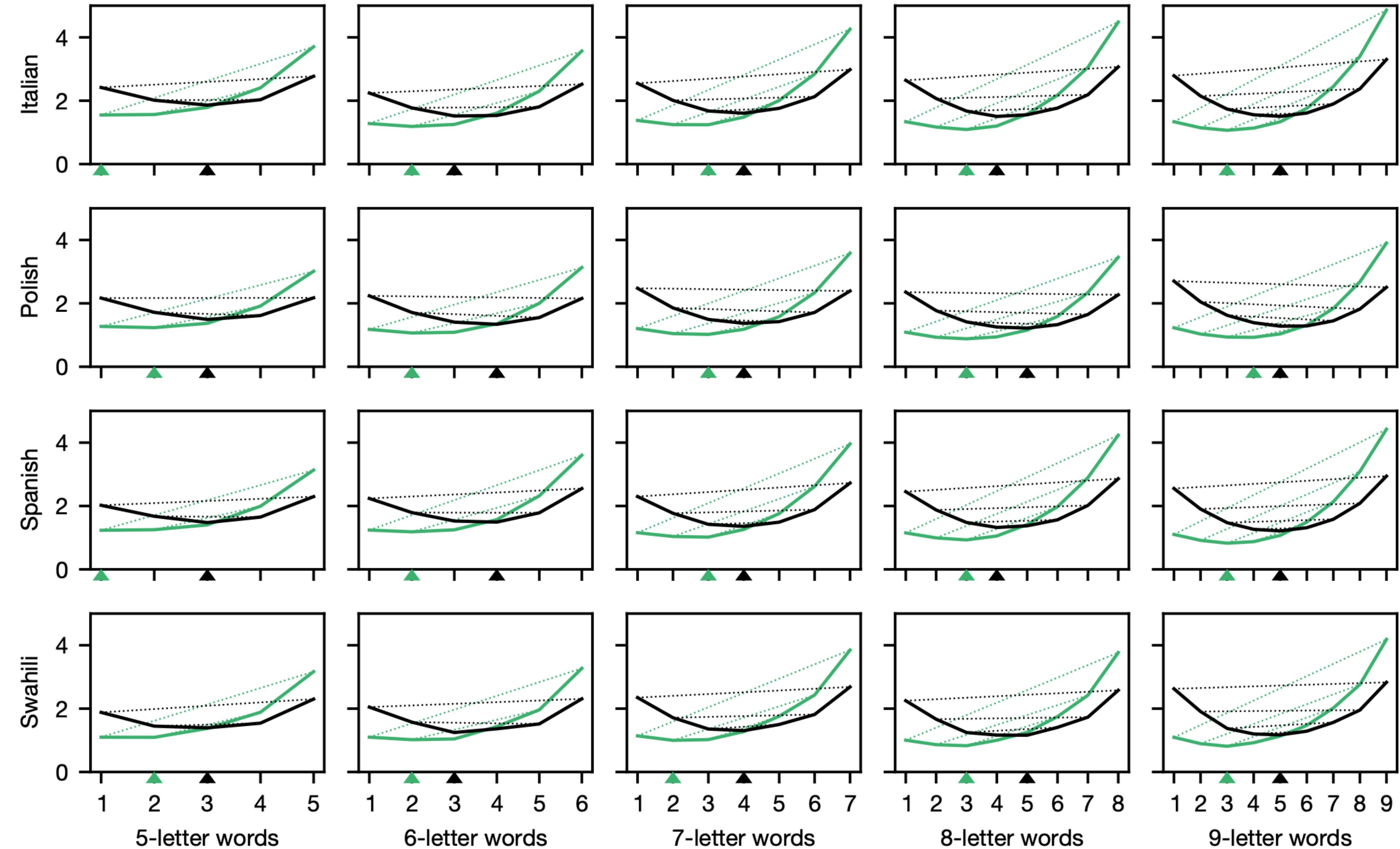
- Study skipping behaviour in L2, as a function of L2 properties vs. L1 behaviour

What one can do with MECO (and similar stuff)

- Study skipping behaviour in L2, as a function of L2 properties vs. L1 behaviour
- Information theory analysis of nGram distribution across languages, and whether/how this affects text scanning

What one can do with MECO (and similar stuff)

- Study skipping behaviour in L2, as a function of L2 properties vs. L1 behaviour
- Information theory analysis of nGram distribution across languages, and whether/how this affects text scanning
- Distribution of information within words



So, what did I learn?

So, what did I learn?

- It's a hell of an effort, as a promoter/coordinator...

So, what did I learn?

- It's a hell of an effort, as a promoter/coordinator...
- ...but not so much as a participant (**networking**)

So, what did I learn?

- It's a hell of an effort, as a promoter/coordinator...
- ...but not so much as a participant (**networking**)
- Quality/quantity trade off

So, what did I learn?

- It's a hell of an effort, as a promoter/coordinator...
- ...but not so much as a participant (**networking**)
- Quality/quantity trade off
- Project management tools, coding toolbox/culture