

Online Experiments for Language Scientists

Lecture 4: Self-paced reading

Kenny Smith

kenny.smith@ed.ac.uk

Matter arising from the grammaticality judgments lab: office hour

- I am setting up a 1-hour weekly slot when I can be available for extra help (for if you are behind on labs only!) – exact time TBC, looking like **Thursday at 10am**, I will confirm

Matter arising from the grammaticality judgments lab: prompt position

- “How do I make the prompt appear **above** the buttons on an html-button-response trial?”

Reminder about Undergraduate Assessment 1

- Due on Thursday 12th November
- If you haven't already, read:
 - The assignment brief
(<https://kennysmithed.github.io/oels2026/assessment/UGAssignmentBrief2026.pdf>)
 - The FAQ
(https://kennysmithed.github.io/oels2026/assessment/oels_assignment_faq.html)
 - The generative AI policy
(<https://kennysmithed.github.io/oels2026/assessment/GenerativeAIPolicy.pdf>)
- No questions after 11am on Monday 9th November!
 - People with extensions please extend this deadline appropriately
- **For postgrad students only (i.e. MSc, PhD students):** this assignment is optional and formative, i.e. doesn't contribute to your grade

Enochson & Culbertson (2015)

Enochson, K., & Culbertson, J. (2015).
Collecting Psycholinguistic Response Time
Data Using Amazon Mechanical Turk. *PLoS
ONE*, 10, e0116946.

Three self-paced reading experiments on
MTurk

- Do small but meaningful RT differences
seen in lab studies replicate online, despite
reduced experimental control and increased
variability in e.g. participant hardware?



Kelly Enochson
(formerly George
Mason University)



Jenny Culbertson
(Edinburgh)

Self-paced reading

Demo with this week's lab code

Sample size, study duration etc

- Self-reported native speakers of English
- N=34 (Exp 1), 82 (Exp 2), 60 (Exp 3)
- 96-120 items per experiment (mainly fillers)
- 20 minutes, **\$1**

Test items and predicted effects (Experiment 1)

Filler-gap (in all sentences)

- ***Which antique** was the maid polishing ____ in the study?*

Full DP vs pronoun

- *Which antique was **the maid** polishing in the study?*
- *Which antique was **she** polishing in the **upstairs** study?*

Agreement attraction

- Which antique** was the maid polishing in the study?*
- ?**Which antiques** was the maid polishing in the study?*

Residual reading times

You would expect e.g. word length and frequency to influence reading time

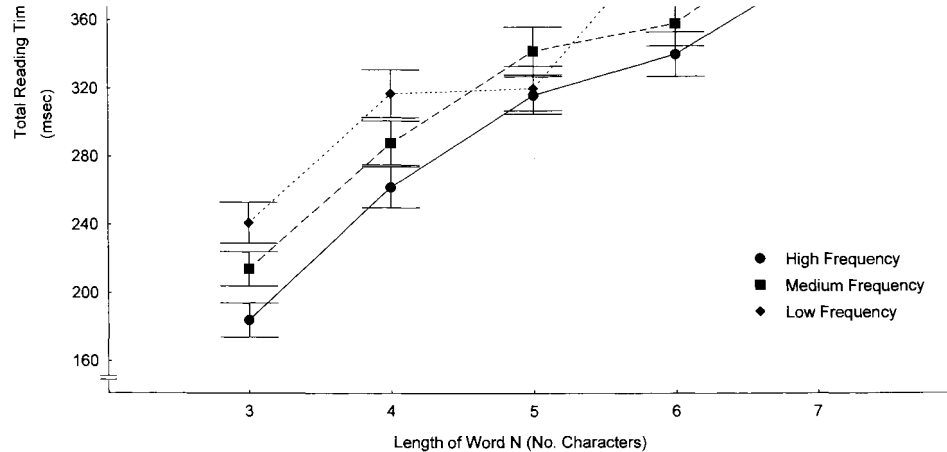


Fig. 1. Effects of length and frequency on word *N* reading time.

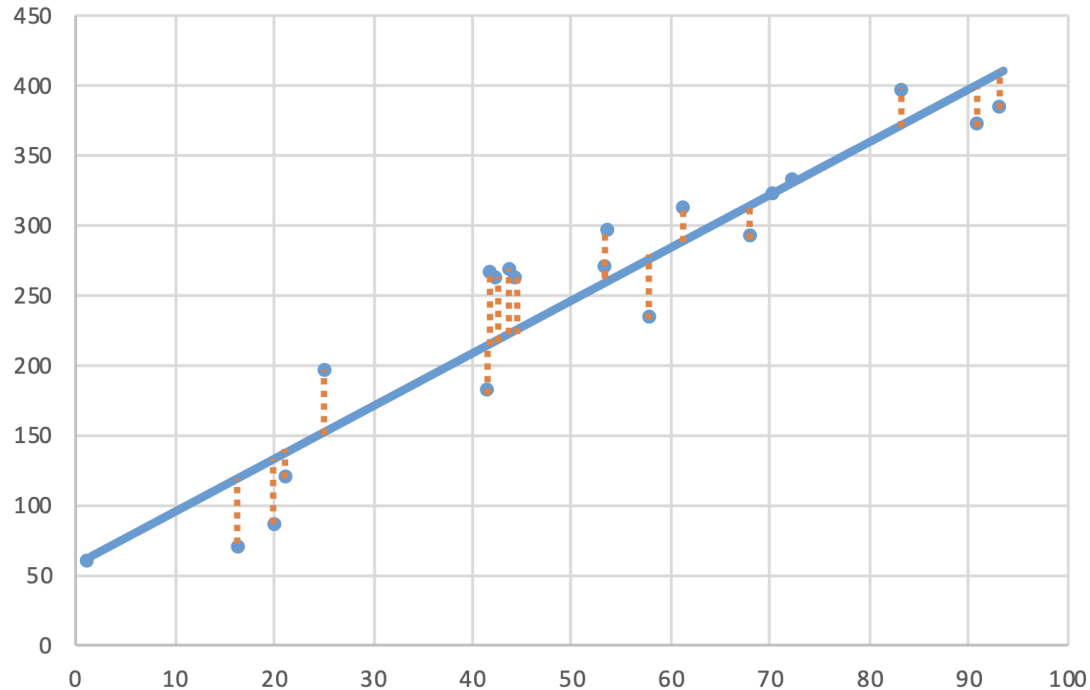
Target word (n) reading times: effects of length and frequency

The mean total reading time for the target word (*n*) as a function of its length and frequency is shown in Fig. 1.

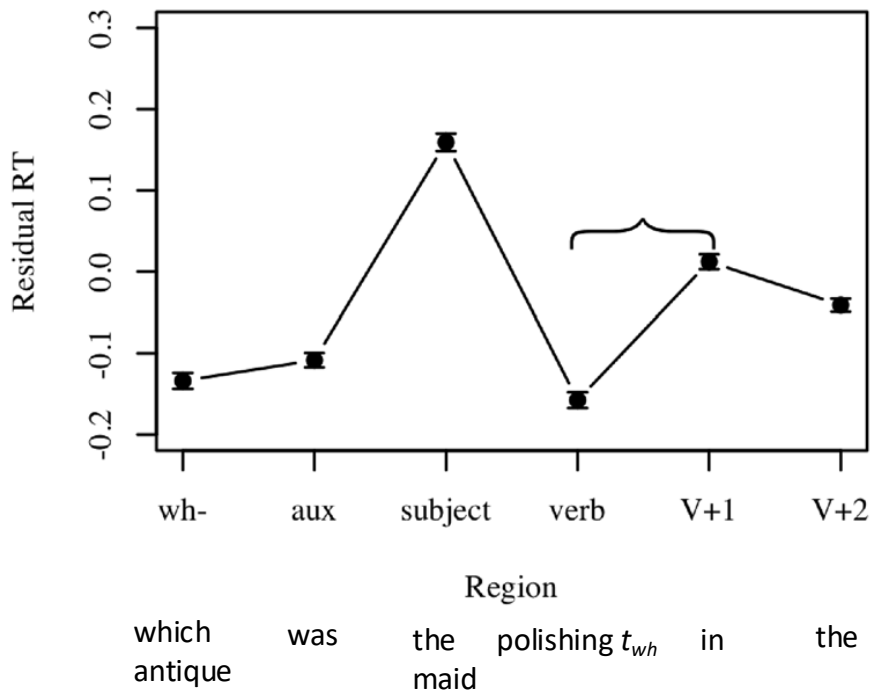
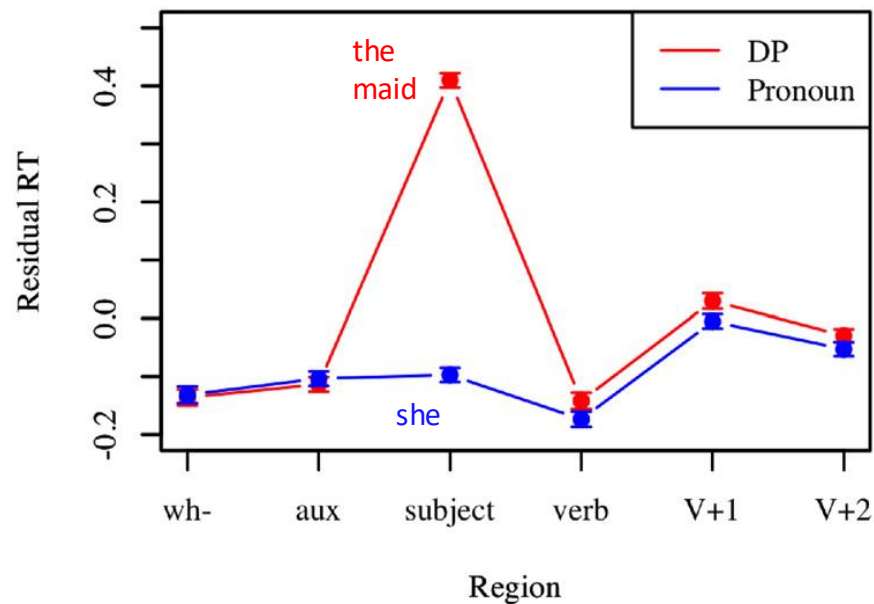
From Underwood, G., Binns, A., & Walker, S. (2000). Attentional Demands on the Processing of Neighbouring Words. In I. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a Perceptual Process* (pp 247-268). Amsterdam: North-Holland.

Residual reading times

Regression line – line of best fit, **minimising residuals**

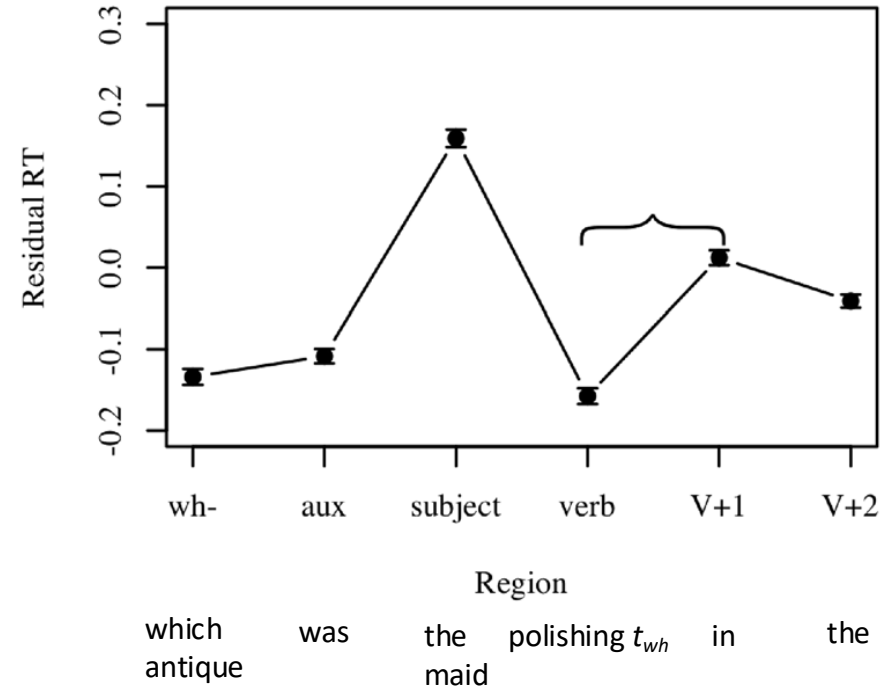


Exp 1 results



A question from the pre-lecture form

“In the filler-gap experiment, I noticed in Figure 3 that reading times fluctuate substantially from the auxiliary to the subject region, and again from the subject to the verb region. Why, then, is the effect of interest defined as the difference between the verb region and the V+1 region?”



A question from the pre-lecture form

“Why didn’t Enochson & Culbertson successfully capture agreement attraction in grammatical wh-fronted constructions? What do you think about that?”

Do any other researchers try to verify this effect using online crowdsourcing tools?”

Agreement attraction

Which antique was the maid polishing in the study?

?Which antiques was the maid polishing in the study?

Test items and predicted effects (Exps 2-3)

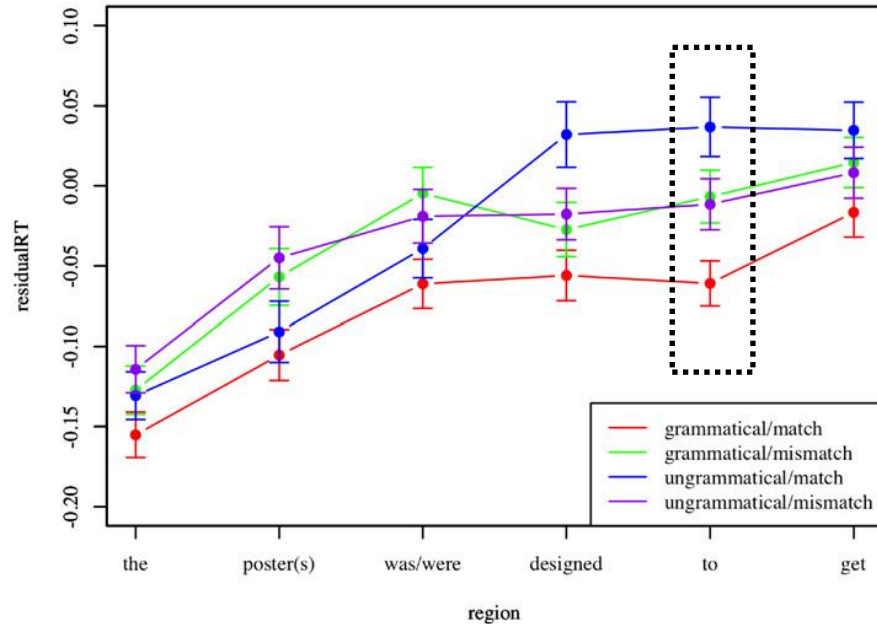
Experiment 2

- *The slogan [on the **poster**] was designed to get attention*
- *?The slogan [on the **posters**] was designed to get attention*
- **The slogan [on the **poster**] were designed to get attention*
- **?The slogan [on the **posters**] were designed to get attention*

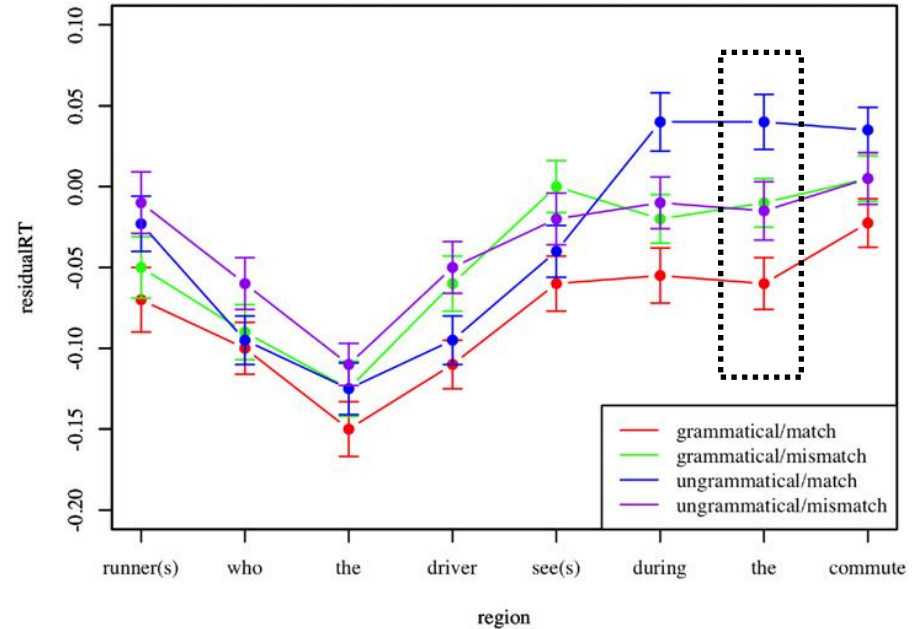
Experiment 3

- ***The runner** [who the driver sees during the commute]...*
- *?**The runners** [who the driver sees during the commute]...*
- ****The runner** [who the driver see during the commute]...*
- **?**The runners** [who the driver see during the commute]...*

The slogan on the poster was designed to get attention
?The slogan on the posters was designed to get attention
**The slogan on the poster were designed to get attention*
**?The slogan on the posters were designed to get attention*



The runner who the driver sees during the commute...
?The runners who the driver sees during the commute...
**The runner who the driver see during the commute...*
**?The runners who the driver see during the commute...*



Enochson & Culbertson's conclusions

MTurk is suitable for collecting reading-time data in self-paced reading tasks

- Similar patterns of results to those seen in lab tasks
- (Paper also includes lab replication of Exp 1)

They also make some suggestions re. batch sizes that I don't necessarily agree with – see my reading notes!

Additional time for Q&A/discussion on this week's reading

Next up

Wednesday lab

- **Venue change this week and next! G.06, 50 George Square**
- Our second proper experiment: self-paced reading
- If you are behind, do your best to get caught up before the lab and bring problems for us to help with!

Next week:

- Word learning