Capturing psycholinguistic processing effects using Amazon Mechanical Turk

Kelly Enochson and Jennifer Culbertson
George Mason University

Summary

Traditional lab-based model for psychology research:
- Resource heavy (infrastructure, subject compensation, time)
- Narrow population (undergraduate students)

New tools: Online services (Crump, McDonnell & Gureckis 2013)
- Less resource instensive (much less infrastructure, money, time).
- More diverse population
- Example: **Amazon Mechanical Turk** (MTurk)
  - Online crowd-sourcing marketplace
  - Post experiments, automatically recruit and compensate participants

Application to psycholinguistics:

Q: Can MTurk be used to measure psycholinguistic effects? With its more diverse population…? Including when **accurate measurements** are crucial…?

A: **Yes.** We replicate three robust psycholinguistic effects requiring precise reading-time measurements on MTurk.

Introduction

Method: Web-based Reading Time (RT) measurements
- ***ScriptingRT*** (Schubert, Murteira, Collins & Lopes 2013)
- Flash-based software embedded in html page
- Captures response times accurately over the web

General Goal: replicate three robust psycholinguistic effects

1. *Subject definiteness* ⟹ pronouns processed faster than DPs (Warren & Gibson 2002; Hofmeister & Sag 2010)

2. *Filler-gap effect* ⟹ processing cost of filler-gap dependencies (Wanner & Marastos 1978, Stowe 1986, Crain & Fodor 1985, Clifton & Frazier 1988, Boland, et al., 1995)

3. *Agreement attraction* ⟹ missing cost of processing spurious (ungrammatical) agreement between plural non-subject DP and verb in:

Prepositional phrases (Bock & Miller 1991, Pearlmutter et al. 1999)
Relative clauses (Bock & Miller 1991, Wagers et al, 2009)
Wh-fronted constructions (Badecker, MS)

Experiment 1

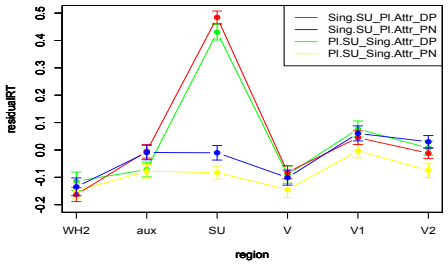
Participants
- 34 adults recruited via MTurk
- Compensated $1 each (for approximately 20 minutes of work)

Materials
- 48 test items, all grammatical *wh*-questions (Badecker MS)
  e.g., *Which cars has the salesman found easy to sell?*
- 72 fillers
- Presented in randomized order (using Scripting RT functionality)

Manipulation

	Wh-Attractor	
Subj,Verb	SG.	PL.
SG.	Which X has he/DP verbed t_{wh}?	Which Xs has he/DP verbed t_{wh}?
PL.	Which X have they/DP verbed t_{wh}?	Which Xs have they/DP verbed t_{wh}?

Results



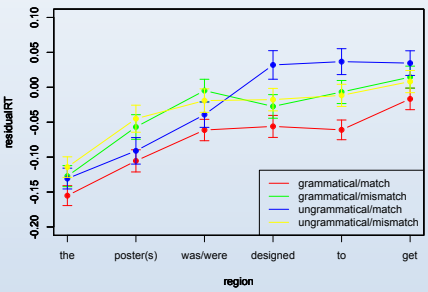
1. Our results show a significant processing cost associated with reading a pronoun as compared to reading a DP ($\beta = -0.26 \pm 0.02$, $p < 0.001$). ***subject definiteness*** ✔
2. Our results show an increase in processing time at the region one word after the verb – where the gap is located ($\beta = 0.397 \pm 0.016$, $p < 0.001$). ***filler-gap effect*** ✔
3. Agreement attraction would present as faster reading times for singular subject / plural attractors than for plural subject / singular attractors. ($\beta = -0.017 \pm 0.053$, $p > 0.1$). ***agreement attraction*** ✖

Experiment 2

Magnitude of attraction effects is smaller in grammatical compare to ungrammatical sentences (Pearlmutter et al., 1999). Experiments 2 and 3 seek to replicate agreement attraction effects in classic **ungrammatical** constructions:
- **Experiment 2: PP modifiers** (Pearlmutter et al. 1999 experiment 1)
- **Experiment 3: RC modifiers** (Wagers et al. 2009 experiment 2)

Participants and Materials
- 82 participants (cf. 80 in Pearlmutter et al., 1999)
- 16 test items, half ungrammatical
  e.g., *The slogan on the poster(s) was/were designed to get attention.*
- 96 fillers

Results

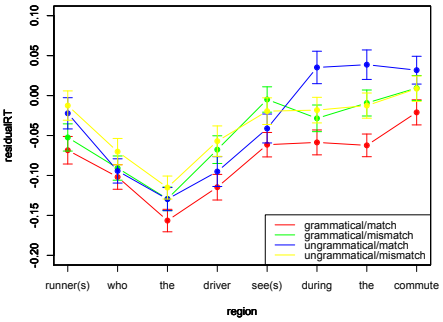


Experiment 2 successfully replicated the attraction effect resulting from a PP modifier ($\beta = -0.09 \pm 0.03$, $p = 0.003$).
Agreement attraction – PP modifier ✔

Experiment 3

Participants and Materials
- 60 participants (Wagers et al. used 30)
- 24 test items, half ungrammatical (Wagers et al. used 48)
  e.g. *The runner(s) who the driver see(s) during the commute…*
- 72 fillers

Results



Experiment 3 successfully replicated attraction effects with RC modifiers ($\beta = -0.02 \pm 0.008$, $p = 0.001$).
Agreement attraction – RC modifier ✔

Conclusions

Mturk and psycholinguistic research
Several robust processing effects were replicated using similar numbers of trials and participants as traditional lab studies.

✓ ***subject definiteness***
✓ ***filler-gap effect***
✘ ***agreement attraction – (grammatical) wh-fronted questions***
✓ ***agreement attraction – PP modifiers***
✓ ***agreement attraction – RC modifiers***

Resources required
- Experiments typically completed within a few days of posting
- Cost: $1.30 per participant

References

Badecker, W. (MS) Agreement in *Wh*-Questions: Subject-oriented Retrieval Mechanisms in a Working-Memory Retrieval Model of Sentence Production.

Pearlmutter, N., Garnsey, S., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language* **41,** 427–456.

Schubert TW, Murteira C, Collins EC, Lopes D (2013) ScriptingRT: A Software Library for Collecting Response Latencies in Online Studies of Cognition. *PLoS ONE* 8(6): e67769. doi:10.1371/journal.pone.0067769

Wagers, M, Lau, E., and Phillips, C. (2009). Agreement Attraction in comprehension: Representations and processes. *Journal of Memory and Language,* **61,** 206-237.

Contact Information

Kelly Enochson
kenochso@gmu.edu
mason.gmu.edu/~kenochso

Jennifer Culbertson
jculber4@gmu.edu
mason.gmu.edu/~jculber4