9.59J / 24.905 Laboratory in Psycholinguistics (CI-M)
2017

Instructor: Ted Gibson

pre-requisite: 9.00

Other useful classes: 24.900;

Probability and Statistics

Goals

Students will learn to:

1. read the primary literature in language research and design informative experiments;

2. present research orally;

3. write a research paper;

4. gather data on the crowd-sourcing website Mechanical Turk (run by Amazon.com);

5. analyze their experiments using the R programming language

Class 1: Overview

- Mechanical Turk demonstration in class

- Syllabus / requirements information

- The usefulness of quantitative methods in linguistics

- Language information sources and constraints
  - Lexicon; syntax; world knowledge; pragmatics; prosody; working memory; context

Mechanical Turk in class

1. Set up the materials in turkolizer format (Gibson, Fedorenko & Piantadosi (2011)

Particle-shift and length: Locality of syntactic dependencies
Verb-particle shift (-, +) x Length (short, long)

Short NP object:
Local Particle: Joe threw out the documents.
Non-local Particle: Joe threw the documents out.

Long NP object:
Local Particle: Joe threw out the very important documents that he brought home. *
Non-local Particle: Joe threw the very important documents that he brought home out. *

Predictions of dependency locality:
No difference for short NP object; Big difference for long NP object

Mechanical Turk in class

Particle-shift and length: Dependency-locality

particle-length 1 loc-particle-short

Joe threw out the documents.

? Did Joe throw out the documents? Yes

particle-length 1 nonloc-particle-short

Joe threw the documents out.

? Did Joe throw out the documents? Yes

particle-length 1 loc-particle-long

Joe threw out the very important documents that he brought home.

? Did Joe throw out the documents? Yes

particle-length 1 nonloc-particle-long

Joe threw the very important documents that he brought home out.

? Did Joe throw out the documents? Yes

Mechanical Turk in class

2. Run the turkolizer program on the formatted items

produces a .turk.csv file:

- turk materials: look at this file
- Includes information for decoding later:
    a.   linking the presentation order to the appropriate conditions
    b.   correct answers to comprehension questions

Mechanical Turk in class

3. Set up the Turk template in Mechanical Turk "Create"

•Instructions

•Format for ratings (1 - 5 or 7)
•Rating then question(s)

Mechanical Turk in class

4. Post the survey

•Go to M Turk: go to Create
•New Batch with an existing project

5. Wait 45 – 60 mins

6. Get the data from "Manage":
•Download data

Topics and Texts

Constraints on language processing: Lexicon, syntax, world knowledge, pragmatics	The R programming language
Language as communication: words, syntax	Descriptive statistics & plotting
The domain specificity / generality of language	Significance testing & confidence intervals
Pragmatics of language use	Regression and mixed effects regression
Behavioral methods and issues in experimental design	Amazon.com's Mechanical Turk

Readings: *Analyzing Linguistic Data*, Harald Baayen textbook; Wickham & Grol (2016). *R for Data Science;* 2-3 journal papers every class; 10-15 for final paper. (Approximately: 50 total)

Assignments and grading

4 programming assignments in R:	15%
16 discussion notes:	15%
Paper 1: Evaluation of an experiment in the literature (40% for draft 1; 60% for draft 2):	10%
Paper 2: Project proposal, plus bibliography:	5%
Oral presentation of proposed project:	5%
Paper 3: Replication writeup:	10%
Oral presentation of final project:	10%
Paper 4: Final paper:	20%
Class/lab participation (obligatory attendance):	10%

Late policy: 10% off each day late, down to 50% off. Then you can hand it in later for 50% credit. (If you contact us ahead of time, with reasons for needing to be late, then we may be able to avoid the deductions entirely)

Psycholinguistics Lab topic areas

- Language above the word level
  - Lexicon, syntax, semantics, pragmatics, discourse, world knowledge, working memory constraints
- Language as communication: Language as rational inference
- The domain specificity / generality of language

Class 1: Overview

- Mechanical Turk demonstration in class

- Syllabus / requirements information

- The usefulness of quantitative methods in linguistics

- Language information sources and constraints
  - Lexicon; syntax; world knowledge; pragmatics; prosody; working memory; context

Preliminaries:
Properties of Human Language

(1) Discreteness / hierarchical structure

- Language is made up of little units that combine to make bigger units

Unit	Number	Meaning
phonemes	20-40	none
morphemes	10,000+	single chunk
words	50,000+	simple combinations
sentences	infinite	complete thought

Preliminaries:
Properties of Human Language

(2) Productivity - Language is not just a memorized set of sentences

Colorless green ideas sleep furiously.
*Furiously sleep ideas green colorless.

➢ There are *rules* that govern sentence structure

Preliminaries:
Properties of Human Language

What kind of rules?   Not prescriptive rules.

Prescriptive rules: What an old-fashioned English teacher might tell you

1. *"Don't say 'ain't' "*

2. *"Don't end sentences in prepositions"*

   Counterexamples:
   - *That is something which I cannot put up with.*
   - **That is something up with which I cannot put.*

3. *"Don't split infinitives"*
   Counterexamples:
   - *To boldly go where no man has gone before.*
   - *? To go boldly where no man has gone before.*
   - *? Boldly to go where no man has gone before.*

Preliminaries:
Properties of Human Language

Descriptive Rules: Rules obeyed implicitly.

Sentence formation rules:
S → NP VP
"A sentence (S) consists of a noun phrase (NP) and a verb phrase (VP)"
VP → V NP
VP → V that S

The Red Sox beat the Yankees.
Mario said that the Red Sox beat the Yankees.
Jill thought that Mario said that the Red Sox beat the Yankees. ...

*(Note: this is a **recursive** rule: The category S expands to another S further along)*

Differences between 24.900 (intro to linguistics) and 9.59/24.905 (this class)

- Methods / evidence:
  - 24.900: within some domains (syntax & semantics): acceptability judgments on a few individuals
  - 9.59/24.905: experiments using many items and participants; many different dependent measures:
    - Acceptability judgments (how good does this sound?);
    - Accuracy on comprehension questions;
    - Sentence completions;
    - Reaction times:
      - Reading times;
      - Looking times to visual scenes, given auditory input;
      - Lexical decision times
    - Brain imaging:
      - Event-related potentials (EEG)
      - Magneto-encephalography (MEG)
      - Functional MRI

When do we need an experiment?

- Acceptability ratings ("How natural does this utterance sound?")
    - ➢ E.g., compare:
        - • "the cat"
        - • "cat the"

- Do we need an experiment to decide that determiners / articles come before the noun?

- How could we convince ourselves without an experiment that this was the case?

- What if you didn't speak the language? Wouldn't you want some quantitative data to provide evidence?

- In addition, most current theoretical questions depend on more complex examples, where the judgments aren't so clear

Behavioral measures *

- Ratings ("How natural / normal does this utterance sound?")
  - ➢ E.g., compare:
    - • "Mary wondered who bought what."
    - • "Mary wondered what who bought."
    - • "Mary wondered what who bought when."

- Careful about the notion of "grammatical" vs. "ungrammatical": is there a binary choice? Or is it continuous?
- Reading times / reaction time
- Response accuracy to questions about the content of a sentence

  - ➢ All of these measures are **noisy**, probably because there are many factors that contribute to them, and we are generally only investigating one or two

  - ➢ That's why we do statistical analyses of the behavioral results: If there are reliable differences in the measure across materials and participants, then the factor in question may affect the dependent measure

  - ➢ But of course there can always be other confounding factors that we didn't consider: *it's difficult to design good experiments*

Syntax & Semantics

- Standard method in the field of syntax:
  - ➤ Acceptability judgment method
    - Single-subject / single-item

  - ➤ **Weaknesses** (Schütze, 1996; Cowart, 1997; Wasow & Arnold, 2005; Ferreira, 2005; Featherston, 2007; Myers, 2009; Gibson & Fedorenko, 2010, 2011; Gibson, Fedorenko & Piantadosi, 2013)

    - small number of experimental participants (typically 1);
    - small number of experimental stimuli (typically 1);
    - cognitive biases on the part of the researcher and participants

Syntax & Semantics: Quantitative methods

- The advantages of quantitative methods (controlled experiments or corpus analyses)

  - ➢ enable the use of inferential statistics to evaluate the likelihood of particular hypotheses;
  - ➢ experimental participants are naïve with respect to the hypotheses;
  - ➢ experimental materials are presented in such a way so as to avoid context effects (i.e. in a random order, varying orders across participants).

Mechanical Turk in class

6. 45 mins later: download the data from "Manage"

7. Analyze data: edit an R analysis file

•Check quality of participant work (and if they have done previous surveys with the same name): look at variable data_summ

•Possibly reject bad participants (to ensure the quality of the participant pool)

8. Plot the results from each experiment: look at means / variance / individual data

9. run statistics (if needed)

9.59J/24.905J Lab in
Psycholinguistics Spring 2017

For information about citing these materials or our Terms of Use, visit: