UNIVERSITY OF BIRMINGHAM | COLLEGE OF ARTS AND LAW

Experimental methods and behavioral data in linguistic research

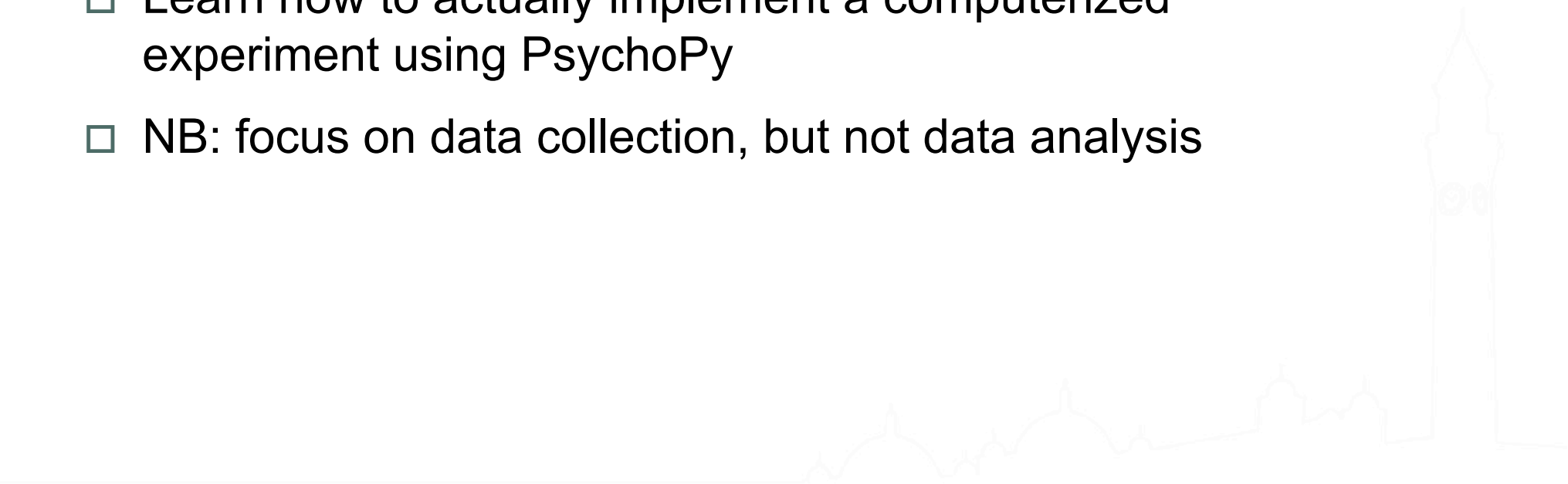
Florent Perek

Tutorial at ICCG-10, Paris

19 July 2018

Aims

- Understand what experiments are and how they relate to other methods
- Review the use of experiments in linguistics, and describe several common experimental designs
- Learn how to actually implement a computerized experiment using PsychoPy
- NB: focus on data collection, but not data analysis



Programme

- Session 1: Designing experiments
- Session 2: A review of some experimental methods
- Session 3: Introduction to PsychoPy

- All materials downloadable at:

  http://www.fperek.net/tutorial.html#four

- For the hands-on session: please download and install PsychoPy from www.psychopy.org

UNIVERSITY OF BIRMINGHAM | COLLEGE OF ARTS AND LAW

Designing experiments

Florent Perek

Tutorial at ICCG-10, Paris

19 July 2018

Overview

- What are experiments?
- How do they relate to other methods?
- Why are they useful?
- Basic principles of experiment design
- Terminology

Research methods

- Main purpose of research = deriving new knowledge
- Deduction: using existing knowledge to infer particular facts (through logic)

  All men are mortal. Socrates is a man.

  Therefore, Socrates is mortal.
- Induction: using observation to make generalizations

  All life forms we know of are water-based.

  Therefore water is necessary for life to exist. (probably)

The scientific method

- Modern science is mostly inductive
- Aim = find the best possible explanation for a phenomenon

  NOT find the "truth"!
- The scientific method = test whether an explanation is supported by empirical observations

The scientific method

"[T]he best we can do is to develop possible explanations supported by empirical findings. The goal of science is then to find the best available explanation for a phenomenon, with 'best' qualified as being the least likely to be false. This is not the same as saying that it is most likely the correct explanation, but only that it is better than any other under consideration. It also means that it too can be dislodged the moment another explanation is found that better accounts for the target phenomenon."

Gonzalez-Marquez et al. (2007: 59-60)

Gonzalez-Marquez, M., Becker, R., & Cutting, J. (2007). An introduction to experimental methods for language researchers. In Gonzalez-Marquez, M., Mittelberg, I., Coulson, S., & Spivey, M. (eds.), Methods in Cognitive Linguistics (pp. 53-86). Amsterdam: John Benjamins.

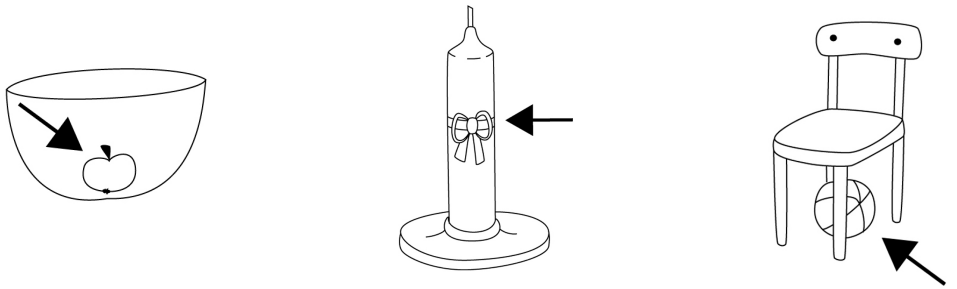
Observation

- Observations can be natural or elicited
- Natural: happen spontaneously, they would have occurred even without the research

  Example in linguistics: corpora

- Elicited: data produced specifically for the research, requested and controlled by the researcher

  Example in linguistics: picture naming in typological fieldwork



Natural observation

- Main advantage: produces fully authentic data
  - For language: reflects “normal” linguistic behavior of speakers
- Potentially avoids the observer’s paradox
  - No aspect of the data can be an artefact of data collection
- But no control over the data
  - You only get spontaneously occurring events
  - Less appropriate to study rare events

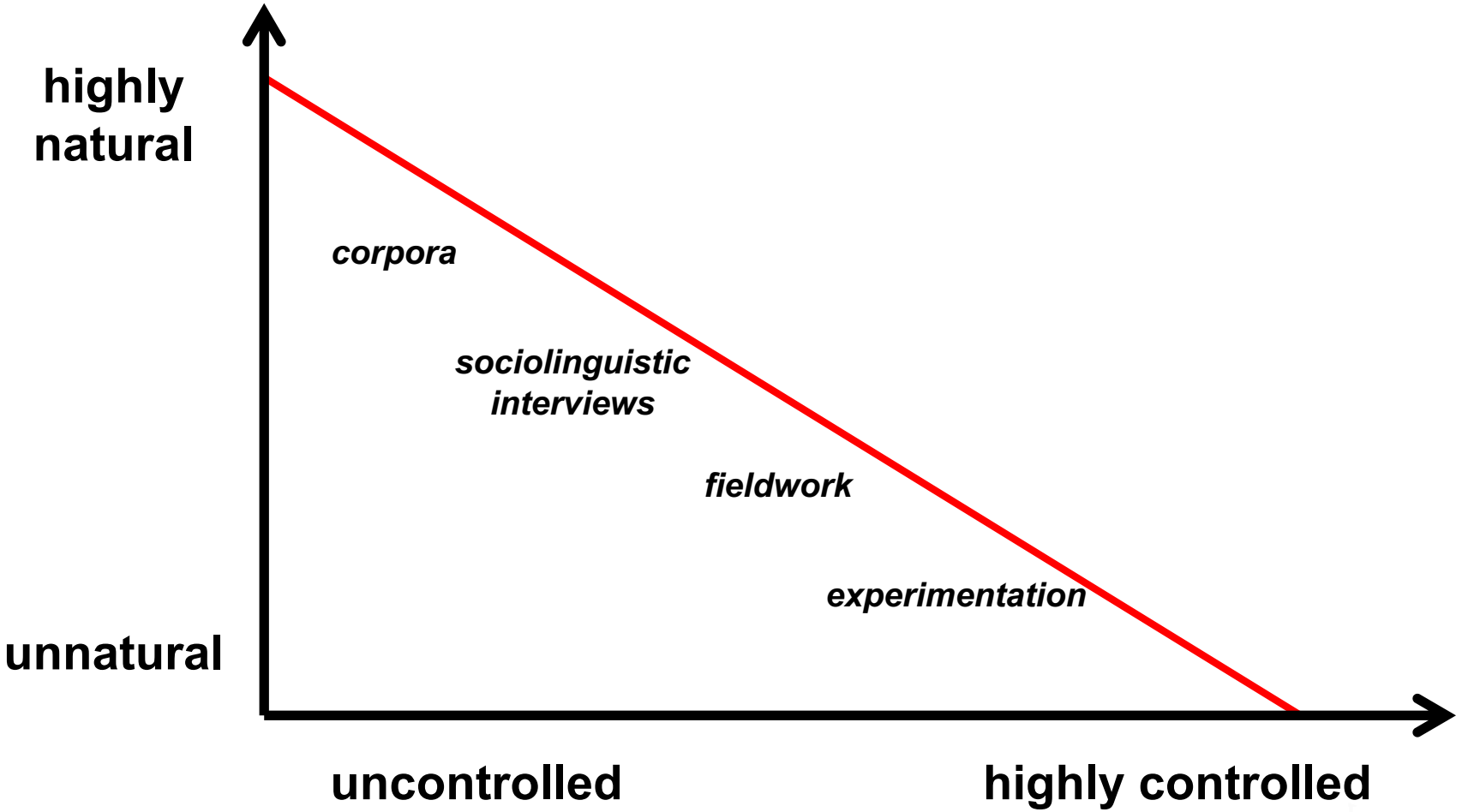
Problems with natural observation

- Induction suffers from epistemic uncertainty
  - Dependent on the data: a theory can only be assumed to be true until it is proven wrong by observations
  - But natural observation can never be considered exhaustive
- One solution = to *cause* data to occur: elicitation
  - Produces data that might not easily be found naturally
  - Allows for control, i.e., to manipulate the conditions under which a certain behavior is expected to occur (or not)

Natural observation vs. elicitation

- In linguistics, natural vs. elicited observation is not a dichotomy but rather a cline
- Always a trade-off between naturalness and control
  - More natural data allow for less control
  - Conversely, more control entails less natural conditions
- Natural and elicited observation should be seen as complementary; not an either/or choice

The naturalness-control cline



Cf. Gilquin, G. & S. Gries (2009). Corpora and experimental methods: a state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1). 1-26.

UNIVERSITY OF BIRMINGHAM | COLLEGE OF ARTS AND LAW

Experimentation in linguistics

What's an experiment?

"A scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact."

Oxford Dictionary of English

"The researcher manipulates an independent linguistic variable to control some aspect of language processing and then measures the effect of the manipulation on a dependent variable of interest."

Garrod (2006: 251)

Garrod, S. (2006). Psycholinguistic Research Methods. In Brown, K. (ed.), *Encyclopedia of Language and Linguistics, 2nd Edition* (pp. 251-257). Oxford: Elsevier.

Experimentation

- The most controlled form of data elicitation
- Complementary to more natural data (e.g., corpus data)
- Often necessary because factors are often confounded in natural observation

Experiments in linguistics

- Primary method of data collection in psycholinguistics
  - Psycholinguistics is interested in how language is represented in the mind and how it is processed
  - Natural linguistic productions contain little trace of this
  - Some aspects of these questions can only be studied in highly controlled conditions and with adequate equipment
  - Methods from cognitive psychology
- Now used in many other areas

Typical properties of experiments

- Protocol

  Carefully planned procedure applied consistently

- Manipulation

  Different conditions created by the experimenter

- Control

  Minimize the influence of variables external to the explanation

- Sampling

  Carry out the procedure on a group of representative subjects (in behavioral sciences: human participants)

(NB: so-called “thought experiments” are NOT experiments!)

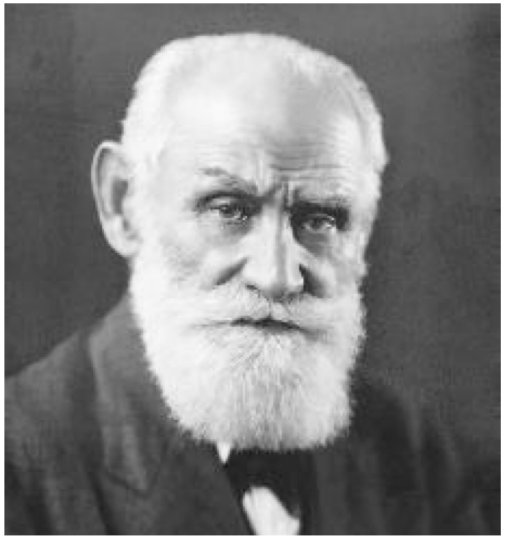
Designing an experiment

- Experiment = a procedure designed to test a hypothesis
- First step: formulate a testable hypothesis from the research question
- The hypothesis can usually be defined at two levels
  - Research hypothesis = what we wish to test
  - Experimental hypothesis = what we measure
  - Concept vs. operationalization

Research hypothesis

- More specific than research question
  - Should be a yes/no question
  - Often the result of an exhaustive literature review
- Formulated as a relationship between two (or more) variables
- Makes predictions about this relationship (if → then).
- Must be of explanatory value
- Must be falsifiable

Example: Pavlov's dog

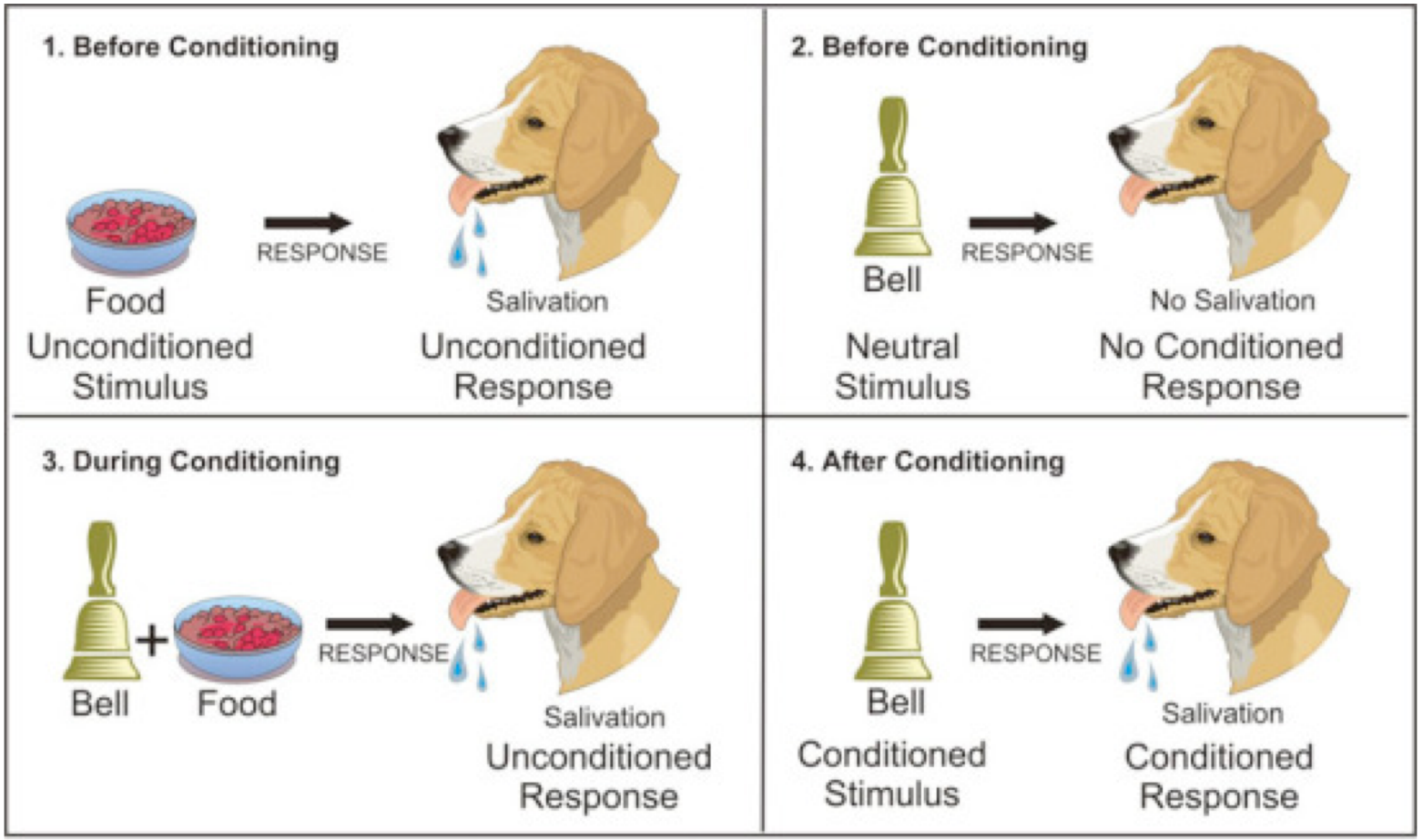


Ivan Pavlov (1849-1936)

Pavlov's research question: can a dog learn to produce a certain behavior as a reaction to some stimulus?

(aka. classical conditioning)

Pavlov's experiment



Source: http://fos.cmb.ac.lk/blog/wp-content/uploads/2017/07/pavlovdog.jpg


WATCH WHAT I CAN MAKE PAVLOV DO. AS SOON AS I DROOL, HE'LL SMILE AND WRITE IN HIS LITTLE BOOK.

Pavlov's research hypothesis

- Research hypothesis:

  If a dog is subjected to the repeated association between the unconditioned and the neutral stimulus, then it will learn to produce the natural response as a reaction to the unconditioned stimulus alone.

- Behavior = salivation (natural response to food = unconditioned stimulus)
- (Neutral) stimulus = bell ringing

Experimental hypothesis

- Research hypothesis = what we wish to test
- Experimental hypothesis = what we can measure
- Designed to verify or invalidate the research hypothesis
- Presented in the form of measurable variables
- Specifies a testable relationship between variables

Pavlov's experimental hypothesis

- Conditioning: the number of times the subject is exposed to food and bell ringing together
- Response: salivation, measured as the amount of secreted saliva

- Experimental hypothesis

  If we increase the number of exposures to the unconditioned stimulus (food) and the neutral stimulus (bell ringing), then salivation of the subject is increased.

Hypothesis testing

- For every experimental hypothesis, there is a corresponding null hypothesis
  - I.e., that the relationship between variables does not exist
  - And that any results obtained are thus due to chance
- The experiment tests the experimental hypothesis
  - Not by proving that it is true
  - But by proving that the null hypothesis is false

Hypothesis testing

- The experiment collects evidence against the null hypothesis
- Inferential statistics weigh the data against the hypothesis
  - Given the data, how likely is the null hypothesis?
  - Significance tests measure how likely results are due to chance
  - Expressed as a p-value; in psychology, 0.05 is the conventional significance threshold
  - If $p < 0.05$, then the null hypothesis is rejected
- If the null hypothesis is rejected, the experimental hypothesis is accepted as a better explanation of the data

Variables

- Measurable properties of an entity or phenomenon
- Experiments are attempts to inductively establish relationships between variables
- Different classifications of variables:
  - Nature: behavioral / stimulus / subject
  - Role: independent / dependent / extraneous / constant
  - Type (ontological): categorical / discrete / continuous / ordinal

Variables: nature

- Where does the variable come from? To what aspect of the experiment is it related?
- Behavorial variables: participants' responses

  Grammaticality rating, reaction time, eye movements, …
- Stimulus variables: factors that influence responses

  Word length, grammatical structure, frequency, …
- Subject variables: unchangeable properties of the participants

  Sex, age, country of origin, …

Variables: role

- How is the variable used with respect to the hypothesis
- Dependent variable(s) (usually just one)

  The “outcome” variable that is being observed

  E.g., time taken to recognize a word

- Independent variables (or predictors)

  Variables that are made to vary throughout the experiment and whose effect on the dependent variable is being considered (“manipulations”)

  E.g., word length, word frequency, age of acquisition, etc.

Variables: role

- Extraneous variables
  - Variables that might have an effect on the dependent variable but are not part of the hypothesis
  - They must be controlled (e.g., by balancing them or making them constant), or their effect cancelled
  - Can be factored out in the analysis using mixed effects regression modeling
- Constant variables

  Variables that do not vary throughout the experiment

Variables: role

- Distinction between extraneous and constant is not always clear and might require further investigation
- Dependent variable is usually a behavioral variable
- Independent variables can be stimulus variables or subject variables (or a combination thereof)
- Experimental hypothesis states a relation between the dependent variable and one or more independent variable(s)

Variables: type

- What type of entity is the variable itself?
- Categorical: falls into a limited number of qualitative values (called “levels”)

  e.g., person (1st, 2nd, 3rd), voice (active vs. passive), sex, native language, etc.

  Binary variables (true/false) are a subcase

- Discrete: a quantity that can only be a integer (i.e., a non-decimal number)

  e.g., frequency, word length (in letters), age, number of exposures to a stimulus, …

Variables: type

- Continuous: a quantity that can also be a decimal number

  e.g., reaction time, height

- Ordinal: categorical variable whose values are ordered but do not follow a regular scale

  e.g., differently sized frequency classes: 1-100, 100-1K, >1K

Variables: type

- The type of variable matters for statistical analysis!
- For a continuous (or discrete) dependent variable

  t-test (two groups), ANOVA (more than two), linear regression (recommended)

- For a categorical dependent variable

  Logistic regression (logit): preferably for a binary outcome; techniques exist for 3+ outcomes but they are more complex

- Cf Núñez (2007)

Núñez, R. (2007). Inferential statistics in the context of empirical cognitive linguistics. In Gonzalez-Marquez, M., Mittelberg, I., Coulson, S., & Spivey, M. (eds.), *Methods in Cognitive Linguistics* (pp. 53-86). Amsterdam: John Benjamins.

Operationalizing variables

- How are variables measured? Is it a good measurement?
- Concepts must be transformed into behaviors that are observable and measurable
- The criteria for operationalization must be transparent
- Coding criteria must be explicit and constant, and they must be described when reporting the experiment

UNIVERSITY OF BIRMINGHAM | COLLEGE OF ARTS AND LAW

Designing and running the experiment

Three components

- Designing an appropriate procedure to test the hypothesis
- Creating a set of stimuli used in the procedure
- Recruiting a suitable group of participants

Procedure

- Task(s) that will allow to measure the dependent variable while manipulating the independent variable(s)
- The procedure must be precisely defined, applied consistently every time, and reported in detail
- Consists of many **trials**, each producing one datapoint
- Trials of the same type can be organised into **blocks**

Procedure

- Between-subjects design

  Each subject is assigned to one group, and each group is tested in one condition

  In such a design, there tends to be a control group (if appropriate): a group that it is not submitted to manipulation

- Within-subjects design

  Each subject is tested multiple times in different conditions

- The two designs can be combined

  E.g., two groups with a different subject variable submitted to the same conditions

Discussion

What are the advantages/disadvantages of between-subjects vs. within-subjects designs?

Can you think of cases where you would *have* to use one or the other?

Stimuli

- Stimulus = variable part of each trial
- Stimuli operationalize the conditions of the experiment
- Each exemplify one value of every independent variable
- All variables must be exhaustively represented
  - Discrete/continuous variables: wide range of values, or values grouped into bins (e.g., low/mid/high)
  - Categorical variables: all combinations of levels
- Same number of items for each condition

Stimuli

Example: time to recognize a word as a function of ...

... Etymology (native vs. loanword) and morphological complexity (number of morphemes)

	1 mrph.	2 mrph.	3 mrph.
Native	5	5	5
Loanword	5	5	5

3 x 2 combinations
x 5 items
= 30 different stimuli

... Corpus frequency and morphological complexity

	< 100	100-500	500-2000	2000-20000
1 mrph.	4	4	4	4
2 mrph.	4	4	4	4
3 mrph.	4	4	4	4

3 x 4 combinations
x 4 items
= 48 different stimuli

Stimuli

- Choosing or creating stimuli can be constrained by the language itself (e.g., high-frequency trimorphemic words?)
- Extraneous variables should be considered and controlled
  - E.g., word length: should be kept constant or balanced both within each cell and across all cells
  - And/or: factored out as a random effect in the statistical analysis (mixed effects regression analysis)

Stimuli

- Filler items are stimuli that will not produce datapoints
- Not always necessary but they can serve as distractors from the stimuli of interest
  - Make the independent variable(s) less noticeable
  - E.g., if the stimuli are sentences with *to*-infinitive vs. gerund complements, fillers should use a range of other structures

Participants

- A representative, meaningful sample
- Compensation can be an incentive: course credit, cash, voucher, participation a in raffle, etc.
- Research with human subjects is usually restricted
  - Legislation depends on the country
  - You may have to request permission (e.g., IRB in the US)
  - Usually three principles of ethics: well-being, informed consent, confidentiality

Principles of ethics

- Well-being

  They should not experience physical or emotional discomfort.

- Informed consent

  They should receive enough information about the experiment: what they will see, do, etc., but NOT the hypothesis!

  They typically sign a document documenting their consent

- Confidentiality

  Participation should be anonymous

  No personal data can be linked to participants' names: use ID numbers to uniquely identify participants

Evaluating an experiment

- You should always have a critical look towards your experiment and acknowledge its potential limitation
- Reliability
  - To what extent can the results be trusted?
  - An experiment is reliable if it produces consistent and replicable effects (test-retest & inter-observer reliability)
- Validity: do the results show what they are meant to?

  Construct, ecological, internal, external validity

Evaluating an experiment

- Construct validity
  - Concerns the operationalization of the object of analysis in the experiment
  - Does the task appropriately create the behavior under study?
  - Are the variables adequately operationalized/measured?
- Ecological validity
  - How is it representative of behavior outside the lab?
  - The task in the experiment should not be too artificial

Evaluating an experiment

- Internal validity
  - Were all explanatory variables accurately identified or controlled for?
  - Seriously consider and discuss alternative explanations
- External validity
  - How do the results generalize to the entire population?
  - Typical subject population = undergraduates, 18-24, educated, etc. → is it representative wrt. the object of study?