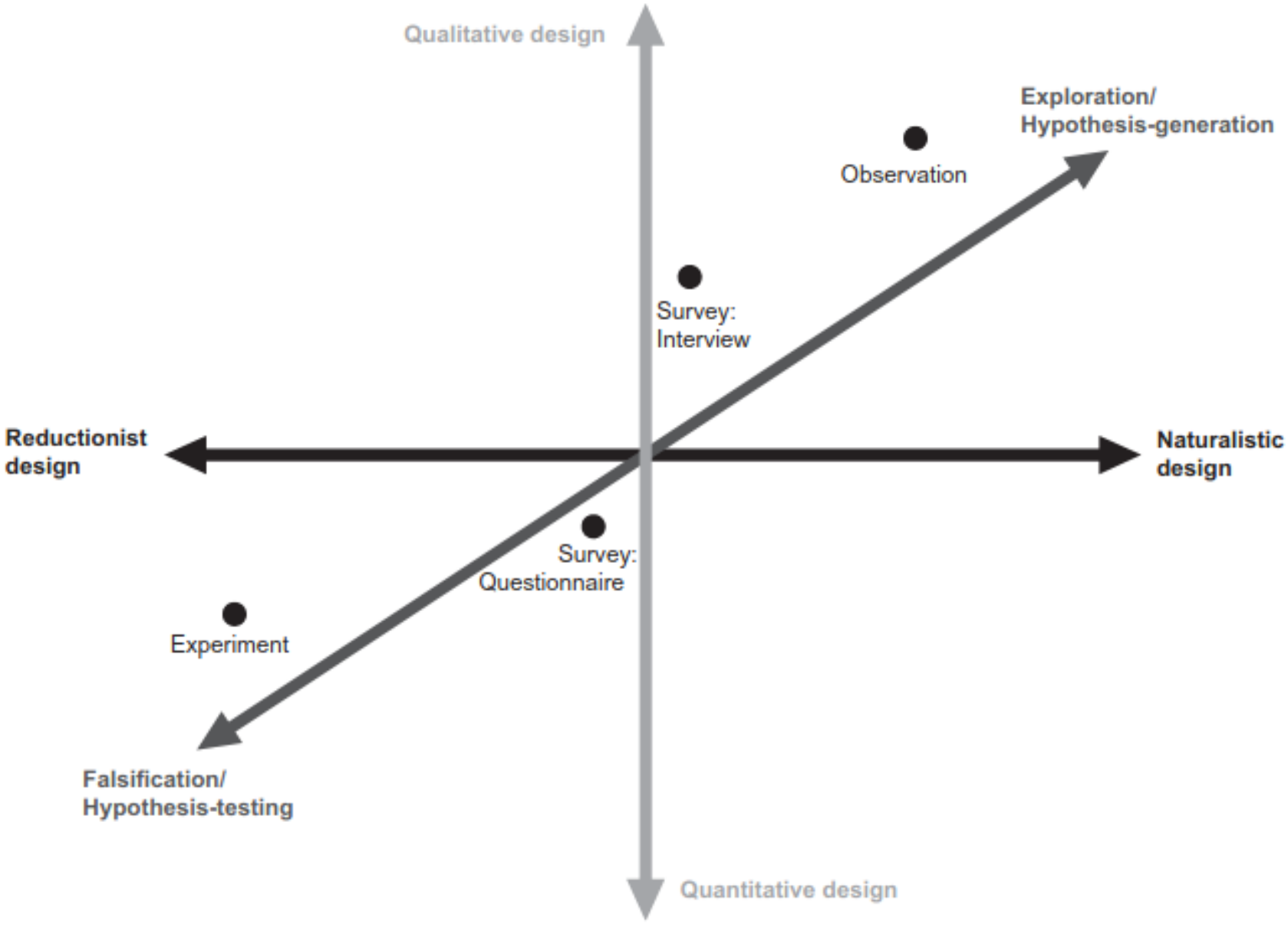
Empirical methods in linguistics [04]

Different types of data and operationalization, coding, analysis

Porf. Dr. Martin Schäfer
Department of English Studies, Universität Leipzig

Key concepts

- General/basic approaches
  - reductionist (controlled / “lab”) vs naturalistic (“uncontrolled”)
  - quantitative vs qualitative
  - exploratory vs hypothesis-testing
- Methods
  observation/survey/experiment or mixed methods
- Sampling/Acquisition
  How do I get my data? How much?
- Data analysis
  How will the data be analysed? How do I need to prepare them?



Prototypical usages of research methods according to three dimensions Völkel and Kretzschmar (2021, Figure 2.1, page 44)

Observation, Survey, and Experiment

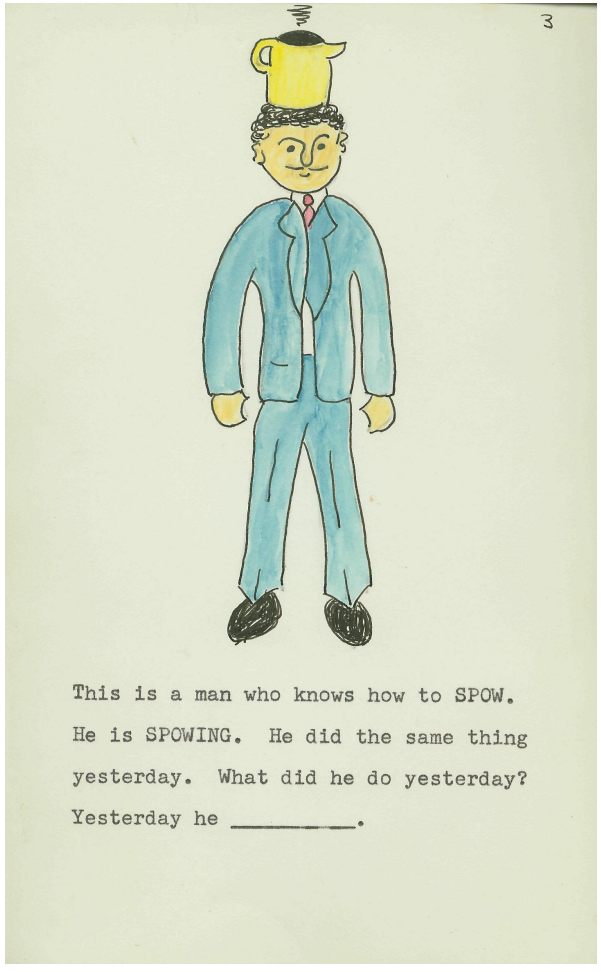
- Observation: “collection of spontaneous, perceivable data without systematic verbal interaction between researcher and participants.” Völkel and Kretzschmar (2021, p. 47)
- Survey: “research method utilising communication between researcher and participants with the goal of collecting verbal information from the participants on a specific research topic.” Völkel and Kretzschmar (ibid., p. 53)
- Experiment: “the researcher systematically varies or manipulates the levels of at least one independent variable in order to observe its influence on at least one dependent variable.” Völkel and Kretzschmar (ibid., p. 61)

Example 1: past tense *-ed*

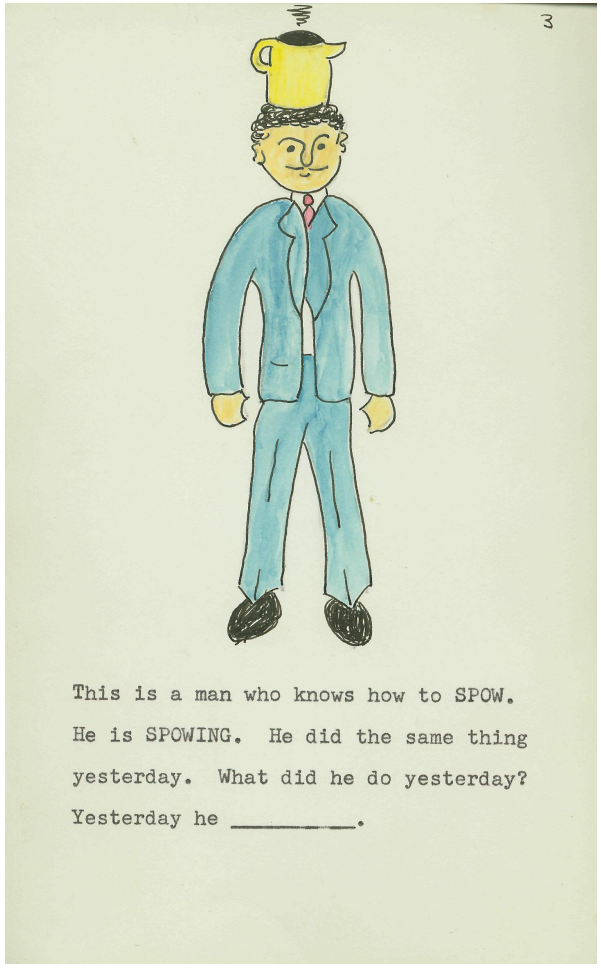
Suppose you want to know at what age English infants "know" the rule for regular past-tense formation in English (rather than just imitating *kissed*, *spilled* or *waited*).

- What kind of observation would you need to see in the children's spontaneous language use to be sure?
- Why are surveys (e.g. asking the parents) less helpful here?
- How could you answer your question based on a simple experiment?

The wug-test


3
This is a man who knows how to SPOW.
He is SPOWING. He did the same thing
yesterday. What did he do yesterday?
Yesterday he _____.

The wug-test



Berko (1958) (verbs PT: *bing*, *gling*, *rick*, *melt*, *spow* ...)

The colored picture was taken from the following website:

`https://allthingslinguistic.com/post/168977408695/sunday-morning-exercise-take-the-wug-test`

[copyright attributed there to *Jean Berko Gleason 2006*]

Example 2

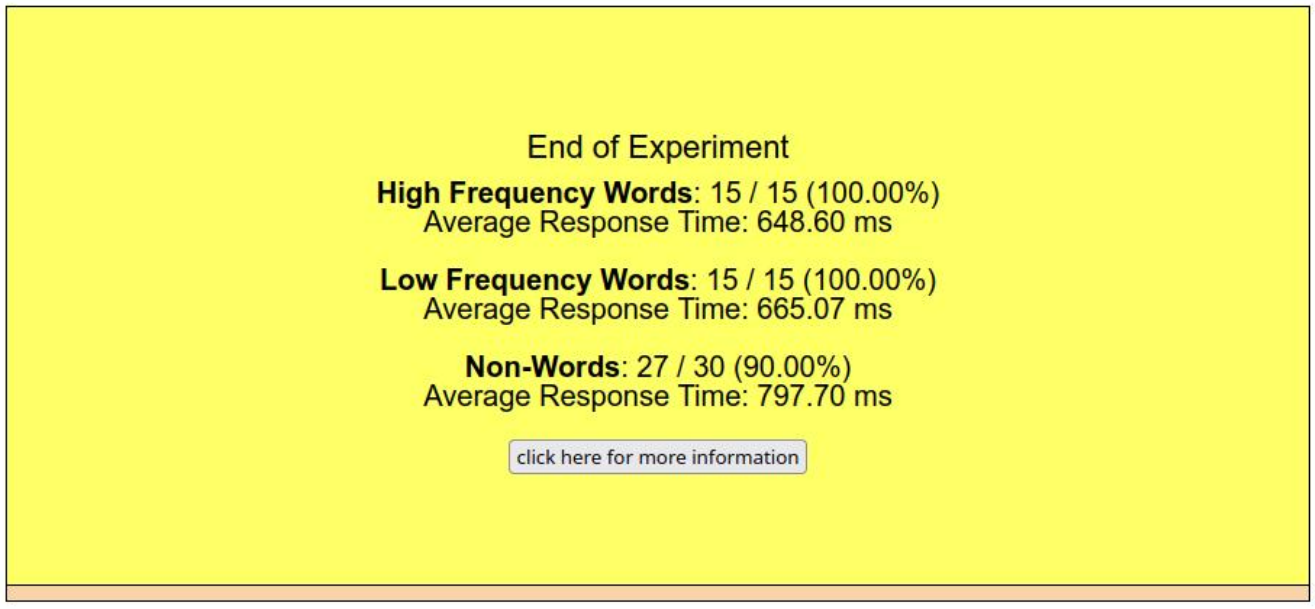
You would like to know how fast (and reliably) people recognise English words and what determines the speed for different words.

Let's do an experiment!

Please visit the following page:
`http://www.intro2psycholing.net/experiments/expt_index.php`

- Select **Visual word recognition** by clicking.
- On the next screen, select **Word frequency** by clicking
- **Carefully read the instructions :)**
- Do the experiment.

Results


End of Experiment
High Frequency Words: 15 / 15 (100.00%)
Average Response Time: 648.60 ms
Low Frequency Words: 15 / 15 (100.00%)
Average Response Time: 665.07 ms
Non-Words: 27 / 30 (90.00%)
Average Response Time: 797.70 ms
click here for more information

Analysing experimental data, step 1

- Open the RT dataset in your spreadsheet software
- Inspect the data (look at the documentation, look at the actual data, use the filter-function)
- Measurement scale?

Analysing experimental data, step 1

- Open the RT dataset in your spreadsheet software
- Inspect the data (look at the documentation, look at the actual data, use the filter-function)
- Measurement scale?
- categorical [different labels]
- ordinal scale [order along some continuum]
- interval scale [same difference between scale points]

Analysing experimental data, step 2: Hypotheses

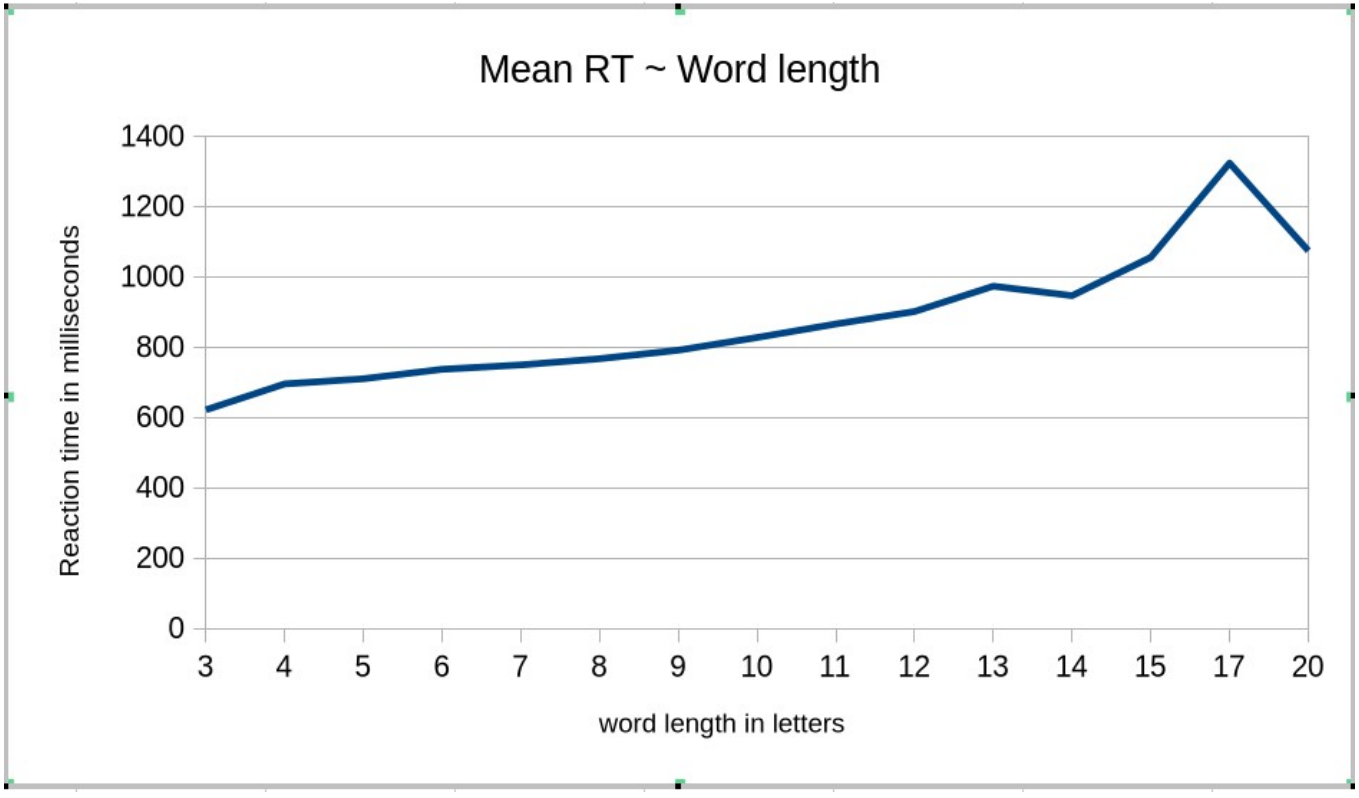
Formulating hypotheses (including IV/DV)

- Formulate one research hypothesis per independent variable. Decide for yourself where you suspect directionality —otherwise make it non-directional.

Step 3, analysis

- Let us look at the relationship between word length and reaction time
- Create a pivot table (insert pivot table, drag **Length** to **row fields**, **Mean_RT** to **data fields**, change type to **average** [doubleclick + select])
- Create a chart
  - preparatory step (building on what we learnt so far): copy data of interest from pivot table
  - then: insert chart), select line chart, etc.
  - use meaningful names for legend and axes

Step 3, analysis



Further steps

- What about frequency? Try to construct a pivot table and then a chart for the relationship between frequency and RT
- Use **FreqClass2**!

Next week

- See HA on Moodle! (Hahaha! A fine joke :))
- see next slide instead :)

HA: Getting access to corpus data

For your projects, and for following along in the next two classes, you need to acces large databases, so-called corpora, of English. Those revelant for our class are here:
`www.english-corpora.org`. For licensed access, you need to register as a student user and follow a few official steps to make this work. For the purposes of this class, unlicensed access is completely fine.

- Scenario 1: Moodle is running again by next Wednesday: Please find a step-by-step registration guide for the corpus website on Moodle.
- Scenario 2: no Moodle etc.: Go to www.english-corpora.org, select "my account" in the navigation menu at the top and the option "register". Use a private email address, i.e. not your usual university account. In "category", choose "Undergraduate student". You do not have to provide your university. A registration link will be sent to your email $\rightarrow$ complete the registration process.

References I

- Berko, Jean (1958). "The child's learning of English morphology". In: *Word* 14, pp. 150–177. URL: http://childes.psy.cmu.edu/topics/wugs/wugs.pdf.
- Völkel, Svenja and Franziska Kretzschmar (2021). *Introducing Linguistic Research*. Cambridge University Press. DOI: 10.1017/9781316884485.