

# Designing an empirical (quantitative) research project

## the bare essentials

### 1 General workflow

Planning and conducting an empirical research project requires you to complete the steps / answer the questions outlined below. I realise this can look like an intimidating task at first glance, so don't hesitate to come talk to me during [office hours](#) if you need help.

- I. Finding a research question and formulating (a) concrete hypothesis/es.
  1. Decide on a general topic or feature you want to focus on.
  2. Consult relevant publications to get information about the feature(s). Consider things such as history, evolution, theoretical approaches, use in particular varieties or social groups, and so on. (Not necessarily all of them.)
  3. Formulate a question that expresses what you are interested in.
  4. Formulate one (or several) hypothesis/es. Hypotheses are *concrete predictions* of what you expect to find in your data. See below for some more details.
  5. Show how your hypothesis connects to / follows from / fills a gap in existing research.
- II. Deciding on a methodology.
  1. What kind of resources / databases are available?
  2. Are these data suitable for answering your research question / testing your hypothesis?
  3. How can you analyse these data in a way that will either support or refute your hypothesis?
  4. How can you operationalise your variables? (How will you code your results?)
    - **What** exactly will you count, measure ...?
    - **How** exactly will you count, measure ...?
  5. What kind of tools will you need?
  6. Is the task manageable in the time frame available?

### III. Carrying out the research.

1. Collect the data.
2. Analyse and visualise the results.
3. Draw conclusions regarding your hypothesis/es.
4. Write everything up. (IMRaD format!)

## 2 Examples of research questions

- Is variety A changing with respect to feature X?
- Is feature X used more / differently in variety A than in variety B?
- What are the constraints governing the use of feature X?
- Do people talk differently depending on their addressee?
- What kind of social information do we ascribe to different forms in a language or different language varieties?
- Does the phonological context determine the realisation of a phoneme?
- Is the frequency of a word relevant for its realisation?
- Is there a pitch difference between stressed and unstressed syllables?

These are general examples. Work in a specific sub-discipline might require or exclude a specific kind of research question.

## 3 Characteristics of valid hypotheses

- General statement concerned with more than just a singular event.
- *If ..., then..., or the ..., the ...* structure (sometimes implicit).
  - **If part** of hypothesis contains the independent variable (often, but not necessarily the cause for the effects). This can be a linguistic or non-linguistic (e.g. social) feature.
  - **Then part** of hypothesis contains the dependent variable (and often its predicted values, distribution and so on). This will be a linguistic feature of some sort.
  - Example: *Men talk more than women.*
    - \* “If you compare men and women, then you will find that the former talk more.”
    - \* Independent variable (cause): gender.
    - \* Dependent variable: amount of speech (as measured in words or minutes).

- Falsifiability.
  - *Reducing the minimum age to obtain a driver's licence in a country may double the number of accidents.* → Not falsifiable, no valid hypothesis.
  - *Reducing the minimum age to obtain a driver's licence in a country will double the number of accidents.* → Falsifiable, valid hypothesis.

You can definitely refute / disprove a given hypothesis or corroborate it with supporting evidence. It is not possible, however, to confirm a scientific hypothesis beyond a doubt. This is because, at least in theory, there is always another explanation. It follows that **you can never prove your hypothesis**, so please don't tell me you did. Use *support*, *corroborate*, *provide evidence for* etc. rather than *prove* when you talk about your hypothesis.

## 4 Further advice

- *Your paper does not have to produce new results no one has ever heard of before.* Replicating existing research or providing further evidence for an existing claim is perfectly fine.
- I'd rather you perform a thorough analysis of one feature than a superficial investigation of several.
- You should, however, try to get as much out of the data you have as possible. Look at subgroups of speakers, consider different (linguistic) contexts, and so on.
- The easier it is for you to acquire a suitable dataset the more elaborate I will expect your analysis to be. For example, someone working with ready-made corpora can be expected to spend more time and effort on analysing their data than someone who needs to transcribe 5 hours of interviews first before they can actually do their phonological analysis.
- In most cases (exceptions apply), your analysis should be (moderately) quantitative, which means you will have to collect at least a certain amount of data. Anecdotal examples will not be enough. How much data you need depends on your topic and methodology. Do come and talk to me about this.

## 5 Resources

There are two ways of getting data for your research project:

1. Collecting your own data.
2. Using an existing database.

Data collection can happen in a number of ways, from handing out questionnaires to automatically storing tweets to conducting your own interviews. Talk to me about what is feasible in

a given context. Your data do not necessarily have to come from English if you do your own data collection, *though this depends on the corresponding seminar and your specific project*.

If you don't want to (or don't have the time to) collect enough data yourself, it is a perfectly valid option to use one of the many databases that are freely available over the internet. Some of them will contain audio files (with or without transcriptions), while others only contain transcribed texts. In many cases you will have to process the data in one way or another (transcribing, aligning, reformatting...) before you can actually start working with them. Here's a non-exhaustive list:

- [Millenium Memory Bank](#)
- [BBC voices](#)
- [The speech accent archive](#)
- [International Dialects of English Archive](#)
- [Talk of the toon](#)
- [SCOTS corpus](#)
- [American English Dialect Recordings](#)
- [English-corpora.org](#)
- [International Corpus of English](#)
- [Santa Barbara Corpus of Spoken American English](#)
- [Corpus of Regional African American Language](#)
- [Freiburg Corpus of English Dialects](#)
- [Vienna-Oxford International Corpus of English](#)
- [TalkBank](#)

## Further reading

Meyerhoff, Miriam, Erik Schleeff, and Laurel MacKenzie (2015). *Doing Sociolinguistics. A practical guide to data collection and analysis*. London: Routledge.

Wray, Alison and Aileen Bloomer (2006). *Projects in Linguistics. A Practical Guide to Researching Language*. London: Arnold.