

Planning and writing an empirical (quantitative) paper

the bare essentials

1 Designing a project

1.1 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 [subsection 1.3](#) 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, see also [subsection 2.2](#))

1.2 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.

1.3 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.

1.4 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.

1.5 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 harvesting texts from the internet 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 numerous existing databases, many (though not all!) of which 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](#)

You might find useful information on additional English-language corpora in the [Corpus Resource Database](#) hosted at the University of Helsinki. It is also possible to use the web itself as a corpus in a number of ways – be sure to comply with copyright and data privacy regulations, though.

2 Writing it all up

2.1 Formatting term papers (special rules apply for final theses)

2.1.1 Title page

- Top left corner:
 - “Ruhr-Universität Bochum”
 - “Englisches Seminar”

- Title of class
- Term
- Name of instructor
- Center: title of term paper
 - Horizontally centred
 - Larger font, bold printing optional
- Bottom right corner:
 - Name of student
 - Student ID number
 - E-mail address

2.1.2 Table of contents

- Opens on new page.
- Chapters, sections, subsections are numbered and given with the page they start on.
- Can be generated automatically.
 - Optional, but highly recommended.
 - Google something like “create a table of contents word” for detailed instructions.

2.1.3 Text body

- Opens on new page.
- Length: 3500–4000 words (excluding graphs and tables).
- Font: Times New Roman (size 12) or Arial (size 11).
- Line spacing: 1.5.
- Page margins:
 - 2 cm on the left (inner margin for double-sided printing).
 - 4 cm on the right (outer margin for double-sided printing).
 - 3 cm at the top and bottom.
- Page numbers: cover sheet is counted, but not visibly numbered; page numbering starts with “2” on the contents page.
- Example words appear in *italics*.
- Numbered headline and new paragraph for each (sub-)section.

- Indented paragraphs (about 1.25cm) within sections (but not after a headline, table, or graph).
- No space between paragraphs.
- Paragraphs fully justified (*Blocksatz*).
- Tables and graphs are marked as *table*, or *figure* and numbered continuously.
- Use footnotes parsimoniously and NEVER for citation!.

2.1.4 References

- Opens on new page.
- Contains all sources (books, articles, websites, ...) used in the writing of the term paper.
- Hanging indent (about 1.25cm).
- For detailed formatting instructions of the individual entries see the [Academic Style Sheet](#) on the website of the English seminar.
- Please note that you are required to apply the “Citation Conventions in Linguistics” as spelled out under section 9 of the “Academic Style Sheet”.
- Most important requirement: consistency!

2.2 Structure of main text

2.2.1 Introduction

- Theoretical background and motivation for research question / hypothesis.
- Research question or hypothesis explicitly spelled out.
- General structure of paper outlined.

2.2.2 Literature review

- Sometimes appended to introduction without its own heading.
- Summary of research relevant to the question or hypothesis under scrutiny.
- Discussion and critical evaluation of existing research, your own thoughts on the matter.
- The more your paper is based on data that you have collected yourself, the more basic your literature review may be. However, even a heavily empirical paper should be grounded in previous research and therefore also needs at least a few paragraphs of literature review.
- Everything should be relevant with respect to your research question / hypothesis!

2.2.3 Method

- Detailed description of all practical aspects of your study, such as:
 - Amount and source and kind (questionnaire, interview, electronic corpus...) of data.
 - Operationalisation ('technical definition') of relevant features and categories used in analysis.
 - If applicable: materials (experimental stimuli, questions, images...) used in data collection.
 - If applicable: technical equipment and software used in data collection and analysis.
- Should allow reader to replicate your study.

2.2.4 Results

- Systematic presentation of results, often relying on graphs and tables.
- Describe results in running text...
- ...but withhold interpretation as far as possible to let readers form their own ideas.

2.2.5 Discussion

- Spells out the conclusions drawn from your analysis, interprets the results.
- Re-assesses your research question / hypothesis.
- Addresses problems and limitations (methodology, representativeness...) of your paper.
- Presents motivated ideas on what one could do next.
- Ends with a summarising conclusion (either as separate section or simply as last paragraph of the discussion).

2.3 Citation

- Using someone else's thoughts and ideas without citing the source is plagiarism.
- A plagiarised paper is a fail.
- Direct (literal) and indirect (re-phrasings, summaries) quotations both need attribution.
- References appear in running text using the author-year format:
 - direct quotation: "This is literally what I said myself" (Smith 2015: 112).
 - indirect quotation: The author explained that this was essentially the claim that she had made herself (cf. Smith 2015: 112).
- Please sign the *Declaration of Authenticity* (last page of this document; taken from the [Academic Style Sheet](#)) and attach it to your paper.

2.4 DOs and DON'Ts

2.4.1 DOs

- Be precise and provide **RELEVANT** details (your reader should be able to replicate your results on the basis of your report).
→ Make sure everything in the write-up is relevant with respect to your hypothesis.
- Don't stop at the most basic comparison/analysis possible, but also have a look at sub-groups, individual test items etc. – especially, but not only, if your hypothesis is not confirmed.
- Use tables and graphs where appropriate.
- Consider basic linguistic (phoneme, allomorph ...) and statistical (p-value, t-test...) concepts as known – no need to explain them in your write-up.

2.4.2 DON'Ts

- Avoid prescriptive vocabulary.
- Don't "prove" your hypothesis because you can't.
- Don't write a chronological narrative but provide a systematic overview
→ **REPORT/SUMMARISE** retrospectively.
- Don't digress (too much) from the required structure/formatting. For example, problems that occurred – and how you solved them – go into the discussion section, not the method section.
- Avoid "personal" or emotional statements (like if the project was fun or not).
- Don't report only those data that confirm your hypothesis.
→ Don't try to hush things up; just state what the problem was (and how it could be avoided next time).
- Don't give long lists of t-tests etc., but use tables.
- Don't just let your graphs and tables speak for themselves – explain what you can see in your analysis/results section!
- Don't use [pie charts!](#)

2.5 Handing in your paper

You will have to hand in an electronic version of your paper. The following formal requirements apply:

- Word or pdf format; only ONE file, please!
- File name: LAST NAME_first name_semester_paper.doc/pdf
(example: SCHMIDT_Sarah_WS2017_paper.doc)

Please upload your file via Moodle – do not send it as an email attachment!

Further reading

Meyerhoff, Miriam, Erik Schleef, 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.

Declaration of Authenticity

I hereby declare that the work submitted is my own and that all passages and ideas that are not mine have been fully and properly acknowledged.

I am aware that I will fail the entire course should I include passages and ideas from other sources and present them as if they were my own.

date

signature