Teaching materials for introducing linguistic research

Franziska Kretzschmar & Svenja Völkel (2021)
Link: https://doi.org/10.14618/ids-pub-10454

Purpose of the slides and how to use them

- The teaching slides accompany the following textbook:

  Svenja Völkel & Franziska Kretzschmar (2021): *Introducing linguistic research*. Cambridge: Cambridge University Press

- The slides follow the structure of the book chapters and can be used for teaching in class.

- They include the basic information per chapter and exercises to work on in class or as homework. More detailed information, additional exercises, suggestions for research projects and recommendations for further reading can be found in the textbook.

- The slides are licensed under CC BY which means you are allowed to use the contents for your purposes ‘as long as you credit the original creation’ (https://creativecommons.org/licenses/?lang=en). The slides can be cited as follows:

  Franziska Kretzschmar & Svenja Völkel (2021): *Teaching materials for introducing linguistic research*. https://doi.org/10.14618/ids-pub-10454

Purpose of the slides and how to use them

- The structure of the textbook and the teaching slides resulted from a **teaching schedule** that we designed at the Johannes Gutenberg University of Mainz, Germany**.** Thus, our syllabus is based on the average duration of a German semester which typically comprises about 13 to 15 weeks of teaching with two sessions (90 minutes each) per week, plus lecture-free time of 8 to 12 weeks:

Chapter 1:	approx. 4 sessions or 6 hours
Chapter 2:	approx. 3 sessions or 4.5 hours (1 session/90 minutes per method)
Chapters 3-8:	approx. 9 sessions or 13.5 hours (1-2 sessions per chapter/subdiscipline)
Chapter 9:	approx. 1 session or 90 minutes

- In the final part of the course (i.e., the remaining sessions) and the beginning of the lecture-free period students should test their new theoretical and practical knowledge in their own small research project (about 4–6 weeks) and finally present it in a research-relevant way (e.g., poster presentation).
  Thus, the introductory class on empirical linguistics is intended to end with a project phase, including the following:

Joint project meetings:	approx. 3 sessions or 4.5 hours (ideally 1 session/90 minutes biweekly)
Autonomous work on projects:	remaining sessions/hours
Possible additional tutorials (not covered in the textbook or slides):	e.g., statistics (4 sessions or 6 hours)
Exam (project presentations):	1 session or 90 minutes

Purpose of the slides and how to use them

Symbols and cross-references

 see page 1	section in the textbook that should be consulted for further information
cf. Section 1.1	cross-reference to other sections of the teaching materials
	exercise
	advantages
	disadvantages

General outline

PART I – Basics

Empirical research 1

Basic methods for data collection 2

General outline

PART II – Subdisciplines

Language documentation and descriptive linguistics 3

Language typology 4

Corpus linguistics 5

Sociolinguistics and anthropological linguistics 6

Cognitive linguistics and psycholinguistics 7

Neurolinguistics 8

General outline

PART III – Conclusion

Insights from linguistic research 9

General outline

PART IV – Appendix

- A1 Ideas for empirical projects
- A2 Research diary
- A3 Poster presentation and publication
- A4 Feedback rules and exam

General outline

References

References cited on the slides — R1

Image credits — R2

Empirical research in linguistics

Chapter 1

Chapter 1: Outline

Research basics

Research, empiricism and theory, quality criteria and ethics

1.1

Research process

Components and important aspects of research

1.2

Why do we need research?

> Einstein once said: '... common sense is actually nothing more than a deposit of prejudice laid down in the mind prior to the age of eighteen.‘
> (Barnett 1948: 58)

What exactly is research?

Research:

- systematic search for new knowledge
- built upon existing knowledge
- examination on the basis of concrete experienced facts
- adequate methodological procedure; scientific standards

Basic requirements:

- curiosity and interest
- analytical understanding
- common sense and intuition

Research as an evolutionary process consisting of three components (Popper 1973)

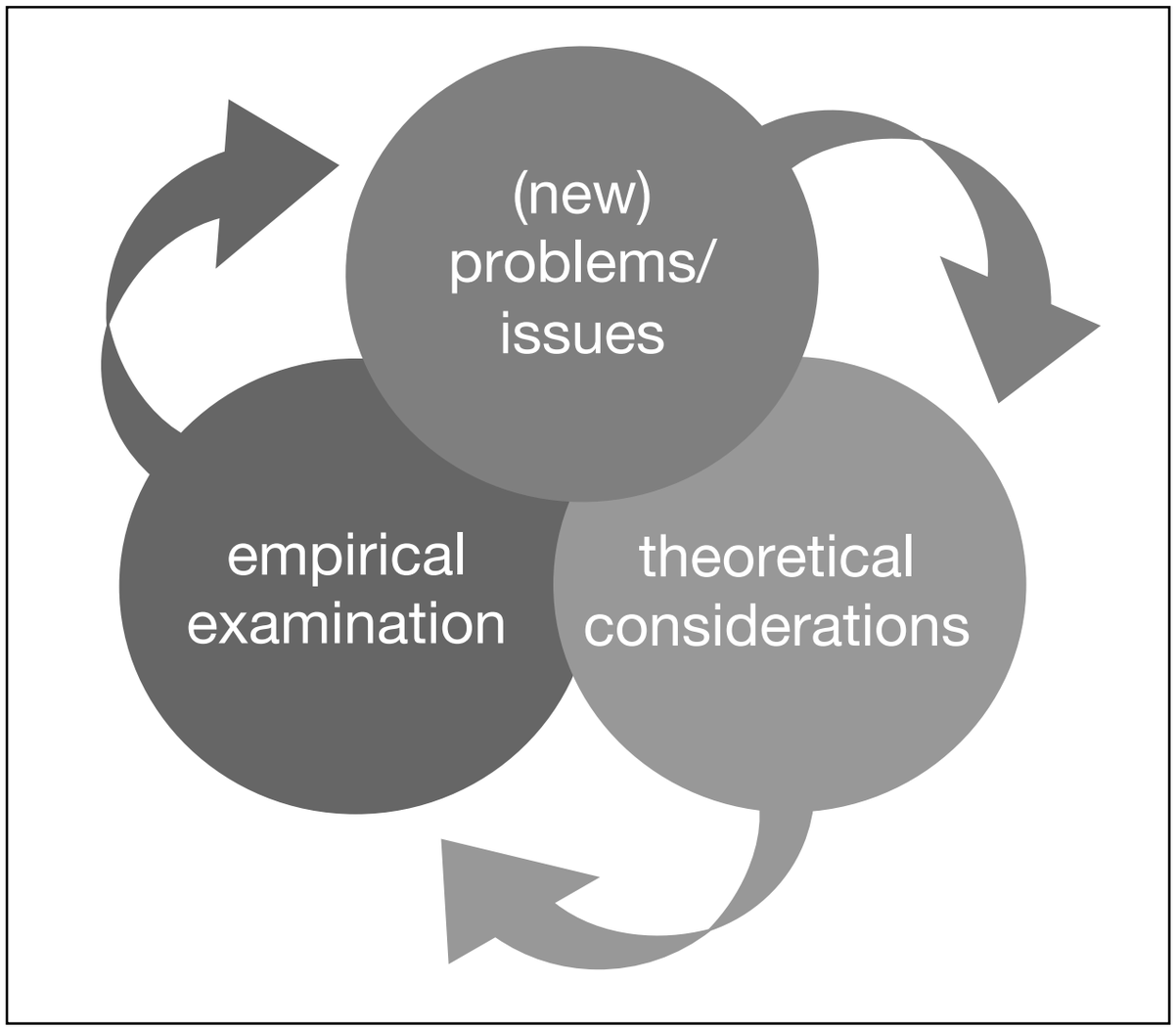


Figure 1

Theory ...	Empirical findings ...
describes and explains a phenomenon **based on existing knowledge**.	**confirm or disprove a theory** or necessitate modifications**.**
makes appropriate **predictions**.	are in any case an **improvement of knowledge** in a continuous evolutionary process**.**
needs to be **verified** by systematic **data-driven research**.	lead to **new issues**.

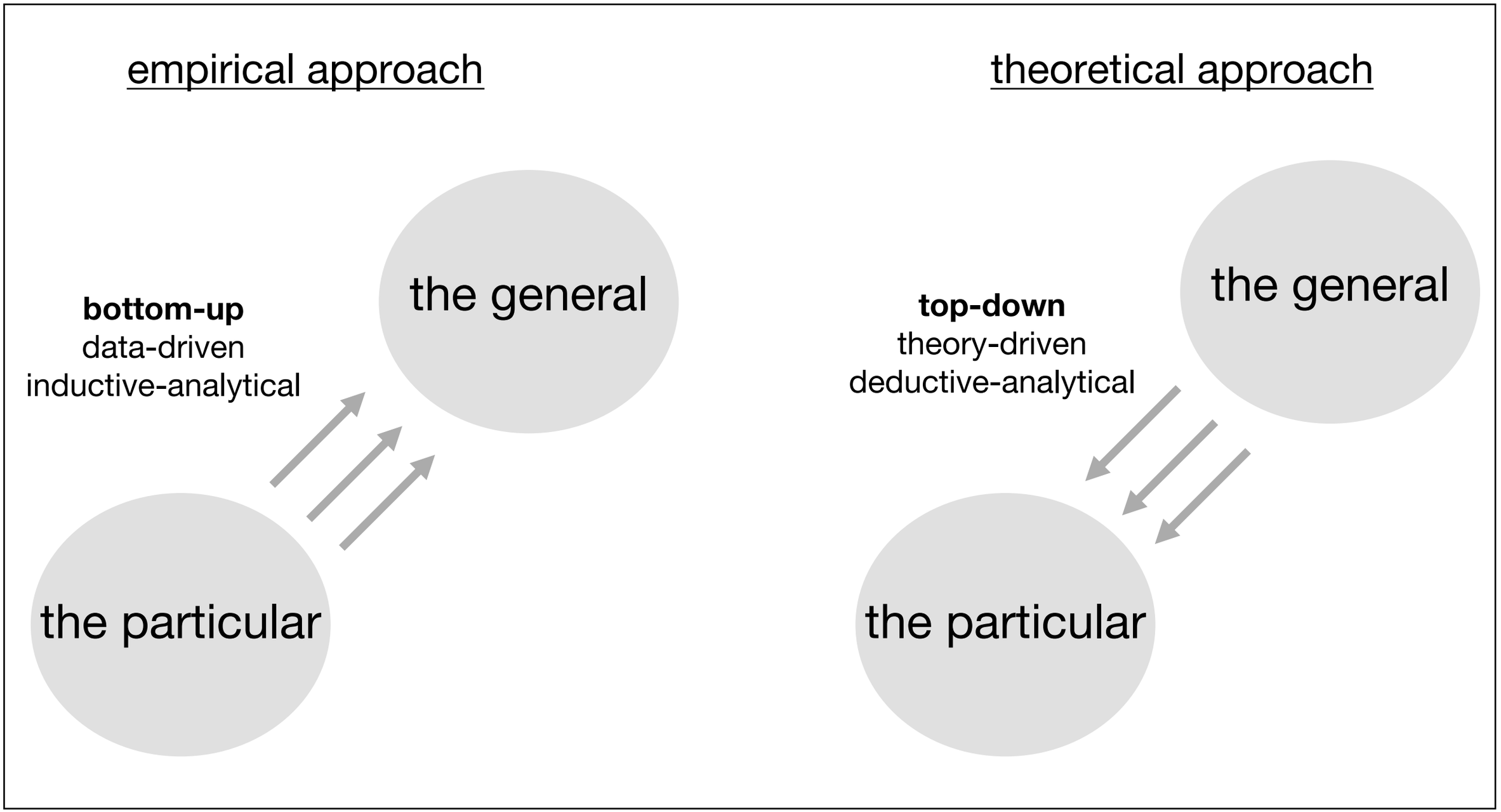

empirical approach
bottom-up
data-driven
inductive-analytical
the general
the particular
theoretical approach
top-down
theory-driven
deductive-analytical
the general
the particular

Figure 2

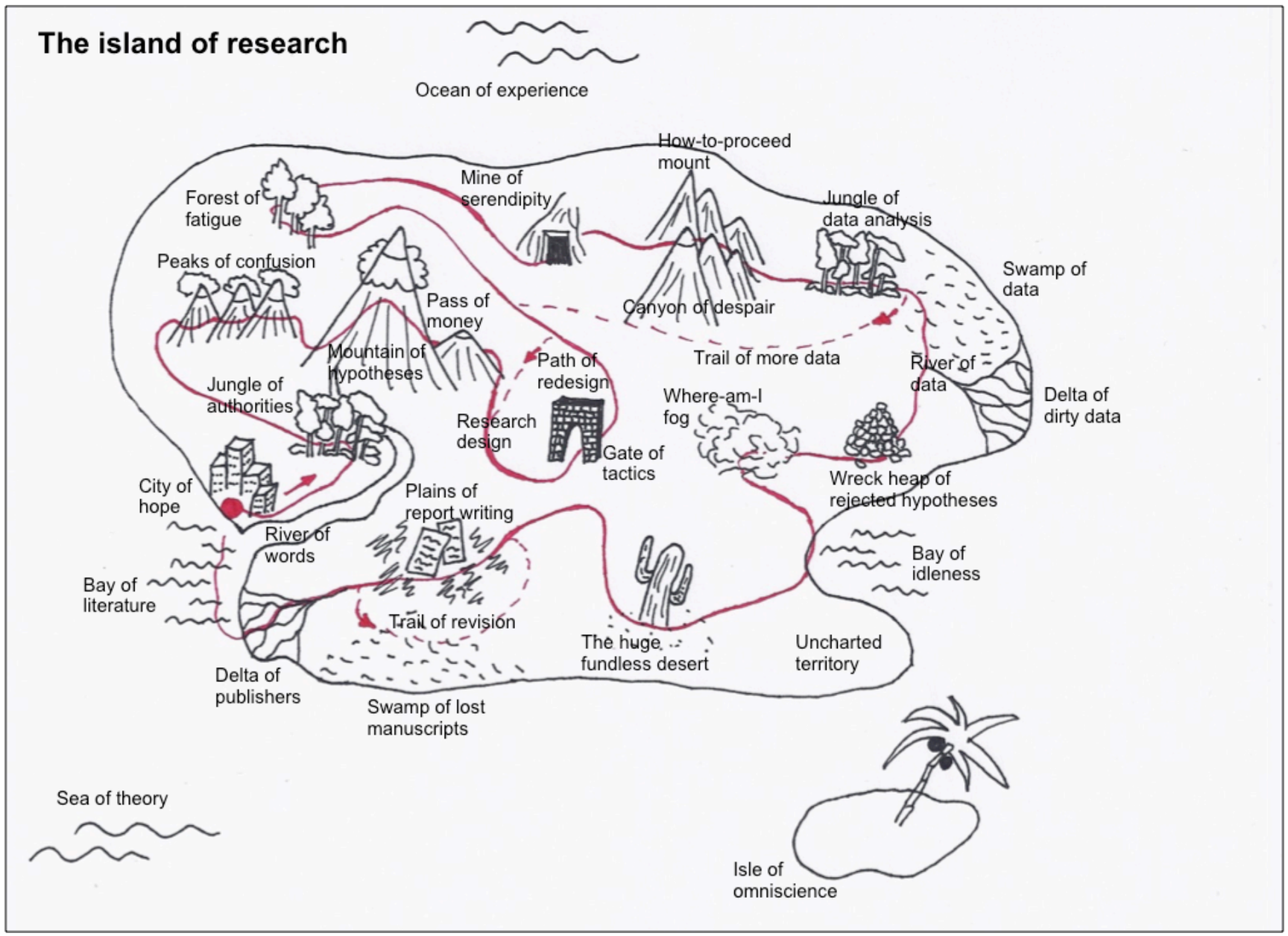


Figure 3
Alemann (1984[2]: 152f, redrawn by Angelika Morgh, with modifications and translations by the authors)

Main stages of the research process:

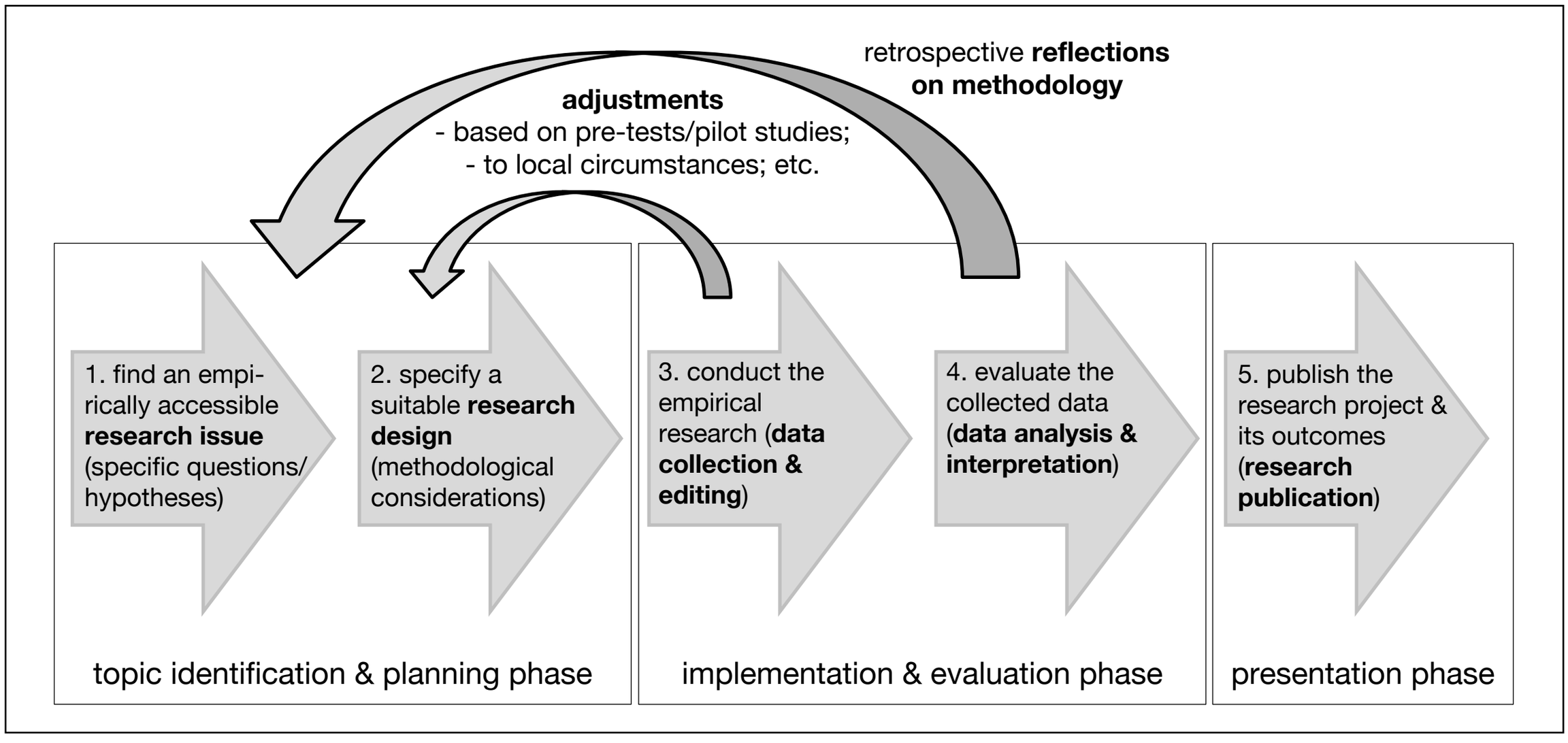


Figure 4

Research components:

Researcher (who)

- single person or group of multiple researchers
- observers, interviewers, experimenters, analysts
- linguistics: linguists

Introspection

researcher = participant

Participants (whom/who)

- one person (case study) up to large numbers
- speakers/signers, writers, hearers & readers
- informants, interlocutors or respondents
- native speakers, particular subgroups or learner groups

Research aim (why)

- generally: contribution to knowledge building
- linguistics: understanding of language in all aspects

Research issue (what)

- research topic
  → research question/hypothesis that is to be answered
- object of research: language (past and present languages and their varieties; single ones or multiple)

Research environment (where)

- field vs. laboratory vs. office vs. combined

Research data

- primary vs. secondary vs. tertiary data
- language data (natural vs. research-generated) vs. language-related data
- confidential vs. anonymous data

Research design (how)

- empirical approach and methodological procedure
- linguistics: fundamental differences in distinct linguistic subdisiplines

Temporal framework (when)

- time in which the research is carried out (from planning to publication)
- data collection: at a certain point in time vs. repeatedly over a long period

Figure 5

Basic kinds of research:

- quantitative vs. qualitative research
- diachronic vs. synchronic research
- cross-sectional vs. longitudinal research
- field vs. laboratory research
- monolingual vs. comparative vs. multilingual research

Basic kinds of research per linguistic subdiscipline

Language documentation & descriptive linguistics	Field research, collection and analysis of primary language data
Language typology	Comparative/cross-linguistic research, usually by use of secondary/ tertiary data (grammars)
Corpus linguistics	Analysis of natural language, generally by use of secondary data (ready-made corpora)
Sociolinguistics & anthropological linguistics	Field research, analysis of language data in the context of socio-cultural data of the speakers
Cognitive linguistics & psycholinguistics	Laboratory (or field) research and/or analysis of language data or typological findings
Neurolinguistics	Laboratory research

Quality criteria to meet the required scientific standard:

- **objectivity:**

  The same research conducted by other researchers should deliver the same results (independence from the researcher).

  → more realistic aim: **intersubjectivity** and **transparency**

- **reliability:**

  The research method produces consistent results that are reproducible with the same methods under the same or comparable conditions (**replicability**).

- **validity**

  → internal validity: The research procedure actually needs to measure what is intended and the research successfully answers the research question/ hypothesis.

  → external validity: generalisability of research outcome; **applicability** to natural settings (ecological validity), to other people (population validity), or other times (historical validity)

Researcher bias & reciprocal effects:

- **researcher bias:**

  bias resulting from the researchers' subjective viewpoint (selective perception) – their personal beliefs, thoughts, expectations, feelings, or attitudes

- **reciprocal effects:**

  (unwanted) research effects that result purely from the researcher's presence; these may be intensified by the use of technical equipment.

  → observer's paradox (response effects of the observer's presence; see Section 2.2)

Read Miner (1956) and discuss the observational description of 'Nacirema'.

Who is the object of observation and what does the author want to illustrate with his description?

Research ethics and rules of conduct:

- respectful and responsible behaviour
- protection of personal rights and well-being:
  no harm, no compulsion, respecting their will
- sensitivity and empathy
- expression of gratitude for their cooperation, offer of compensation and access to research data

  → official statutes:

  international rights & local laws/informal rules, ethical policies of research institutions & funding agencies

  → research approval

- sharing scientific knowledge
- respect of intellectual property: no plagiarism
- presentation of clean data and transparent analyses

generally

vis-à-vis the research participants

vis-à-vis the scientific community

Basic kinds of research & research questions:

- **descriptive studies**: description of a phenomenon → *'What is the case?'*
- **explanatory studies**: search for an explanation for a phenomenon → *'Why is s.th. the case?'*

- **explorative studies**: first overview of a relatively unstudied topic → general and open research questions
- **hypothesis-testing studies**: specific research focus, built upon previous knowledge → *'Is s.th. the case given specific parameters?'*

- **qualitative studies**: textual descriptions of a research issue → *'How is s.th.?'*
- **quantitative studies**: numeric presentations of a research issue → *'How frequent is or to what extent occurs s.th.?'*

Hypotheses: empirically verifiable/falsifiable assumptions or predictions

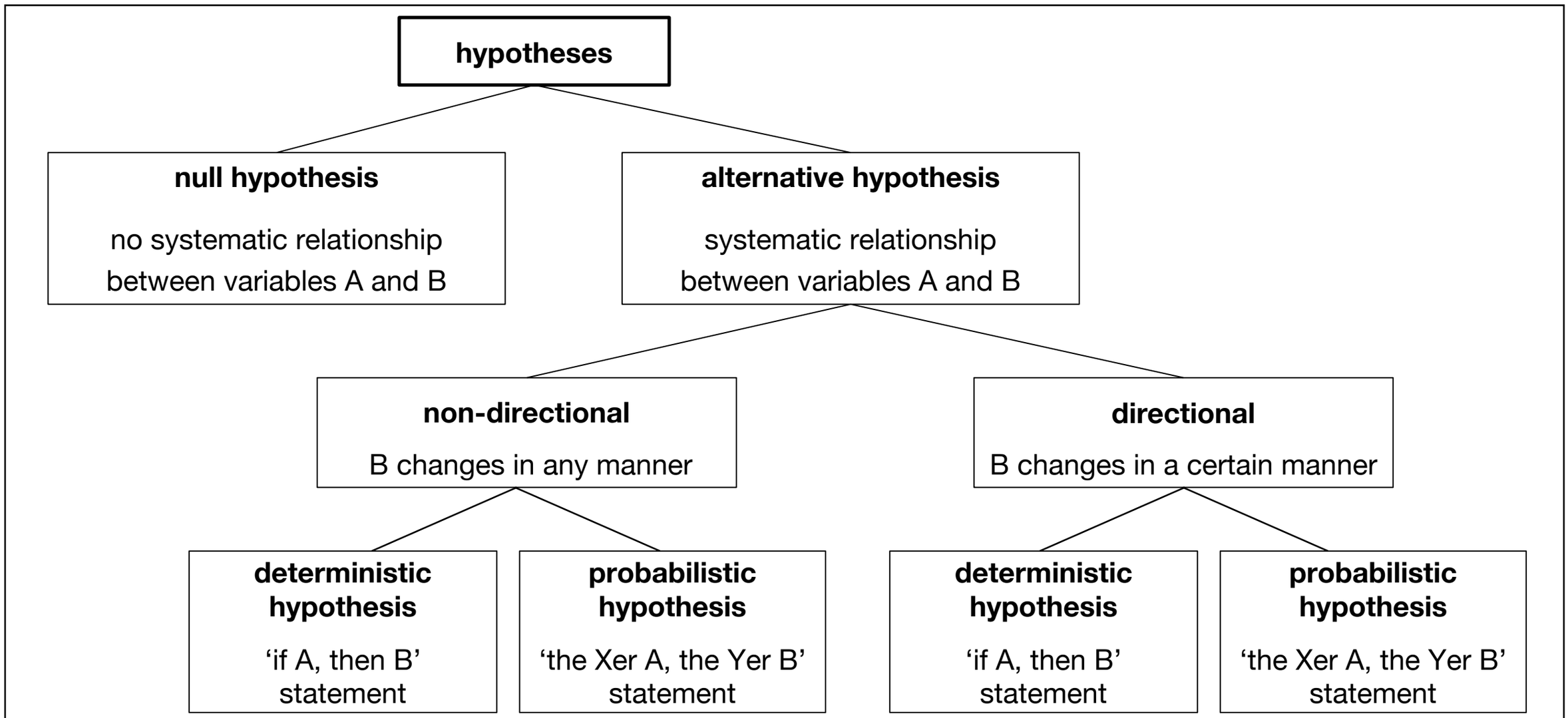


Figure 6

Formulate different kinds of hypotheses with A (sex) and B (amount of used interjections).

Basic research questions per linguistic subdiscipline

Language documentation & descriptive linguistics	– How do native speakers linguistically behave in various naturally occurring contexts? – What are the underlying structural patterns of an unstudied or less-studied language?
Language typology	– What are the common linguistic features and what are the differences between the languages of the world and how are the different features distributed?
Corpus linguistics	– What are quantitative or qualitative patterns of language use in natural situations?
Sociolinguistics & anthropological linguistics	– How is language used by different groups of speakers or in different social/cultural contexts? – In which way do linguistic forms and practices reflect culture-specific/social meaning?
Cognitive linguistics & psycholinguistics	– Which cognitive conceptualisations are reflected in language, and which mental processes are active (and when) during language production/comprehension/acquisition? – Is there an interrelation between language and thought?
Neurolinguistics	– When and where is language processed in the brain? – What is the genetic basis of human language?

overall aim of linguistic research: **find answers to open issues/questions**

What could be researched in linguistics?

Find interesting, relevant and empirically investigatable research questions or hypotheses that are feasible within the project framework.

Advices:

- You can start your research diary (see Appendix A2) with this loose collection of research issues. It can serve as an initial pool of ideas for developing your own research project.
- Write down a list of general topics, and narrow them down to concreate and clear-cut research questions, and/or specific hypotheses.
- Consult linguistic resources on the chosen topic to ground the question/hypothesis in the current state of research.
- Instead of developing your own research project from scratch, it might be easier to replicate other studies with marginal changes regarding a single parameter.

Making a research question/hypothesis empirically measurable:

Hypothesis:
Man rather pronounce the suffix ‘-ing’ as in ‘going’ [-in], while women rather pronounce it [-iŋ]. (cf. Labov 1972)

- identification of **variables** (basic parameters or research units as relevant for the study)
  - A: independent variable
  - B: dependent variable

variable A: ???;
variable B: ???

- definition of their **values** or defined levels, which need to

be *disjoint*

and *exhaustive*

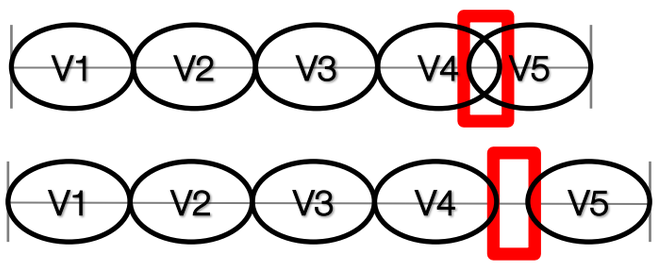


Figure 7

values of A: ???;
values of B: ???

- selection of **research objects** carrying these values

A: ???
B: ???

Making a research question/hypothesis empirically measurable:

Hypothesis:
Man rather pronounce the suffix ‘-ing’ as in ‘going’ [-in], while women rather pronounce it [-iŋ]. (cf. Labov 1972)

- identification of **variables** (basic parameters or research units as relevant for the study)
  - A: independent variable
  - B: dependent variable

variable A: sex;
variable B: pronunciation of ‚-ing’

- definition of their **values** or defined levels, which need to

be *disjoint*

and *exhaustive*

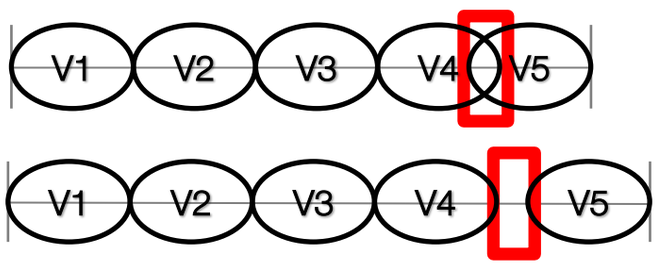


Figure 7

values of A: male, female;
values of B: [-in], [-in]

- selection of **research objects** carrying these values

A: humans
B: human speech

Kinds of variables:

- **independent variables** (A): assumed to have an impact on B
- **dependent variables** (B): assumed to change under the impact of A
- **intervening variables**: other parameters with an impact on B, apart from A

- **categorical variables**: with a finite number of values without a logical order/sequence – dichotomous (2 values, e.g., sex) vs. polytomous (more than 2 values, e.g., gender: masc., fem., neu.)
- **discrete variables**: numeric variables with discrete non-decomposable values (e.g., number of siblings: 2 or 3 siblings)
- **continuous variables**: numeric variables with continual values (e.g., age: 2.5 years)

- **manifest variables**: pertain to observable properties
- **latent variables**: pertain to properties that are not directly observable

Operationalise the continuous variable ‘age’ in terms of possible values.

Discuss different options and consider for which kind of linguistic research question/hypothesis they are suitable.

page 20-21

Levels of measurement:

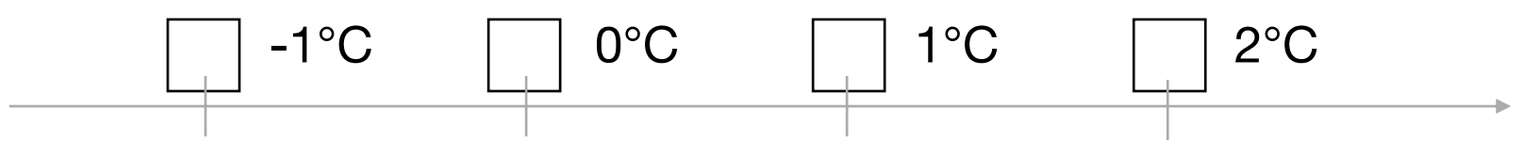
- nominal scales: alternative values without a ranking of the differences

☐ male
☐ female

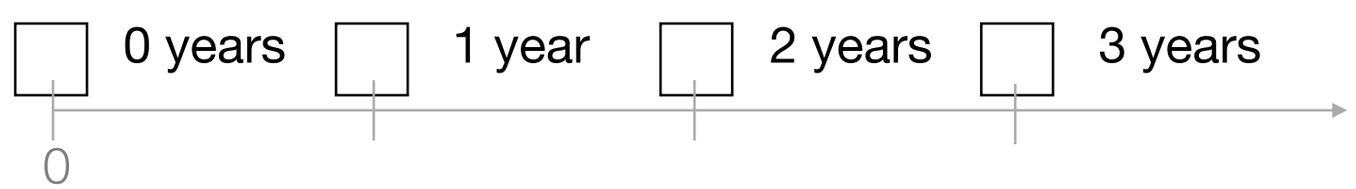
- ordinal scales: hierarchically ordererd alternative values, but without metric intervals

- metric scales: alternative values represent metric intervals, i.e. intervals of the same size

  - interval scales: without zero point



  - ration scales: starting from zero



→ important for the application of statistical measures (see below: statistics)

Data as needed for studying the specific research question/hypothesis

- primary vs. secondary vs. tertiary data
- language data (natural vs. research-generated) vs. language-related data

Sampling: work with a representative unbiased selection of data (e.g., languages, research participants or language items)

- *sample size*: one instance (a selection but not really a sample) up to a huge number of instances
- *selection procedure*:
  - **random sampling**: selection solely on a random basis
  - **quota/systematic/purposive sampling**: selection guided by relevant research criteria

→ interaction of sample size & number of considered criteria (or subgroups) in quota sampling

→ problematic aspect: unequal access to elements of the basic set/population (→ weighting)

Basic methods of data collection

- **observation**: collection of sensually perceivable (particularly visible) information

  → e.g., observation & documentation of natural language data

- **survey**: collection of information via verbal interaction (questions and answers)

  - **interview** (... in an oral way)
  - **questionnaire** (… in a written way)

  → e.g., *elicitation* in the sense of obtaining conveniently generated language data by systematic survey

- **experiment**: collection of systematic data under controlled conditions (generally in the laboratory)

  → e.g., experiments with stimuli and tasks investigating language skills or attitudes

→ more detailed information will follow in Section 2

Basic techniques of data documentation

- written notes: protocols or detailed reports, handwritten or in electronic version

  → during data collection and/or in the immediate aftermath

- ???
- ???

Discuss advantages.

- ???
- ???

Discuss disadvantages.

- recordings: audiotapes, video recordings, photographs, recordings of special devices for laboratory experiments, etc.

- ???
- ???

Discuss advantages.

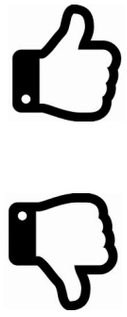
- ???
- ???

Discuss disadvantages.

Basic techniques of data documentation

- written notes: protocols or detailed reports, handwritten or in electronic version

  → during data collection and/or in the immediate aftermath



- possibility to capture background information and the overall setting
- usability in situations in which electronic recordings are impossible or inappropriate

- insufficient for capturing large volumes of data in a short period of time
- no access to data that were not recognised the first time

- recordings: audiotapes, video recordings, photographs, recordings of special devices for laboratory experiments, etc.

- possibility to capture a large amount of information in a short period of time
- repeated access to the data for a more detailed perception

- Devices create a non-natural atmosphere and enhance the observer's paradox.
- only a selected focus can be recorded (limited by positioning & perspective of devices)

General rules for data editing

be consistent, follow general conventions, chose a variant that fits the particular study

Basic steps of data editing

- **transcription**: systematic transfer of oral data into a written form (without adjustments/corrections)
  - *phonetic/phonemic transcription*: generally IPA-based
  - *orthographic/graphemic transcription*: based on a writing system
- **transliteration**: transfer of a written text into another writing system
- **translation** of transcripts/written data into a widely known language
- **annotation**: addition of linguistic information
  - *Part-of-Speech (POS) tagging*: categorisation of linguistic units based on grammatical information (see Section 5)
  - *interlinear/morpheme-by-morpheme glossing*: based on the Leipzig glossing rules or GRAID (see Section 3)

Search in linguistic studies for examples of each of these editing processes.

Basic research procedure per linguistic subdiscipline

Language documentation & descriptive linguistics	– recordings of natural language data; transcription, annotation and translation of these recordings – elicitation and/or corpus analysis of recorded, transcribed and annotated texts
Language typology	– cross-linguistic analysis of data from language descriptions or surveys (mainly questionnaires)
Corpus linguistics	– analysis of natural language data
Sociolinguistics & anthropological linguistics	– participant observation and surveys (mainly interviews) and/or analysis of linguistic data in relation to socio-cultural data
Cognitive linguistics & psycholinguistics	– laboratory or field experiments (cognitive tasks) and/or language analysis with regard to underlying concepts and processes
Neurolinguistics	– laboratory experiments; quantitative analysis of neurobiological basis of language processing

Data analysis: systematic search for specific patterns

Units of analysis: entities about which you want to make a statement as determined by the research question/hypothesis; they may be but are not necessarily the same as the units of measurement

Some analytical methods in linguistics:

- **distribution analysis**
- **feature analysis**
- **component analysis**
- **conversation analysis**
- **discourse analysis**
- **content analysis**
- **network analysis**
- **corpus analysis**
- **contrastive analysis**
- **error analysis**
- **comparative analysis**
- **...**

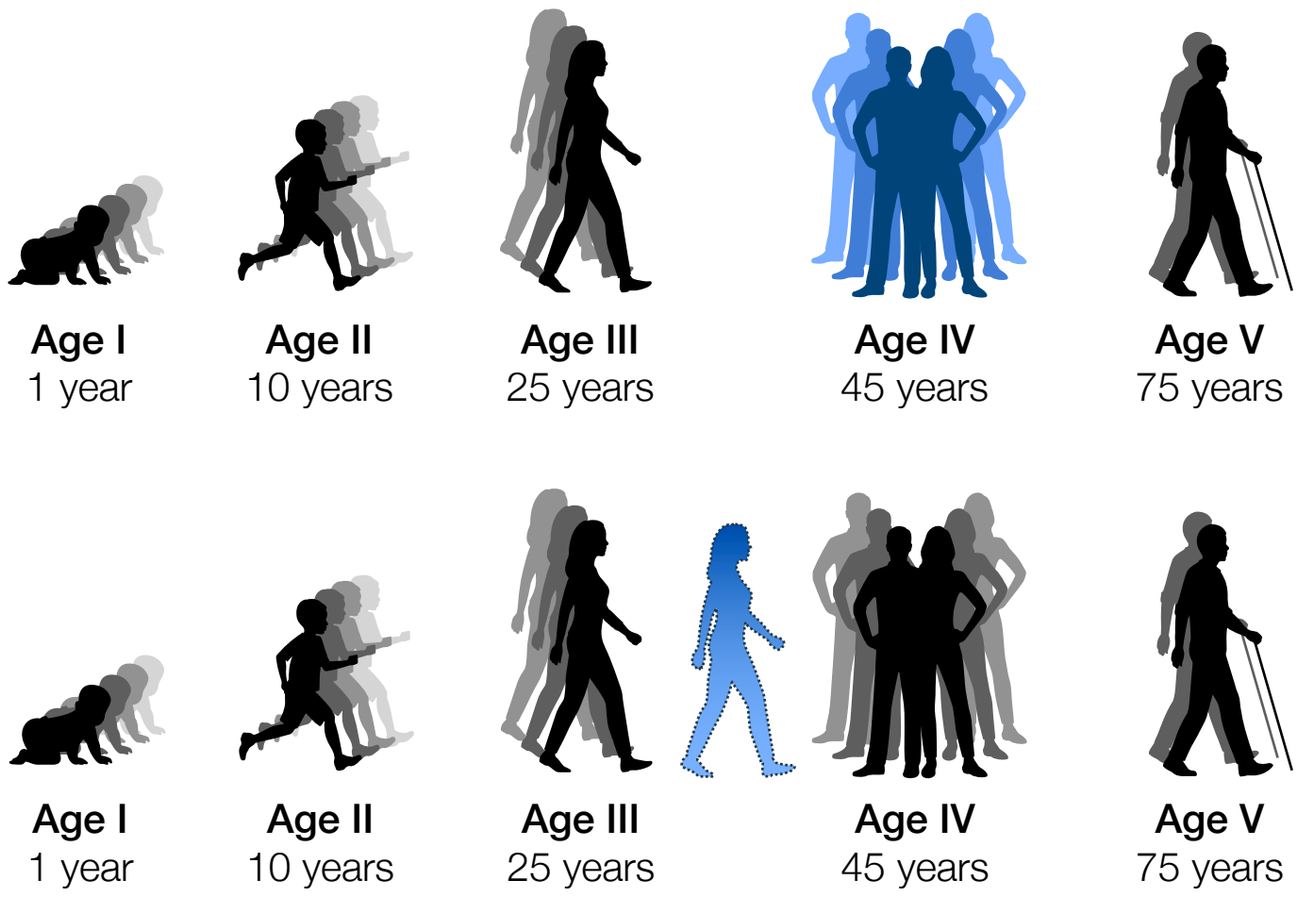
Statistics

- **descriptive statistics**: describe systematic patters in any data set

  - frequencies: absolute frequency (e.g., 10 participants are female), relative frequency (e.g., 1/4 or 25% of the participants are female) } i.e., 10 of 40 participants
  - measures of central tendency: mode, arithmetic mean, median
  - measures of dispersion: range (between the two most extreme data points), variance, standard deviation, standard error

- **inferential statistics**: calculate the probability of patterns; draw inferences about the representativeness and generalisability of the findings

  - significance level: e.g., p-value < .05 in most disciplines
  - parametric tests (normal distribution of the data points is required): e.g., t-test, Anova
  - non-parametric tests: e.g., chi-squared test, Wilcoxon rank sum test

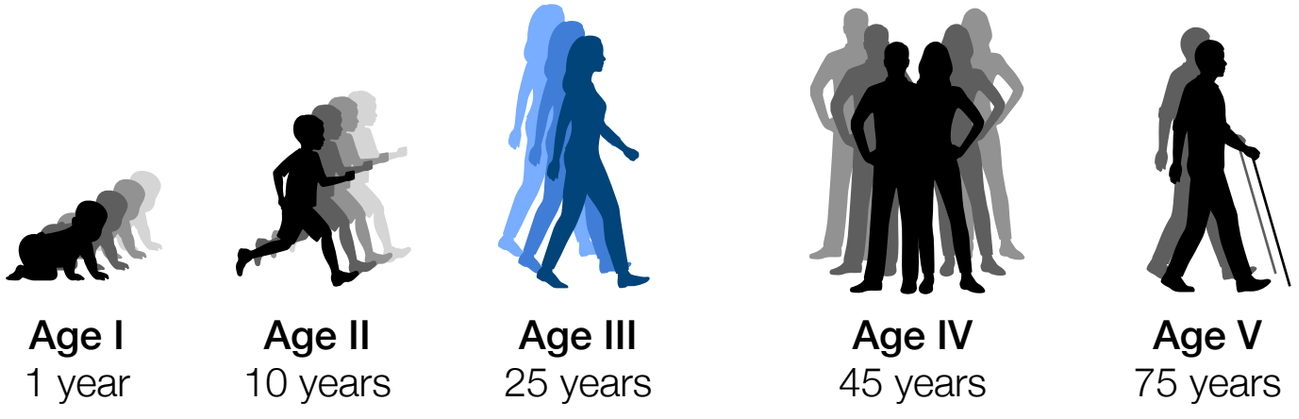


Mode – the value shared by most data points

There are 6 people in group IV. This is a higher number than in the other groups with a count of 2-4. The mode is 45 years.

Mean – the sum of all data points divided by the total number of data points

539 years (4x1 + 4x10 + 3x25 + 6x45 + 2x75) are divided by 19 people.
The mean is 28.37 years.



Median – the value that divides all data points into two even halves

For people in group III there are exactly two age groups younger or older than them. Equivalently, there are exactly 8 people younger or older than the people in group III. The median is 25 years.

Figure 8

Statistical measures depending on the level of measurement (Meindl 2011:98)

	Nominal scale	Ordinal scale	Interval scale	Ratio scale
Interpretation	qualitative differences (e.g., male vs. female)	order relation/ranking (e.g., some, many, all)	length of the interval between values (e.g., -1°C, 0°C, 1°C)	ratio of values (e.g., 0 years, 1, year, 2 years, 3 years)
Mathematical operations	NONE	NONE (e.g., no sums as you cannot add up 'some' and 'many')	sums & differences (e.g., no products as 20°C is not twice as warm as 10°C, then what about -10°C)	sums, differences, products & ratios
Statistical operations	frequencies & mode	frequencies, mode, median, quartiles	frequencies, mode, median, quartiles, arithmetic mean, range, standard deviation	frequencies, mode, median, quartiles, arithmetic mean, range, standard deviation
Visualisation tecjniques	bar/pie charts	bar/pie charts, boxplots	bar/pie charts, boxplots & scatter diagrams	bar/pie charts, boxplots & scatter diagrams

Outcome of data analysis: information that the data has revealed

Retrospective reflections

- Are the data and the research outcome plausible?
- Do the results answer the research question?
- Have all kinds of bias been ruled out?
- Does the empirical study meet scientific standards (i.e., quality criteria and research ethics)?

Interpretation of analysis results: information which the data itself does not show but which can be deduced or concluded from it

Basic research outcomes per linguistic subdiscipline

Language documentation & descriptive linguistics	– edited text corpora, (reference) grammars, and/or dictionaries
Language typology	– universals (cross-linguistic commonalities) & rara (cross-linguistic particularities), typologies (cross-linguistic variation), and distributional maps
Corpus linguistics	– patterns of language use, generally language-internal variation
Sociolinguistics & anthropological linguistics	– relationships between cross-linguistic or language-internal variation and social parameters of the speakers or the cultural context
Cognitive linguistics & psycholinguistics	– mental conceptualisations as encoded in language, patterns of language behaviour (e.g., speed and accuracy), and relationships between language and thought
Neurolinguistics	– patterns of brain activation in time and space

Publication of academic research - a basic requirement, leading to feedback

Formats of publication:

- conference and workshop presentations (**posters** and **talks**)
- publications (**articles** in scientific journals or edited volumes and **monographs**)

Basic elements of presentations & publications:

- title
- author & affiliation
- table of contents
- glossary
- abstract
- introduction
- main part: methods, results, etc.
- conclusion
- references
- appendix

→ for more detailed information on posters see Appendix A3

Basic research methods for data collection

Chapter 2

Chapter 2: Outline

Research design
Basic considerations on data collection

2.1

Observation
Definition, types and specifics of conducting an observation

2.2

Survey
Definition, types and specifics of conducting a survey

2.3

Experiment
Definition, types and specifics of conducting an experiment

2.4

Mixed-methods
Definition and types

2.5

The research design is the starting point of an empirical project and when it comes to selecting the method(s) for data collection.

Basic approach

- Quantitative, qualitative, or both? Explorative or hypothesis-testing? Research question, hypothesis, and variables?

Sampling

- Cross-sectional or longitudinal? Sampling procedure and sample size?

Method to collect (primary) data

- Observation, survey, experiment, or mixed-methods?

Data analysis

- Which analysis approach is appropriate to analyse the data?

Data set: #01
Participant ID: MA01 (JE02)
Date, time, location: 01 Sep 21
Researcher ID: J. Doe
Session: 01

Section 1: ...
Section 2: additional informant JE02 joined the conversation
Section 3: MA01 got tired
Section 4: not recorded

Additional remarks: recording device failed unnoticed after section 3
→ check criteria for data exclusion

Figure 9

A **protocol** supplements data collection and includes (at least) the following information:

- unique identifier (ID) for data file (e.g., participant or data set ID)
- date, time, location of data collection
- ID of researcher/person collecting the data
- specifics of the recording session (e.g., sequence of recordings and pauses, number or behaviour of informants/participants)
- additional, noteworthy remarks, such as:
  - unexpected interruptions (situational, technical, etc.)
  - participant feedback (e.g., on experimental task)
  - additional informants joining interview or observation
  - personal reflections

Primary data collection methods are prototypically used along three dimensions:

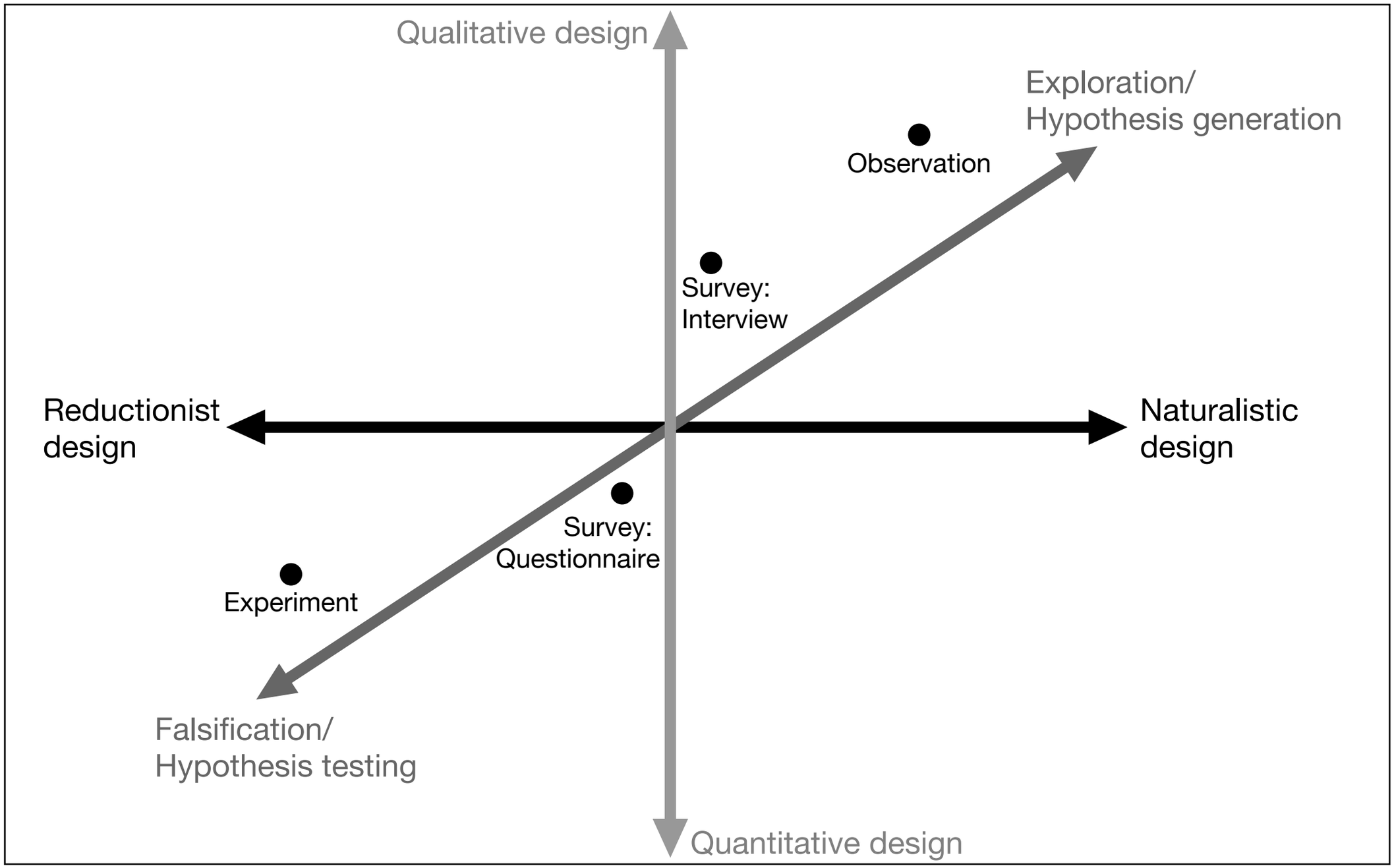


Figure 10

Preferences for data types, analysis approach and hypothesis treatment depend on the method for the collection of primary or secondary data.

	Primary data				Secondary data
	Observation	**Survey**		**Experiment**	**Corpus methods**
		Interview	**Questionnaire**		
Type of data and analysis	natural authentic data, mainly qualitative analysis			reductionist data, mainly quantitative analysis	natural authentic data, mainly quantitative analysis (with ready-made corpora)
Hypothesis treatment	hypothesis (or variable) generation	hypothesis (or variable) generation and hypothesis testing	hypothesis testing	hypothesis testing	hypothesis testing

Observation

Definition and observer's paradox

2.2

Observation is a collection of spontaneous, perceivable data without systematic verbal interaction between the researcher and participants.

- systematic scientific method mainly to explore hypotheses or variables, but also for hypothesis testing
- object of observation: language behaviour of speakers in its natural setting
- high degree of naturalness and authenticity
- comprehensive data, containing a multitude of different variables that can be analysed mainly qualitatively, but also quantitatively

Observer's Paradox

> The aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation.
> (Labov 1972: 209)

Observation

Definitional criteria and resulting types

2.2

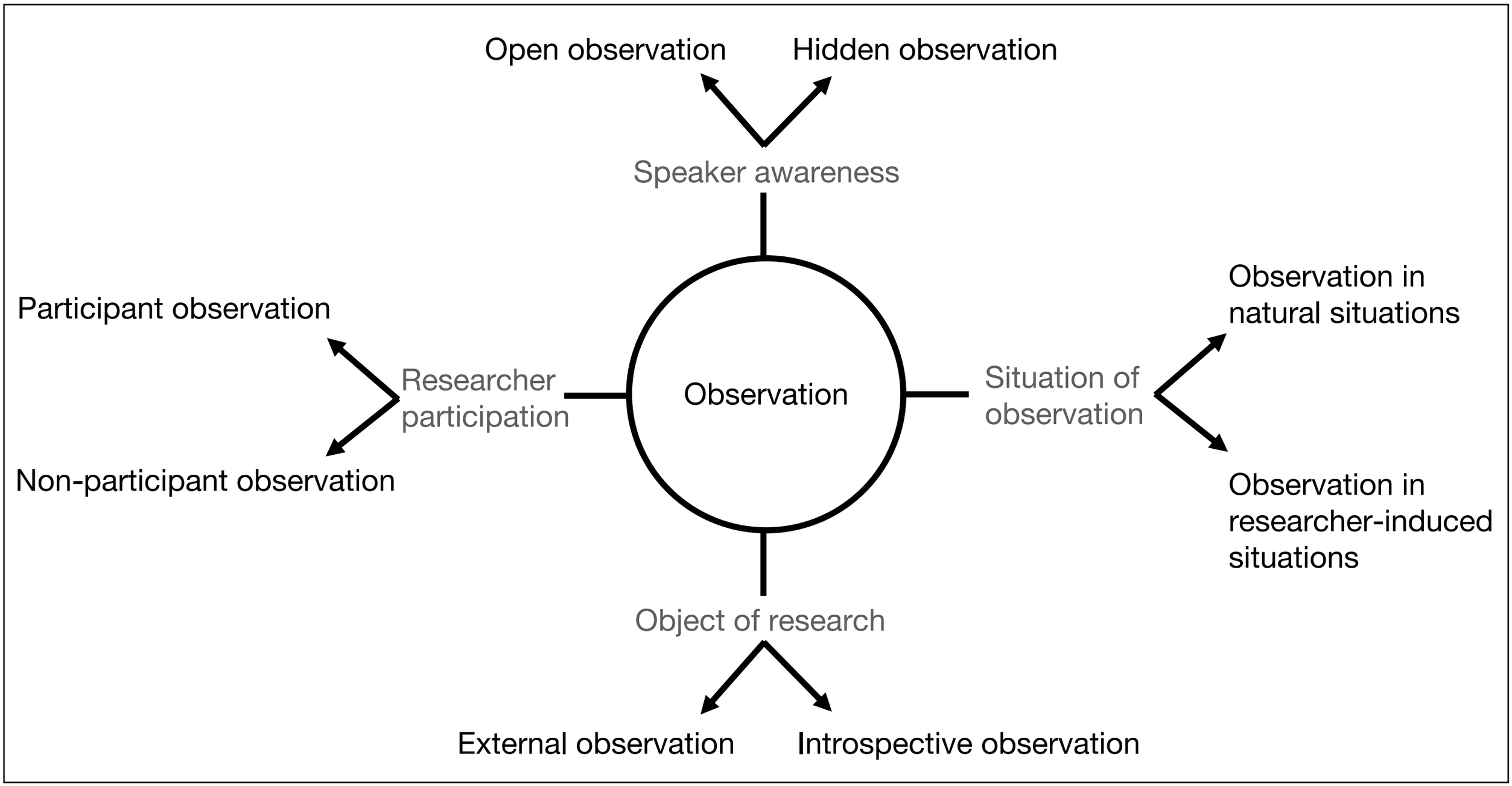


Figure 11

Researcher attitudes can lead to two kinds of **personal bias** that may negatively affect the research process:

- **observer bias**: expectations that researchers can develop because of their awareness of the hypothesis being tested (Cordaro & Ison 1963)
- **researcher bias**: distinction between two perspectives during data analysis (Pike 1967[2])

  → *emic perspective*: 'insider's view' on the collected data

  → *etic perspective*: external (academic) view only

The **observational research plan** includes at least the following items:

object of research

- Who will you observe and do you have access to the relevant speaker community? How many speakers can you manage to observe?

location and time

- When and where do you collect your data? On how many occasions can you collect data?

longitudinal or cross-sectional design

type of observation

data documentation

- Do you need recording devices in addition to handwritten notes to collect data? Additional devices may intensify the observer's paradox or may be problematic regarding research ethics.

coding scheme

The **coding scheme** is the core of an observational study.

It is a category system including the units of observation as delineated by:

- number and kind of categories to be observed

  → directly observable or indirectly inferable?

  → how many (can be plausibly observed)?

- definition of variables and their values/levels based on categories

Observation

Strengths and weaknesses

2.2

- useful for exploratory research, research with special populations (e.g. non-verbal infants, indigenous people), and in remote areas
- investigation of latent variables or unconscious processes in speakers' minds
- circumvent biased self-reports by speakers

- time consuming in data collection and analysis
- error-prone and sometimes less replicable due to flexibility during data collection
- susceptible to personal bias

Observation

Exercises, preferably as homework

2.2

1. Conduct an observation of about 30 minutes at the market place or at any other public space in your town and write a protocol. What was your observational focus? Which coding scheme did you use and which problems did you face?

2. Conduct an observation sitting in a public place with a newspaper which has a big hole in its centre. Observe the people through that hole and collect information on how you feel in that situation, how people react on you, and what impact these two parameters have on the observation.

3. Conduct an introspective observation on your personal language use for 7 days. Observe which language varieties or registers you use depending on the communicative situation or interlocutor. If you are bi-/multilingual, observe under which circumstances you use either language. Develop a coding scheme which includes variables interesting for an external observation.

4. Observe the language behaviour of a group of 2-3 friends or fellow students first by conducting a participant observation and then by conducting a non-participant observation (preferably both in the same communicative situation). Spend about 30 minutes for each observation and write a protocol. Which coding scheme did you use and did the type of observation influence your findings (if so, in what way)?

5. Think about socio-linguistic variants (phonetic, syntactic, or lexical) in your native language. Discuss the advantages or disadvantages of investigating them in researcher-induced situations or in fully natural situations. If possible, collect pilot data on a variant in both kinds of situations.

Survey

Definition

2.3

Survey is a research method utilising communication between researcher and participants. Surveys aim to collect verbal information from the participants on a specific research topic.

- collection of elicited language data (including rare phenomena) with a qualitative or quantitative approach
- suited for exploratory and hypothesis-testing approach
- suitable for collection of not directly perceivable data, i.e. speakers‘ beliefs and attitudes, their behaviours, and demographic characteristics
- less natural data compared with observational methods

Survey types are distinguished based on their mode of administration.

- written administration: questionnaire
  - further subtypes based on medium of communication
- oral administration: interview
  - further subtypes based on question format and number of informants

→ Questionnaire and interview tend to correlate with either a quantitative or a qualitative research design/analysis.

pages 53-56

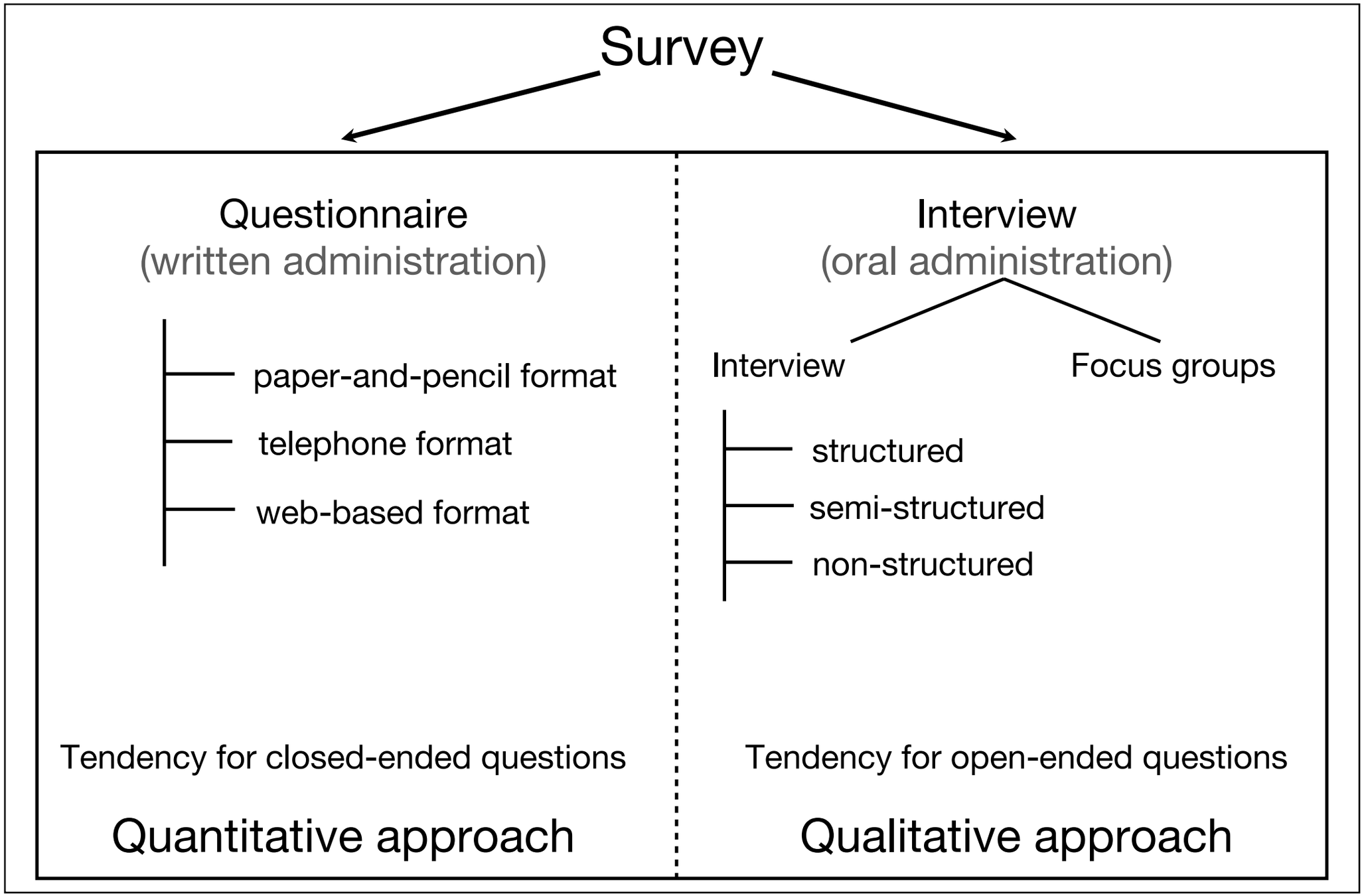
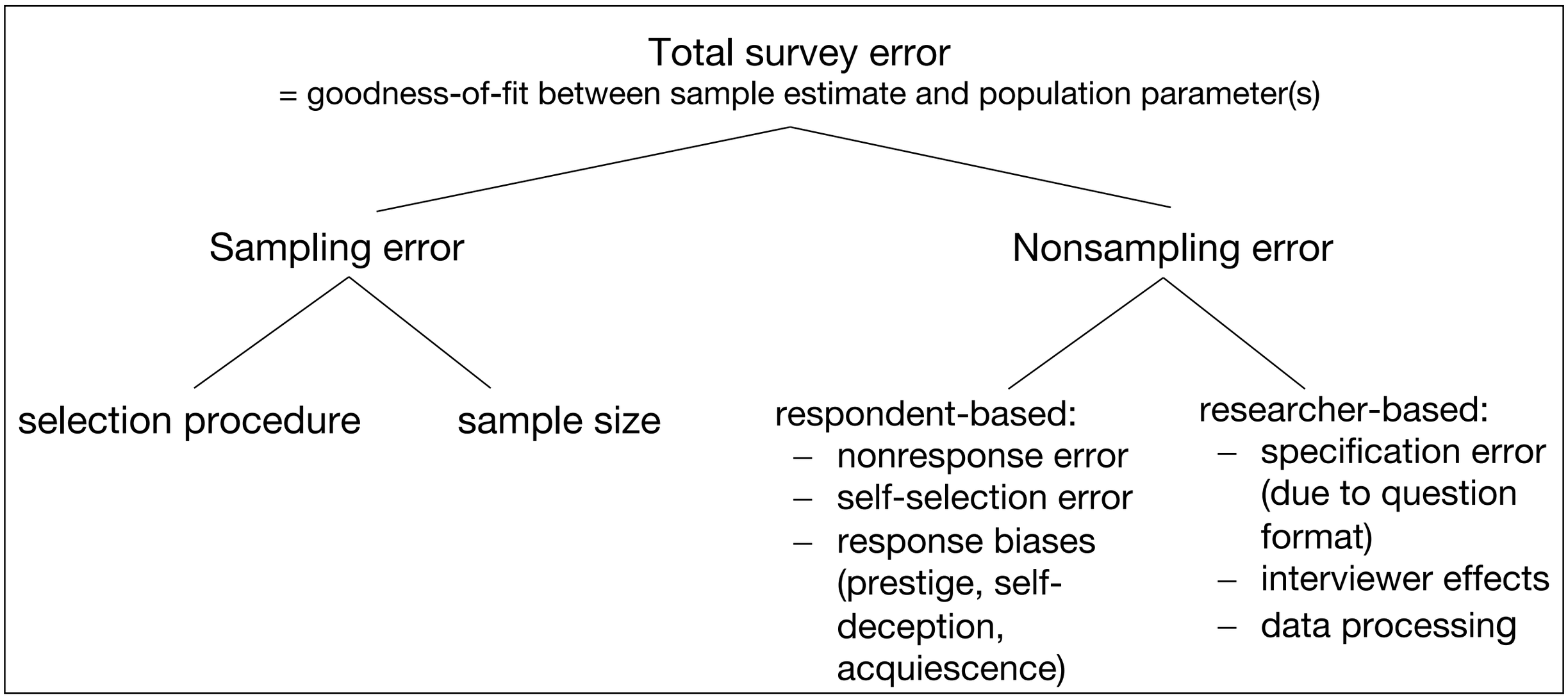

Survey
Questionnaire
(written administration)
paper-and-pencil format
telephone format
web-based format
Tendency for closed-ended questions
Quantitative approach
Interview
(oral administration)
Interview
Focus groups
structured
semi-structured
non-structured
Tendency for open-ended questions
Qualitative approach

Figure 12

The quality of a survey depends on how large the total survey error is. Researchers strive to minimise the total survey error.



→ for more information about sampling see Section 1.2

Figure 13

The question format affects the internal validity of the survey and so requires particular attention during survey preparation, especially regarding 4 parameters:

kind of question

- closed-ended questions vs. open-ended questions
- type of closed-ended question: alternative (single) vs. multiple responses, binary vs. polytomous response categories, ranking vs. rating scales
- type of rating scale: e.g., Likert scale with uneven or even number of categories

wording

- avoid ambiguity and unnecessary complexity
- use a language variety or register appropriate for the sample

number

- avoid too many redundant questions
- open-ended questions may take more time to answer

order

- more general or open-ended questions should precede more specific or closed-ended questions
- do not begin with socio-cultural or personally sensitive questions

Discuss the amount of information obtainable with the different question formats below. Can you think of particular problematic aspects of using the question formats for certain types of research questions?

closed-ended questions

number of response choices

(a) single choice: What is your mother tongue? ☐ - English ☐ - Other

(b) multiple choice: Do you speak any of the following languages? ☐ - Hindi ☐ - English ☐ - Punjabi

number of response categories

(a) binary: Do you speak French? ☐ - yes ☐ - no

(b) polytomous: Do you speak French?
☐ - yes (native) ☐ - yes, fluently ☐ - yes, somewhat ☐ - no, not at all

rating scales

Rate the acceptability of the following sentence: 'The horse raced past the barn fell'.

(a) 7-point Likert scale: very acceptable ☐ - ☐ - ☐ - ☐ - ☐ - ☐ - ☐ very unacceptable

(b) 4-point Likert scale: very acceptable ☐ - ☐ - ☐ - ☐ very unacceptable ☐ - no preference

open-ended questions

(a) List the languages you speak according to your level of proficiency.
Begin with your native language(s): _____

(b) How often do you use swear words in public? _____

Survey

Strengths and weaknesses

2.3

- helpful to collect large amounts of data within shorter period of time
- useful to collect data on psychological constructs that cannot be directly observed
- can be combined with various types of research designs (quantitative, qualitative, explorative, hypothesis-driven) and be adjusted for various groups of respondents/informants

- susceptible to several sources of errors that potentially corrupt the data
- less natural situation of data collection
- certain level of language awareness, language proficiency or literacy is required for respondents/informants and researchers depending on how the survey is administered (written or oral, speaker community, ...)

Survey

Exercises to do in class or as homework

2.3

1. Prepare a survey on bi-/multilingualism. Choose whether your survey will be carried out as a semi-structured interview with open-ended questions or as a questionnaire with closed-ended questions. Find 10–15 questions and recruit about 5 bi-/multilinguals as respondents. Discuss your experience of designing and conducting the survey in class.

2. Find two bi-/multilingual speakers and conduct a non-structured interview about how they grew up speaking several languages and how they use their native languages today. Compare the interviews and discuss to what extent the results from non-structured interviews may generalise to other speakers.

3. Prepare a questionnaire on taboo words using (mainly) closed-ended questions. Find 10–15 questions and recruit about 5 respondents from different age groups (adolescents, young adults, older adults). Discuss your experience of designing and conducting the survey in class.

4. Find a linguistic phenomenon that shows variation in acceptability or grammaticality according to the literature. Prepare a questionnaire using acceptability rating scales (e.g., 7-point Likert scale) for about 10 instantiations/items of this phenomenon and ask (i) linguists and (ii) non-linguists to rate them. Discuss the influence of expert knowledge and respondent awareness on survey results in class.

Experiment

Definition

2.4

In an 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.

- researcher controls the confounding variables and the setting of data collection.
- reductionist approach, i.e. maximum possible reduction to a few variables, with quantitative focus
- hypothesis-testing approach with the goal to identify a cause-effect relationship or at least significant correlations between variables
- least natural data compared with survey and observation

Experimental types are classified according to the experimental setting and quality of the experimental design.

- experimental setting: laboratory vs. field
- experimental design: three design criteria (Kirk 2003)
  - randomisation
  - replication
  - control of confounding variables


pages 62-64

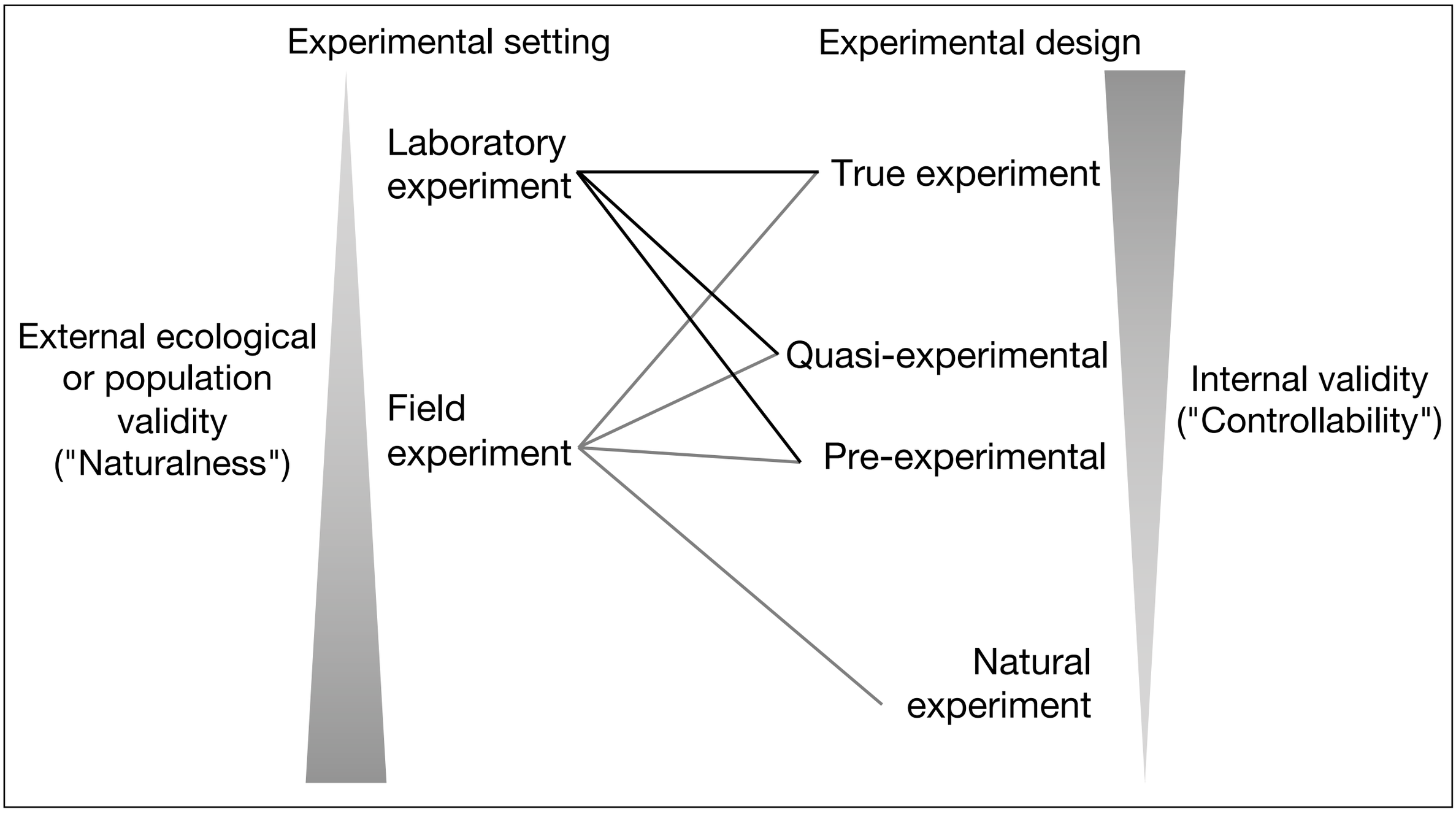

Experimental setting
Experimental design
Laboratory experiment
True experiment
Quasi-experimental
Field experiment
Pre-experimental
Natural experiment
External ecological or population validity ("Naturalness")
Internal validity ("Controllability")

Figure 14

Linguistic experiments include:

experimental design

- Experimental type and control for confounding variables?

technical equipment for data recording (→ type of data)

- Applicable in laboratory or field setting?

} specify the variables

stimuli/test materials

- Which ones and how many of each kind?

experimental task

- Which one(s)?

} implement the levels of the independent variable(s)

implementation of research hypotheses as statistical hypotheses

- Kinds of descriptive and inferential statistics?

Confounding variables negatively affect the validity of experiment.

They result from:

- participants
- researchers
- experimental task and stimuli
- specifics of data collection
- basic experimental design

They can be accounted for by:

- randomisation
- keeping the confounding variable(s) constant
- blocking or matching
- systematic variation
- exclusion of the confounding variable(s)
- blind protocol

pages 67-68

Kinds and function of experimental stimuli/items

- number and kind chosen together with experimental task
- verbal or non-verbal (e.g., pictures, objects, ...)
- used to trigger a response in the participant (= behaviour measured as the dependent variable)
- often complemented by filler or distractor items to mask the nature of the independent variable(s) and/or experimental task

Two basic kinds of experimental stimuli/items, with a fluid transition between the two:

→ **minimal pair**: items implement levels of the independent variable(s) so as to differ systematically in only one aspect, e.g.:

1) voiced vs. unvoiced syllable onset:
ba-da-ga vs. *pa-ta-ka*

2) literal vs. figurative meaning of a word (here: shark):
Sharks live in the ocean vs. *Lawyers are sharks*

3) grammatical vs. ungrammatical sentence:
The airplane took us to the island vs. **The airplane took we to the island*

- reduced to the linguistic domain/variable of interest, i.e. natural variability in, e.g., lexical frequency, length, plausibility or predictability held constant

- presented in isolation and with several repetitions (of different lexical instantiations)

- may involve a violation paradigm, i.e. the comparison of grammatically licensed vs. unlicensed structures (see example 3)

Two basic kinds of experimental stimuli/items, with a fluid transition between the two:

→ **natural (or naturalistic) stimuli**: naturally occurring items implement levels of the independent variable(s) and possibly differ in other aspects as well, e.g.:

1) excerpts from existing text genres or registers (e.g., short stories, newspaper, etc.)
2) items derived from corpus queries
3) (naturalistic) items developed to mimic naturally occurring language

- presented in natural context with little or no modification in terms of, e.g., normalising lexical frequency, plausibility, word length or predictability
- typically exclude violations, unless inherent in the chosen natural materials
- may include minimal pair-items in order to increase their naturalness

Experiment

Exercise to do in class

2.4

This exercise follows the basic procedure of the gating paradigm that is used to study word recognition in psycholinguistics (Grosjean 1980).

Instruct the students to write down the first 2-3 words that come to their minds upon hearing 3 sentence fragments. These sentence fragments vary in how they constrain what word may plausibly end each sentence (cf. Zwitserlood 1985, examples 1-3).

1. Start with a neutral sentence: “The next word is ...” Ask students to read aloud their responses and reflect on whether there is any relationship between them (either in class or in small groups).

2. Now read aloud a medium-constraint sentence: “The men mourned the loss of their...” Ask students to read aloud their responses and reflect on whether there is any relationship between them (either in class or in small groups).

3. Finally, read aloud a high-constraint sentence: “The hurricane had capsized the freighter. The men mourned the loss of their ...” Ask students to read aloud their responses and reflect on whether there is any relationship between them (either in class or in small groups).

Discuss in class how context influences word processing and what this entails for researchers’ choice of using stimuli based on minimal pairs that are presented in isolation vs. natural/naturalistic stimuli.

Experimental tasks specify:

- how participants engage with the experimental stimuli/items
- the kind of response participants give
- how and when participants respond

Experimental tasks can impact data quality:

- task effects: specificity, sensitivity and ecological validity of the task, speed-accuracy trade-off (“be either fast or accurate”)
- task demands: difficulty of the task leading to floor or ceiling effects

Basic kinds of experimental task

1. **decision-related tasks:** grouping of stimuli according to some parameter(s)

2. **memory-related tasks:** memorising or learning of stimuli with subsequent test

3. **elicitation-based tasks:** producing a spoken, signed or written utterance according to some parameter(s)

4. **comprehension tasks:** listening or reading

→ for more information about the kinds of experimental task see Section 7.3

Experiment

Strengths and weaknesses

2.4

- only method allowing to investigate cause-effect relationships between variables
- reductionist, highly controlled design allows for detailed investigation with high internal validity
- suitable to study linguistic knowledge and processes of which participants are not consciously aware

- high control of the experimental setting limits feasibility outside of the lab
- trade-off between internal validity and external validity, often at the expense of external ecological and population validity
- Certain types of experiments are biased towards samples with particular linguistic or educational characteristics, thereby reducing opportunities for cross-linguistic comparison.

Experiment

Exercises to do in class or as homework

2.4

1. Find a research paper with an experimental approach and summarise the main components of the experimental design, i.e., hypothesis, method for data collection, experimental task, possible confounds, etc. Discuss to what extent the study meets the quality criteria of reliability, internal and external validity, and how this influences the experimental results.

2. Find a research question from the domain of phonology, syntax, semantics or pragmatics and develop an experimental design, focusing particularly on the construction of sample items following the minimal-pair logic. Discuss in class what problems you have encountered when developing minimal-pair stimuli for (one of) the above domains and to what extent they may influence data interpretation.

3. Prepare four lists of real words and nonce words. Real words come from two lists: (1) 15 infrequent words (nouns, verbs or adjectives), (2) 15 frequent words from the same language as used in list (1). Nonce words come from 2 lists: (3) 15 pseudowords (strings without a meaning that are pronounceable in the language you chose for lists 1 & 2) and (4) 15 non-words (unpronounceable strings). Mix the four lists with each other so that their members occur randomly. Find three participants and ask them to find the real words “as fast as possible”, giving them a time limit of max. 1.5 minutes (with a paper-and-pencil format). Then ask three further participants to find the real words “as accurately as possible”, giving them no time limit. Record both the total time each participant takes to finish the task and the number of errors (missed real words and nonce words marked as real). Discuss the results from both participant groups in the light of the speed-accuracy trade-off phenomenon.

Mixed-methods designs: ‘a procedure for collecting, analyzing, and mixing quantitative and qualitative data at some stage of the research process within a single study in order to understand a research problem more completely’ (Ivankova & Creswell 2009: 137)

- systematic combination of quantitative and qualitative data, analyses, and interpretation
- suitable for, but not limited to, the validation of findings from single-method designs
- elaborate description and explanation of multifaceted research problems

Mixed methods

Definitional criteria and resulting types

2.5

Three classification criteria (Ivankova & Creswell 2009: 138-139):

- **timing**: sequential vs. concurrent application of quantitative and qualitative data collection
- **weighting**: prioritisation of one data type over the other or not
- **mixing**: kind of integration of quantitative and qualitative data

Resulting types of mixed-methods design (Ivankova & Creswell 2009: 139-145):

- **explanatory design**: from quantitative patterns to qualitative explanation
- **exploratory design**: from qualitative exploration to quantitative validation
- **triangulation** (or convergent) design: concurrent data collection
- **embedded design**: nesting of one design within the other

pages 71-72

Relation between the main study designs and quantitative or qualitative data and analysis

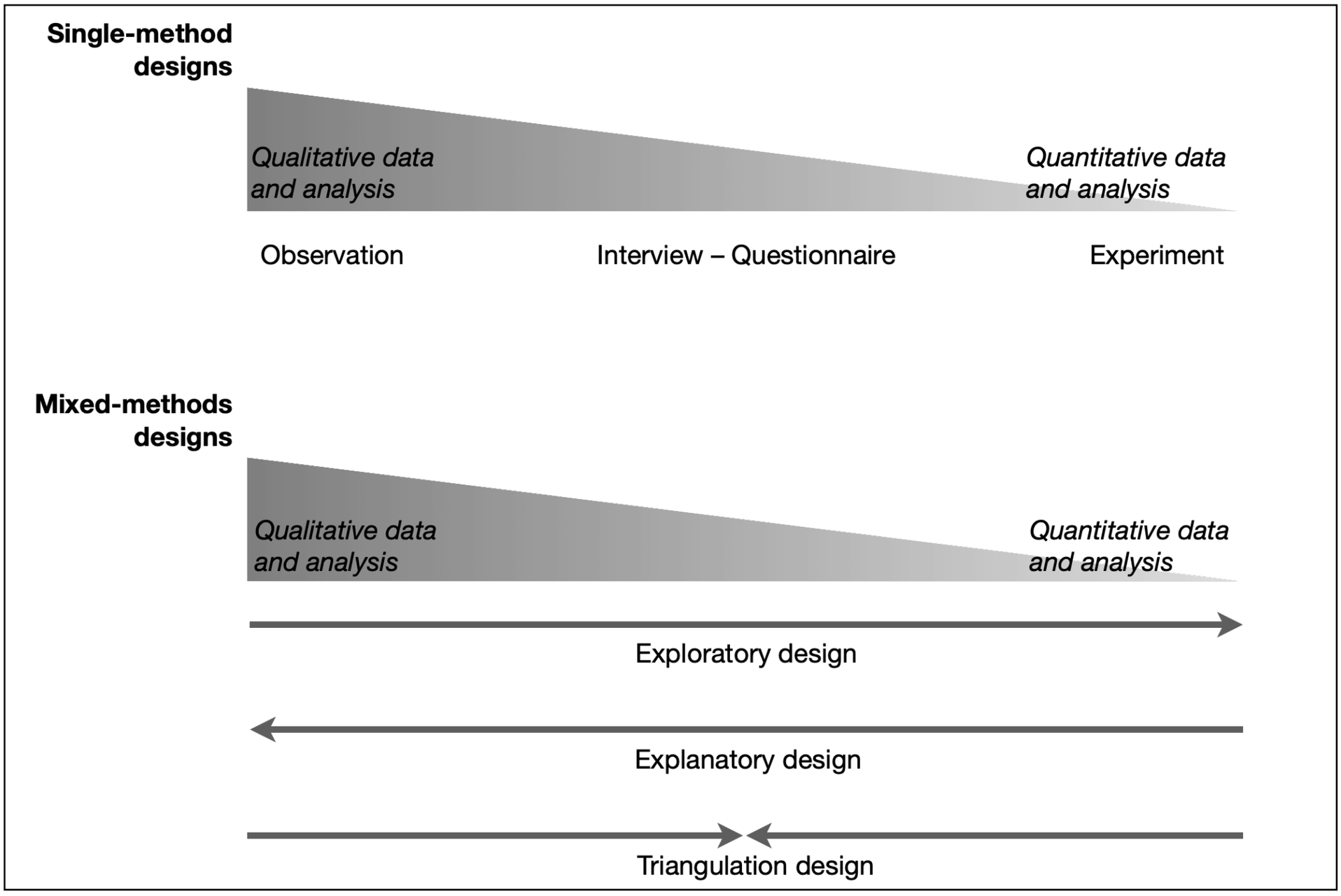


Figure 15

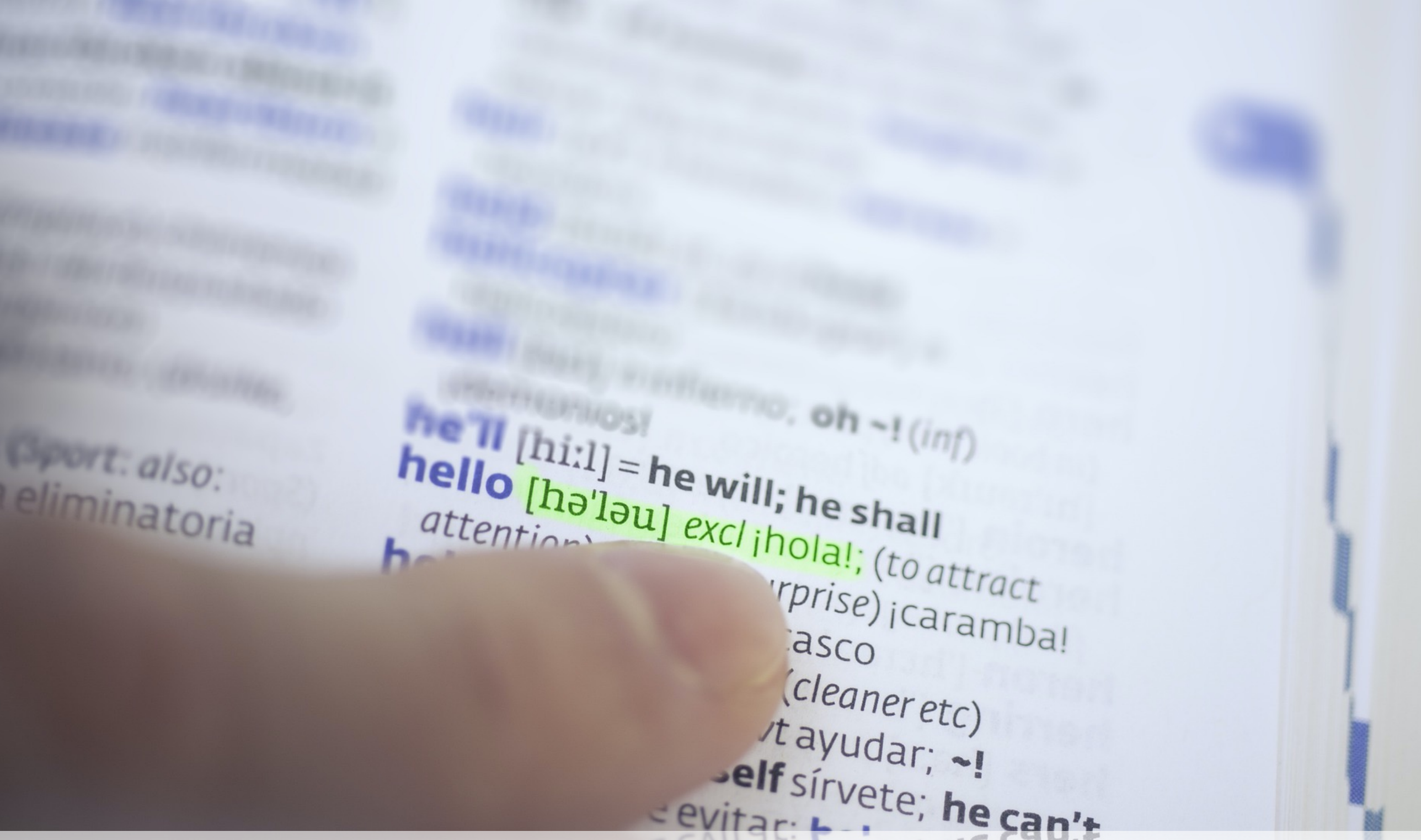
Mixed methods

Strengths and weaknesses

2.5

- allow for compensation of a single method's shortcomings regarding validity or reliability
- integrate multiple perspectives to investigate a complex research problem in a single study

- more complex than single-method designs (data collection, analysis, interpretation)
- require more resources (funding, particular skills, time etc.)



Language documentation and descriptive linguistics

Chapter 3

Chapter 3: Outline

Definitions 3.0

Research aims and questions 3.1

Documentary and descriptive approach 3.2

Methodology 3.3

Endangered languages, fieldwork, compilation & elicitation

Basic research findings 3.4

Definitions

3.0

Language documentation is concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties.
(Gippert, Himmelmann & Mosel 2006: v)

… "**documentary linguistics**", which can briefly be defined as the collection or gathering of linguistic data through a variety of methods and techniques, with a focus on reliability, representativity, and archivability.
(Chelliah & de Reuse 2011: 10f)

Description of a language is an activity (and derivatively, its result), that formulates, in the most general way possible, the patterns underlying the linguistic data. Its purpose is to make the user of the description understand the way the language works.
(Lehmann 1999: 10)

The job of **descriptive linguistics** is to describe individual languages as perceptively and rigorously as possible, with maximal accountability to a naturalistic corpus of data ideally collected within a broad program of language documentation […] to ensure that the full spectrum of language structures are represented.
(Evans & Dench 2006: 3)

Language documentation/documentary linguistics:

- **research aim**: documentation of primary language data in order to preserve it and ensure future accessibility
- **research question**: How do native speakers behave linguistically, i.e., how do they use language to express various facts and topics in different naturally occurring situations and contexts?
- **secondary or applied goals**: reclamation/revitalisation of endangered languages, political or religious purposes

Descriptive linguistics:

- **research aim**: structural description of a language
- **research question**: How is a language structured, i.e., what are the underlying structural patterns of a certain language?

see page 81 for more detailed research questions

Documentary & descriptive linguistics:

two **separate** but **closely interrelated, complementary, or even partially overlapping** subdisciplines

Documentary approach	Descriptive approach
collection of natural language data (primarily endangered languages) for the purpose of preservation and accessibility to others for future purposes → edited **representative text corpora** gained by the recording of oral texts, the collection of written texts & their editing	**analysis of language data** for the purpose of describing the underlying structural patterns → **comprehensive grammars** gained by systematic elicitation and/or the analysis of natural language data
linguistic fieldwork (for the documentation & description of barely studied or unstudied languages)	

Objects of research:

barely studied or unstudied languages or language varieties

- languages in remote areas
- minor languages in areas with dominant major languages
- less studied aspects and varieties of major languages
- **endangered languages** (focus in language documentation)

  → a gradual phenomenon

  → criteria of endangerment:

  1. low absolute/relative number of speakers (decreasing language transmission to infants)
  2. little prestige and benefit of the language (decreasing use)
  3. no written documents, no standard variety (high degree of variability)

How can you find such languages or language varieties?
And what does it mean to study them?

Documentary & descriptive fieldwork:

- collection of primary language data
- work with native speakers (in their local environment)

> [linguistic fieldwork is] Going into a community where a language is spoken, collecting data from fluent native speakers, analyzing the data, and providing a comprehensive description, consisting of grammar, texts and dictionary.
> (Dixon 2007:12)

> Fieldwork describes the activity of a researcher systematically analyzing parts of a language other than one's native language (usually one the researcher did not speak prior to beginning fieldwork) within a community of speakers of that language, prototypically in their native land, living out their existence in the milieu and mental currency of their native culture.
> (Everett 2001: 168)

> We define descriptive linguistic fieldwork as the investigation of the structure of a language through the collection of primary language data gathered through interaction with native-speaking consultants.
> (Chelliah & de Reuse 2011: 7)

> The ideal way to study the language of a traditional community is in situ, living with the village, learning as much of the social customs of the people as possible.
> (Foley 2002: 131)

Kinds of fieldwork – variation along these parameters:

- **language of interaction**: monolingual vs. bilingual fieldwork
- **number of research participants**: fieldwork with single main informants/interlocutors vs. with entire groups
- **research setting**: fieldwork with native speakers in their natural environment (monolingual vs. multilingual settings) vs. in the researcher's environment
- **field site**: rural vs. urban environment
- **language/language variety**: oral vs. written
- **researcher's background**: part of the language community vs. no contact with it prior to fieldwork
- **researcher**: single person vs. research groups
- **research aim**: language documentation vs. descriptive work
- **researcher's involvement in the language community**: participation vs. no participation

Preparing fieldwork - topics to consider:

- travel & accommodation on-site
- medical & political situation
- duration & time of fieldwork
- local contacts & cultural customs
- research equipment & needed skills
- local infrastructure & personal needs
- adequate compensation of research participants & fieldwork ethics
- funding & research permission (from academic institutions and from local authorities)
- psychological challenges of fieldwork
- existing information on the language/language variety/language family to built upon & scholars with experience in the field

Watch the film ‘The Linguists’ (Harrison & Anderson 2008) to get familiar with documentary and descriptive fieldwork. Which challenges are there and how can you deal with them?

Compilation of text corpora:

- data collection: **recording** of oral texts (representative, multiple genres)

  collection of written texts (representative, multiple genres)

  addition of **metadata** (text type, source, date, setting, etc.) to each text file

- data editing: **transcription** of oral data
  → phonemic transcription (IPA-based) vs. orthographic transcription
  → including vs. excluding the encoding of prosodic and paralinguistic features

  translation of texts into a widely known language
  → free translation of the meaning vs. literal translation

  segmentation of the text into annotation units

  annotation of texts
  → interlinear morphemic glossing
  → including/excluding glossing of other information (semantic or pragmatic)

Record a short sequence and transcribe it – using the IPA (www.internationalphonetic association.org/content/ipa-chart). Annotate a written text – using the Leipzig glossing rules (www.eva.mpg.de/lingua/resources/glossing-rules.php).

Software tools: ELAN, TOOLBOX, etc.

→ ELAN - documentation & annotation tool of the Max Planck Institute in Nijmegen/NL:
https://tla.mpi.nl/tools/tla-tools/elan/

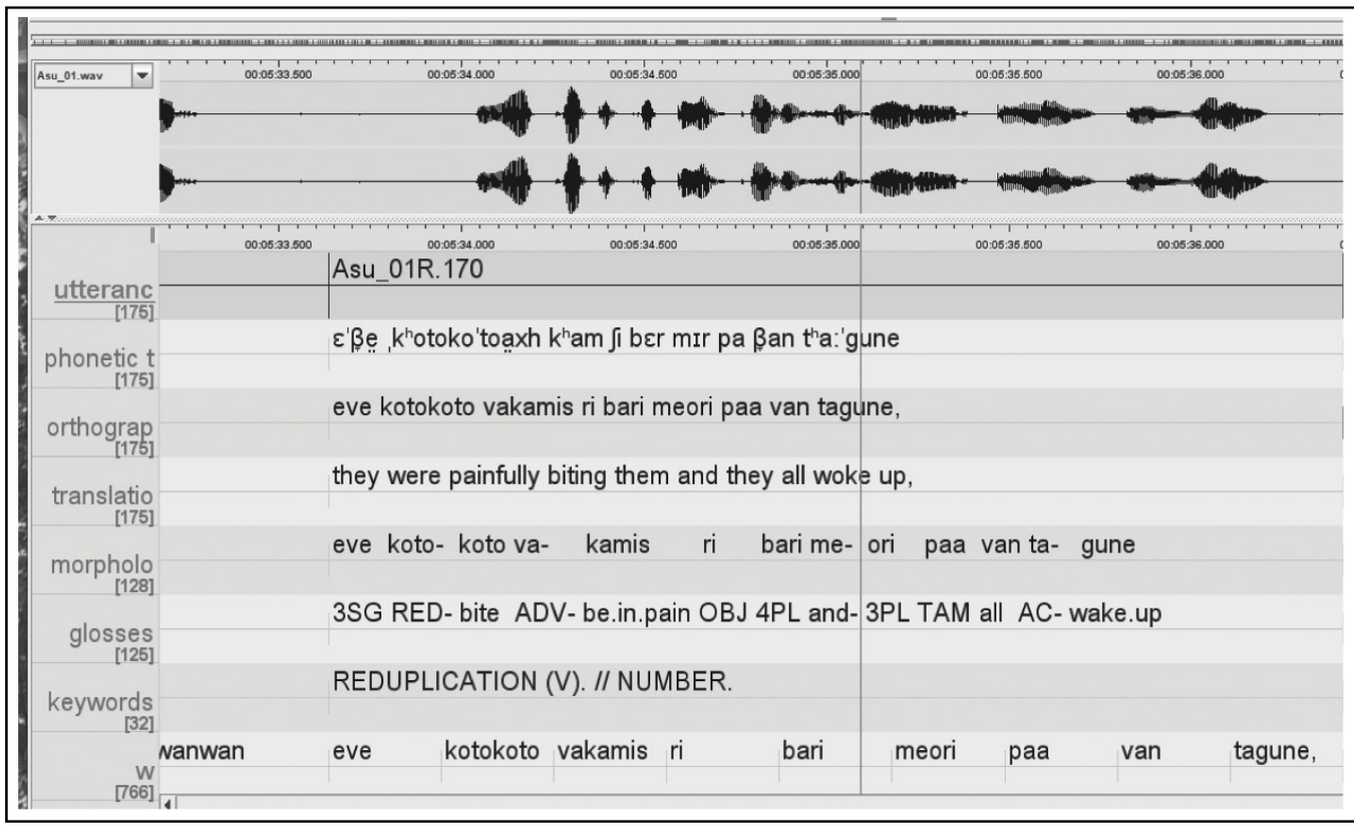


Figure 16

1. transcription time-aligned to sound file
2. translation linked to transcription tier (segmented per sentence)
3. annotation tier (interlinear morphemic glossing) linked to the transcription tier (segmented per morpheme)

Grammar and dictionary writing:

- **corpus analysis** → unconscious language use/performance
- **elicitation** (i.e., systematically controlled data collection) → reflected language competence

Elicitation techniques:

- bilingually: - translation into the target language
  - back-translation from the target language
- monolingually: - target language interrogation
- bi-/monolingually: - naming of items or description of situations, actions, etc.
  - target language manipulation (acceptability judgements, correction, substitution of individual elements, creation of words/ phrases/ sentences/ texts with particular elements, or completion of sentences, etc.

Elicitation advices:

1. start with the elicitation of single lexical items, followed by simple sentences and finally complex sentences and phrase-internal structures;
2. elicit minimal pairs;
3. constantly question your own (provisional) analysis;
4. work with several native speakers on the same issues.

Watch Everett's presentation of monolingual elicitation (Everett 2013) and work out by means of which examples and items he tries to get information on which particular linguistic structures.
How does the elicitation of Everett (2013) differ from the elicitation used by 'The Linguists' (Harrison & Anderson 2008) in India (Chapter 9)?

Research outcomes:

- **edited text collections** (corpus) but also collections of single texts
- **reference grammars** but also grammar sketches or descriptions of single linguistic aspects
- **dictionaries** but also simple word lists

→ cross-references

Requirements for edited text collections (Himmelmann 1998; Woodbury 2011):

- extensive
- representative
- expandable

] exhaustiveness (documentation of the entire language)

- opportunistic
- portable
- preservable

] permanent data availability (for future purposes) → archiving

- comprehensible and transparent
- ethical
- reliable

] data and research quality

Requirements for reference grammars

(Payne 1997; Mosel 2006; Noonan 2006; Payne & Weber 2007):

- extensive or comprehensive, covering
  - general information on the language
  - phonetic/phonological descriptions
  - morphological descriptions
  - syntactic descriptions
  - (- semantic and pragmatic issues)
- transparent
  - ascending vs. descending order
  - semasiological vs. onomasiological structure
- maximally theory neutral → 'basic linguistic theory' (Dixon 2010 & 2012)
- reliable and traceable → including language examples
  - descriptive vs. prescriptive data;
  - corpus-based vs. elicitation-based data

Dictionaries: monolingual vs. bilingual

Language typology

Chapter 4

Chapter 4: Outline

Definitions 4.0

Research aims and questions 4.1

The typological approach 4.2

Methodology
Cross-linguistic comparison, sampling, data sources 4.3

Basic research findings 4.4

Explanation and interpretation of the findings 4.5

Definitions

4.0

In its most general sense, **typology** is [the] classification of languages or components of languages based on shared formal characteristics. [...] typology has the goal of identifying cross-linguistic patterns and correlations between patterns. [... Therefore,] (a) Typology utilises cross-linguistic comparison, (b) typology classifies languages or aspects of languages, and (c) typology examines formal features of languages.
(Whaley 1997: 4,7)

To put it very simply, **linguistic typology** concerns itself with the study of structural differences and similarities between languages. [...] In other words, a driving force is to try to establish recurrent patterns across languages, in order to answer the questions “what is out there?”, “where does it occur?” and “why do we have particular patterns?”
(Velupillai 2012: 15)

The broadest and most unassuming linguistic definition of ‘**typology**’ refers to a classification of structural types across languages. In this [...] definition, a language is taken to belong to a single type, and a typology of languages is a definition of the types and an enumeration or classification of the languages into those types. [...] This definition introduces the basic connotation that ‘typology’ has to contemporary linguists: typology has to do with cross-linguistic comparison of some sort.
(Croft 1990: 1)

Linguistic typology compares languages to learn how different languages are, to see how far these differences may go, and to find out what generalizations can be made regarding cross-linguistic variation.
(Daniel 2011: 44)

Language typology:

- **research aim**: cross-linguistic comparison
- **research questions**:

  a. primary questions:

  - To what extent do the languages of the world share linguistic patterns?
  - And to what extent do they differ structurally?
  - How are the different linguistic features distributed in the languages of the world?

  b. follow-up question:

  - Why do these structures occur which are actually to be found in the languages of the world (and no theoretically possible others)?

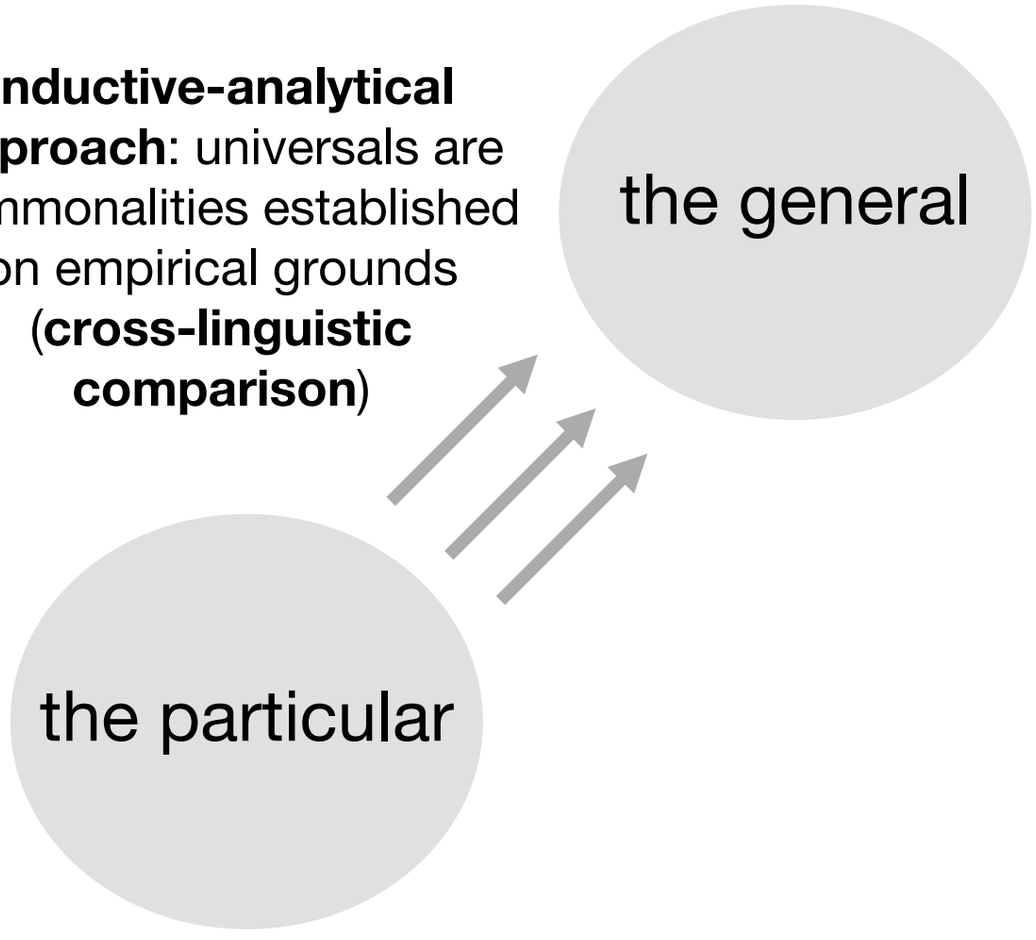
see page 108-109 for more detailed research questions

Universals in language typology

vs.

Universal grammar (UG) in generative approaches

inductive-analytical approach: universals are commonalities established on empirical grounds (**cross-linguistic comparison**)



deductive-analytical approach: UG is a theoretically deduced framework

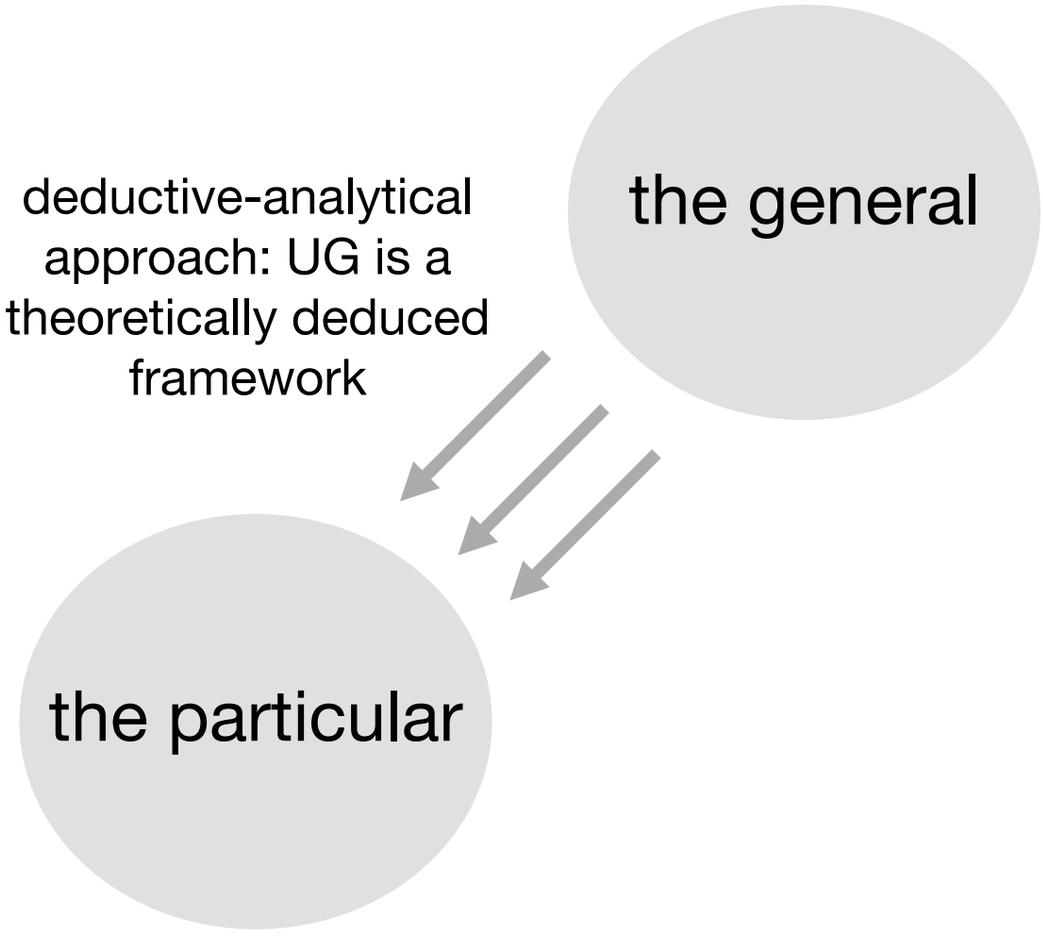


Figure 17

Cross-linguistic comparison

→ comparability of languages:

- **metalanguage** (i.e., description and categorisation tools that are not language-specific but apply to all languages)
- **comparable structural features** according to which each language can be classified

  → identification of parameter/variables (one parameter or two or more correlating parameters)

  → identification of their values (different types regarding the parameters; each languages can clearly be assigned to one of these types)

  → In lack of comparable formal grounds, cross-linguistic comparison must ultimately be carried out on a semantic/pragmatic-cognitive level (Croft 1990:12).
- **simplification** of language representations

Determine the parameters and their values of the universals in Section 4.4.


see page 125 for solutions

How can we study the languages of the world?

- studying all languages is impossible
  (due to the workload, data availability, delimitation language vs. dialect, etc.)
- work with a **language sample** (i.e., a representative selection of languages):
  → **sample criteria** (to avoid bias, i.e., the overrepresentation of certain language groups):
  1. sample size
  2. genetic relatedness (→ language families: Ruhlen 1987, Ethnologue, WALS, etc.)
  3. geographic proximity (→ linguistic areas)
  4. environmental similarity: socio-cultural relatedness & similarity of natural surrounding
  5. typological similarity
  6. data accessibility

Discuss the sample criteria and their relevance/importance for sample building.

- areal typology: languages from a certain geographic area

Genetic relatedness

Macrofamilies according to Ruhlen (1987)
macrofamily [number of known languages: living, extinct]

1. Khoisan [31, 2]
2. Niger-Kordofanian [1064, 4]
3. Nilo-Saharan [138, 0]
4. Afro-Asiatic [241, 17]
5. Caucasian [38, 0]
6. Indo-Hittite [144, 36]
7. Uralic-Yukaghir [24, 3]
8. Altaic [63, 3]
9. Chukchi-Kamchatkan [5, 0]
10. Elamo-Dravidian [28, 1]
11. Sino-Tibetan [258, 10]
12. Austric (incl. Austronesian) [1175, 11]
13. Indo-Pacific [731, 17]
14. Australian [170, 92]
15. Eskimo-Aleut [9, 0]
16. Na-Dene [34, 7]
17. Amerind [583, 271]

Language Isolates: Basque, Burushaski, Etruscan, Gilyak, Hurrian, Ket, Meroitic, Nahali, Sumerian
Unclassified languages (New Guinea): Busa, Messep, Nagatman, Pauwi, Porome, Taurap, Warenbori, Yuri;
(South America): Arara, Carabayo, Chiquitano, Guaviare, Kohoroxitari, Mutus, Yari, Yuwana
Pidgins & Creoles [38]

Which languages should be considered in typological studies and why?

Kinds of samples:

- **probability sample**: to study the distribution of probabilities for different linguistic types

  → extremely vulnerable to bias; each language (sub)family & linguistic area needs to be represented according to its size.

- **variety (or diversity) sample**: to discover the spectrum of linguistic variation

  → Languages from as many distinct language (sub)families and linguistic areas as possible need to be considered.

- **random sample**

  → no sample criteria

- **convenience sample**

  → practical criteria (such as data accessibility) play a major role

Variety sample (e.g., Rijkhoff et al. 1993; Rijkhoff & Bakker 1998: 274):

- **sample criteria**: only genetic relatedness – diversity values (DV) calculated for each macrofamily

Table 7. *The distribution of languages in samples of different sizes using DVs*

Families (Ruhlen 1987)	Sample size 30	40	50	60	70	80	90	100	125	150	175	200	250	300	400	500	1000
1. Afro-Asiatic	1	2	2	3	4	5	5	6	8	9	11	12	16	19	26	32	65
2. Altaic	1	1	1	1	1	1	2	2	2	3	3	3	4	5	7	9	17
3. Amerind	2	5	7	9	12	14	16	18	24	29	35	40	51	61	83	104	209
4. Australian	1	2	3	4	4	5	6	7	9	11	13	15	19	23	31	39	79
5. Austric	2	4	5	7	9	11	13	14	19	23	27	31	39	47	64	80	161
6. Basque (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7. Burushaski (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8. Caucasian	1	1	1	1	1	1	1	1	1	2	2	2	3	3	4	5	10
9. Chukchi-Kamchatkan	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
10. Elamo-Dravidian	1	1	1	1	1	1	1	1	1	1	1	2	2	3	3	4	9
11. Eskimo-Aleut	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	4
12. Etruscan (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13. Gilyak (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14. Hurrian (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15. Indo-Hittite	1	1	2	2	3	3	4	4	5	7	8	9	11	14	18	23	47
16. Indo-Pacific	2	3	5	7	8	10	11	13	17	20	24	28	35	42	57	72	145
17. Ket (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18. Khoisan	1	1	1	1	1	1	1	1	1	1	1	2	2	3	3	4	8
19. Meroitic (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20. Na-Dene	1	1	1	1	1	1	1	1	1	2	2	2	3	3	4	5	11
21. Nahali (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
22. Niger-Kordofanian	1	3	4	5	6	7	8	9	12	15	18	20	26	31	42	53	106
23. Nilo-Saharan	1	1	2	3	3	4	4	5	6	7	8	10	12	15	20	25	50
24. Pidgins & Creoles	1	1	1	1	1	1	1	2	2	2	3	3	4	5	6	8	16
25. Sino-Tibetan	1	1	2	2	3	3	4	4	5	6	7	9	11	13	18	22	45
26. Sumerian (LI)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
27. Uralic-Yukaghir	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	6

Build a sample of 30 languages based on Rijkhoff & Bakker 1998. Use only languages for which data (grammars) are available at your institution.

Considerations:

- Given the following simplified language family tree, which two languages of the language family should (not) be included in the sample from a variety point of view?

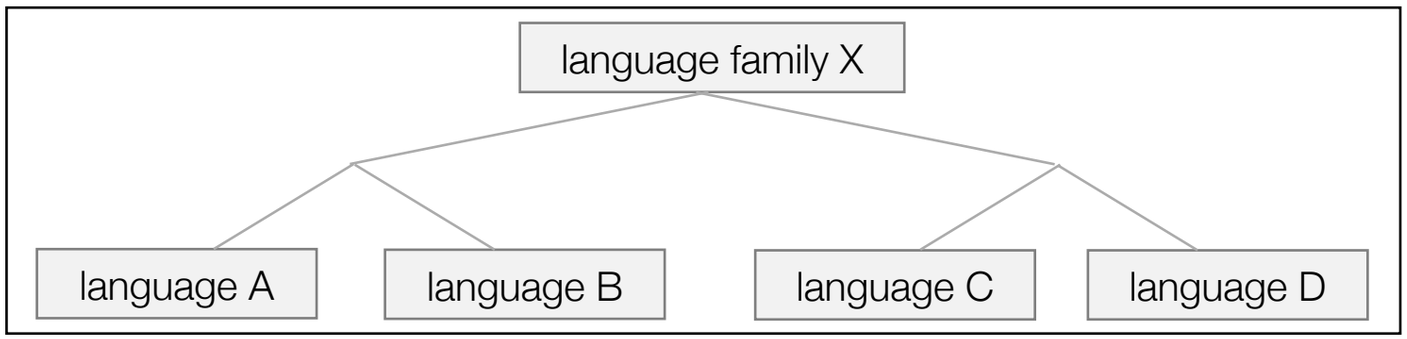


Figure 18

- In case no data is available for a language isolate (LI), which macrofamily would get an additional language included in the sample based on the following diversity values of Rijkhoff & Bakker (1998: 272)?

Afro-Asiatic	55.53	Chukchi-Kamchatkan	2.47	Na-Dene	9.44
Altaic	14.79	Elamo-Dravidian	7.43	Niger-Kordofanian	90.38
Amerind	178.44	Eskimo-Aleut	3.34	Nilo-Saharanan	42.18
Australian	67.58	Indo-Hettite	39.71	Pidgins & Creoles	13.47
Austric	137.41	Indo-Pacific	123.39	Sino-Tibetan	38.52
Caucasian	8.54	Khoisan	6.97	Uralic-Yukaghir	4.93

see page 118-120 for advices

Probability sample (e.g., Dryer 1989, 1992):

- **sample criteria**: genetic relatedness (252 genera) & geographic proximity (6 large areas)

AFRICA:
Northern Khoisan (Xu), *Central Khoisan* (Korana, Nama), *Kordofanian* (Katla, Moro, Masakin, Rashad), *Mande* (Susu, Vai, Mandinka, Gambian Mandinka, Bambara, Mende), *Northern Atlantic* (Fulani, Diola-Fogny), *Ijoid* (Kolokuma Ijo), *Kru* (Seme, Grebo), *Gur* (Toussian, Tenyer, Bimoba, Kirma), *Adamawa-Ubangi* (Mbum, Day, Gbaya, Kaka, Gbeya Bossangoa, Sango, Nzakara, Mba), *Kwa* (Fanti, Twi, Nkonya, Lelemi, Ewe), *Defoid* (Yoruba), *Edoid* (Bini, Engenni), *Igboid* (Igbo, Izi), *Platoid* (Jukun), *Cross River* (Efik), *Kainji* (Duka), *Bantoid* (Noni, Ewondo, Bobangi, Swahili, Luganda, Nkore-Kiga, Luvale, Lamba, Mwera, Shona, Zulu), *Kadugli* (Katcha), *Songhai*, *Saharan* (Kanuri, Tubu), *Maban* (Maba), *Fur*, *Nubian* (Dongolese Nubian), *Surma* (Didinga), *Nera* (Barya), *Nyimang*, *Temein*, *Tama*, *Daju* (Shatt), *Nilotic* (Pari, Dholuo, Acooli, Bor, Bari, Maasai, Karimojong, Sebei, Pokot), *Kuliak* (Tepeth), *Kresh*, *Bongo-Bagirmi* (Sara-Ngambay, Bagirmi, Yulu), *Mangbutu- Efe* (Mamvu), *Balendru* (North Lendu), *Berta*, *Kunama*, *Komuz* (Koma), *Berber* (Berber, Shilha, Tamazight), *Biu-Mandara* (Tera, Margi, Lamang, Gude), *West Chadic* (Hausa, Kanakuru, Angas, Ngizim), *Omotic* (Ometo), *Beja*, *Central Cushitic* (Kemant), *Eastern Cushitic* (Afar, Arbore, Geleba, Somali, Oromo), *Southern Cushitic* (Iraqw), *Semitic* (Chaldean, Biblical Hebrew, Modern Literary Arabic, Colloquial Egyptian Arabic, Sabaic, Ge'ez, Tigre, Amharic, Gouragué, Chaha).

EURASIA:
Basque, *Armenian* (Classical Armenian, Modern Armenian), *Indic* (Pali, Nuri, Welsh, Romany, Sinhalese, Dumaki, Shina, Marathi, Hindi, Urdu, Punjabi, Nepali, Maithili, Bishnupriya, Manipuri), Iranian (Ossetic, Wakhi, Pashto, Persian, Southern Tati), *Albanian*, *Greek* (Modern Greek), *Italic* (Latin, Rumanian, French, Spanish), *Celtic* (Irish, Scots, Gaelic, Breton, Welsh), *Germanic* (Danish, Swedish, Icelandic, Dutch, German, English, Frisian), *Baltic* (Lithuanian), *Slavic* (Russian, Ukrainian, Polish, Czech, Bulgarian, Serbo-Croatian), *Samoyedic* (Nenets), *Ugric* (Hungarian), *Finnic* (Udmurt, Komi-Permyak, Eastern Cheremis, Finnish), *Mongolian* (Dagur, Kalmyk, Buriat, Khalkha), *Tungus* (Evenki, Orok), *Turkic* (Orkhon Turkic, Chuvash, Azerbaijani, Turkish, Uzbek, Karakalpak, Yakut), *Japanese*, *Korean*, *Chukchee-Kamchatkan* (Koryak, Itelmen), *Nivkh*, *Yukaghir*, *Ket*, *Sumerian*, *Hurrian*, *Elamite*, *Kartvelian* (Georgian), *Northwest Caucasian* (Abkhaz), *Nax* (Chechen, Ingush), *Avaro-Andi-Dido* (Avar), *Lak-Dargwa* (Lak, Dargwa), *Lezgian* (Archi, Lezgian), *Burushaski*, *Northwest Dravidian* (Brahui), *Dravidian Proper* (Kolami, Gondi, Koya, Kuvi, Telugu, Tulu, Kannada, Tamil), *Munda* (Kurku, Santali, Mundari, Ho).

SOUTHEAST ASIA & OCEANIA:
Chinese (Mandarin, Hakka, Cantonese, Malayan Cantonese), *Karen*, *Tibetic* (Lepcha, Gurung, Ladakhi, Sherpa, Magari, Kham, Kusunda, Thulung, Limbu, Dafla), *Baric* (Garo, Kachari, Kokborok), *Burmic* (Ao, Bawm, Chingpaw, Burmese, Lahu), *Miao-Yao* (Miao, Mong Njua, Mjen, Pu Nu), Khasi, *Palaung-Khmuic* (Palaung), *Viet-Muong* (Vietnamese), *Katuic* (Katu), *Bahnaric* (Stieng, Chrau, Sre, Brao, Cua), *Khmer* (Cambodian), *Aslian* (Temiar), *Nicobarese* (Car), *Kam-Tai* (Nung, Lao, Thai), *Atayalic* (Atayal), *Paiwanic* (Rukai), *Philippine Austronesian* (Chamorro, Palauan, Central Agta, Balangao, Pangasinan, Kapampangan, Western Bukidnon Manobo, Tagalog, Bikol, Mamanwa, Hiligaynon, Tboli, Tondano), *Sundic* (Sundanese, Toba Batak, Indonesian), *Central-Eastern Malayo-Polynesian* (Mor, Manam, Kaliai-Kove, Patep, Wedau, Iduna, Kilivila, Pokau, Motu, Balawaia, Nissan, Tolay, Tigak, Halia, Mono-Alu, Lenakel, Iai, Dehu, Yapese, Kiribatese, Kusaiean, Mokilese, Ponapean, Puluwat, Sonsorol-Tobi, Sa'a, Arosi, Sakao, Nguna, Rotuman, Fijian, Niuean, Tongan, Samoan, Futuna-Aniwa, Easter Island, Tahitian, Hawaiian), *Andamanese*.

AUSTRALIA - NEW GUINEA:
Finisterre-Huon (Selepet), *East New Guinea Highlands* (Kobon, Agarabi, Gadsup, Usarufa, Awa, Tairora, Hua, Yagaria, Gahuku, Golin, Salt-Yui, Banz, Kewa), *Central and South New Guinea* (Kamoro, Asmat, Auju, Kati, Mombum), *Angan* (Baruya, Kapau), *Marind* (Boazi, Marind, Jaqai), *Sentani*, *Dani-Kwerba* (Lower Grand Valley Dani), *Wissel Lakes-Kemandoga* (Ekari, Moni), *Binanderean* (Guhu-Samane, Suena), *Central and Southeast New Guinea* (Kunimaipa, Koita, Mountain Koiali, Barai, Omie, Yareba, Magi, Daga), *Madang* (Siroi, Amele), *Adelbert Range* (Waskia), *Trans-Fly-Yelmek-Maklew* (Jelmek, Makleu, Kiwai, Jei, Moraori, Kanum), *Kolopom* (Kimaghama), *Torricelli* (Arapesh), *Sepik-Ramu* (Ambulas, Iatmul, Alamblak, Rao, Kire, Autuw), *Bougainville* (Nasioi, Telei), *Yele-Solomons* (Yeletnye), *Mangarayi*, *Nunggubuyu*, *Tiwi*, *Yiwaidjan* (Jiwadja, Maung), *Gunwinyguan* (Ngandi, Gunwinggu, Dalabon, Gunbalang, Wageman), *Maran* (Alawa, Mara), *West Barkly* (Djingili), *Garawan* (Garawa), *Daly* (Maranungku, Malakmalak), *Wororan* (Ngarinjin), *Tangkic* (Yukulta), *Pama-Nyungan* (Muruwari, Gidabal, Ritharngu, Djapu, Uradhi, Anguthimri, Ngawun, Guugu Yimidhirr, Yidiny, Dyirbal, Wargamay, Nyawaygi, Margany, Gumbaynggir, Yaygir, Dharawal, Ngiyambaa, Madimadi, Wembawemba, Pitta-Pitta, Diyari, Alyawarra, Aranda, Garadjari, Yindjibarndi, Thargari, Watjarri, Western Desert, Gugada, Warlpiri).

NORTH AMERICA:
Eskimo-Aleut (West Greenlandic, Yup'ik), *Haida*, *Tlingit*, *Athapaskan-Eyak* (Hupa, Slavey, Chipewyan, Sarcee, Western Apache, Navajo), *Kutenai*, *Wiyot*, *Yurok*, *Algonquian* (Blackfoot, Cree, Menomini), *Chimakuan* (Quileute), *Wakashan* (Kwakiutl), *Bella Coola*, *Coast Salish* (Squamish), *Interior Salish* (Shuswap, Kalispel), *Keresan* (Acoma), *Yuchi*, *Siouan* (Hidatsa, Lakota, Dhegiha, Biloxi), *Caddoan* (Wichita, Pawnee), *Iroquoian* (Mohawk, Tuscarora, Cherokee), *Tsimshian* (Gitksan, Coast Tsimshian), *Chinookan* (Lower Chinook, Kathlamet), *Takelma*, *Coos* (Hanis Coos), *Alsea*, *Siuslawan*, *Klamath*, *Sahaptian* (Northern Sahaptin, Nez Perce), *Wintun* (Patwin, Wintu), *Maidu* (Northeast Maidu), *Yokuts* (Wikchamni, Yaudanchi Yokuts, Yawelmani), *Costanoan*, *Miwok* (Sierra Miwok), *Zuni*, *Atakapa*, *Chitimacha*, *Tunica*, *Muskogean* (Choctaw, Seminole), *Yukian* (Wappo), *Huave*, *Totonacan* (Totonac), *Mixe-Zoquean* (Mixe, Sierra Popoluca, Copainalá Zoque, Ostucacan Zoque), *Mayan* (Tzotzil, Jacaltec, Mam, Ixil, Tzutujil), *Karok*, *Chimariko*, *Shasta*, *Palaihnihan* (Achumawi, Atsugewi), *Pomo* (Eastern Pomo, Southeastern Pomo), *Washo*, *Chumash* (Barbareño Chumash), *Salinan*, *Esselen*, *Seri*, *Yuman* (Kiliwa, Diegueño, Hualapai), *Tonkawa*, *Karankawa*, *Coahuiltecan* (Coahuilteco), *Tequistlatecan* (Huamelultlec Oaxaca Chontal), *Tarascan*, *Tanoan* (Kiowa, Taos), *Numic* (Northern Paiute, Shoshoni, Comanche, Ute, Chemehuevi), *Takic* (Luiseño, Cahuilla), *Hopi*, *Pimic* (Papago, Nevome, Northern Tepehuan), *Taracahitic* (Western Tarahumara, Yaqui), *Aztecan* (Pipil, Huasteca, Nahuatl, Michoacan Nahuatl, North Pueblo Nahuatl, Tetelcingo Nahuatl), *Coric* (Cora, Huichol), *Otomian* (Pame, Otomi), *Mixtecan* (Jicaltepec Mixtec, Ocotepec Mixtec, Peñoles Mixtec), *Popolocan* (Mazatec), *Chinantecan* (Comaltepec Chinantec, Palantla Chinantec), *Zapotecan* (Chatino, Zapotec), *Tlapaneca*.

SOUTH AMERICA:
Yanomam (Sanuma, Central Waica), *Misumalpan* (Miskito), *Talamanca* (Bribri), *Guaymi*, *Itonama*, *Warao*, *Mura* (Piraha$), *Barbacoan* (Cayapa), *Cahuapanan* (Jebero), *Zaparoan* (Zaparo, Iquito, Huao), *Quechua* (Imbabura Quechua), *Aymara*, *Jaqaru*, *Mapudungu* (Araucana), *Patagonian* (Gununa Kune), *Qawesqar* (Kawesqar), *Iranxe*, *Movima*, *Ticuna*, *Nambikuaran* (Nambikuara), Puinave (Hupda), *Tucanoan* (Tucano, Carapana, Southern Barasano, Yebamasa, Siona), *Cayuvava*, *Trumai*, *Salivan* (Saliva), *Candoshi*, *Jivaro* (Achuar, Jibaro), *Cariri*, *Tupi-Guaraní* (Guajajara, Urubu-Kaapor, Guaraní, Wayapi, Cocama, Munduruku, Siriono), *Guahiban* (Cuiva), *Chipayan* (Uru), *Maipurean* (Ipurina, Piro, Axininca, Campa, Machiguenga, Baure, Ignaciano, Island Carib, Resigaro, Goajiro), *Andoke*, *Peba-Yaguan* (Yagua), *Boran* (Muinane), *Witotoan* (Murui, Witoto), *Carib* (Carib, Apalai, Makuchi, Hixkaryana), *Mascoian* (Lengua), *Moseten* (Moseteno), *Guaicuruan* (Abipon), *Mataco*, *Panoan* (Chacobo, Amahuaca, Sharanahua, Cashibo, Shipibo-Conibo), *Tacanan* (Eseejja, Araona, Tacana), *Rikbaktsa*, *Iate*, *Bororoan* (Bororo), *Chiquito*, *Ge-Kaingang* (Kaingang, Xavante, Apinaye, Cayapo, Kraho, Canela-Kraho).

→ to study whether linguistic features are a global or only a regional phenomenon

example of Dryer (1992: 83):

	Africa	Eurasia	SE-Asia & Oceania	Australia & New Guinea	N-America	S-America	Average
OV & Po	**15**	**26**	**5**	**17**	**25**	**19**	**107**
OV & Pr	3	3	0	1	0	0	7
VO & Po	4	1	0	0	3	4	12
VO & Pr	**16**	**8**	**15**	**6**	**20**	**5**	**70**

What are the parameters and their values in Dryer‘s example? How do they correlate and is this a correlation a global or a regional phenomenon?

criteria for not including a language in the sample:

- languages in which OV and VO occur likewise
- languages without data on the topic
- languages in which the investigated categories (e.g., adpositions) do not exist

Data sources:

- secondary/tertiary data: **reference grammars**, dictionaries & descriptive articles on the topic
  → most common data source; different representations of grammars are to be standardised

- elicited primary data: **questionnaires**
  → access to native speakers is mandatory; time-consuming

- tertiary data: **typological databases** - World Atlas of Language Structures (http://wals.info)
  - Glottobank (https://glottobank.org)
  → viable for typological issues which have not already been analysed within the database

- compiled primary data: **multi-aligned parallel text corpora**
  → only very few and special kinds of parallel texts available
  (e.g., the Bible, user manuals, brochures of global organisations).

Research outcomes:

- **universals** (quantitative statements on certainties or probabilities):

  - *absolute universals* (exceptionless statements) vs.
    statistical universals (probability statements)

  - *unrestricted/unconditional universals* (one parameter) vs.
    implicational universals (unidirectional correlations between two logically independent parameters)

  → Universals Archive (https://typo.uni-konstanz.de/rara)

- **typologies** (scope of variation: types of languages regarding linguistic parameters)
- **typological maps** (geographic distribution of typological features)

Universals:

- In most languages (95.77%), the subject precedes the object (SOV, SVO, VSO). Only in 4.23% of the languages, the object precedes the subject (VOS & OVS; OSV does not occur at all) (Tomlin 1986: 22).
- All languages have vowels and consonants.
- If the genitive follows the noun, then the relative clause follows the noun: NG ' NRel (Hawkins 1983: 83).
- No language allows more than four arguments per verb (Pesetsky 1995).
- With overwhelmingly more than chance frequency, languages with dominant order VSO have the adjective after the noun (Greenberg 1963: 85).
- In all languages in which adjectives are inflected, nouns are inflected as well (Moravcsik 1993).

What kind of universals are these? Take the exercise in Section 4.3 regarding their parameters and values into consideration. see page 125 for solutions

Typological findings are motivated by:

- **common experiences & shared conditions** of mankind

  e.g., number of fingers → numerals (decimal or quintal systems);
  gravity → spatial reference;
  physiology of the human vocal tract → sound inventory

- **social, political & environmental factors**

  e.g., culture-specific ideas of social structure → kinship terminologies, social deixis, person names, etc.;
  political power in contact situations → language death, structural transfer, etc.;

- **interacting and competing motivations:**

  - **discourse**: discourse strategies (topic-focus, etc.) → definite/indefinite marking etc.;
  - **processing**: efficient processing → recursion of relative clauses is limited etc.;
  - **economy**: minimal speech effort → frequent forms are/become shorter etc.;
  - **iconicity**: mirroring an experienced item property → alienable/inalienable constructions, etc.

Corpus linguistics

Chapter 5

Chapter 5: Outline

Definitions 5.0

Research aims and questions 5.1

Corpus-linguistic approaches 5.2

Methodology
Corpus types, basic aspects of corpus research 5.3

Basic research findings 5.4

Definitions

5.0

[C]orpus linguistics [...] can be seen as a *pre-application methodology*. [...] [U]nlike other applications that start by accepting certain facts as *given*, corpus linguistics is in a position to define its own sets of rules and pieces of knowledge before they are applied [...]. Corpus linguistics has, therefore, a theoretical status and because of this it is in a position to contribute specifically to other applications.
(Tognini-Bonelli 2001: 1)

However, if the types of linguistic analyses that corpus linguists conduct are examined, it becomes quite evident that corpus linguistics is more a way of doing linguistics, "a methodological basis for pursuing linguistic research" (Leech 1992: 105), than a separate paradigm within linguistics.
(Meyer 2002: xi)

What is corpus linguistics? It is certainly quite distinct from most other topics you might study in linguistics, as it is not directly about the study of any particular aspect of language. Rather, it is an area which focuses upon a set of procedures, or methods, for studying language (although, as we will see, at least one major school of corpus linguists does not agree with the characterisation of corpus linguistics as a methodology). (McEnery & Hardie 2012: 1)

Corpus linguistics is empirical. Its object is real language data. The discourse is the totality of all the texts that have been produced within a discourse community. [...] Corpus linguistics makes general and specific claims about the discourse, based on the analysis of a suitably selected cross-section of it, i.e. the corpus. General claims have to do with rules or with probabilistic expectations. They fall within the field of grammar or variation or language change, and also into the field of lexical meaning [...]. (Teubert 2005: 3-4)

- **research aim**:
  - identify how speakers use their language(s) in natural contexts (i.e., uninfluenced by the researcher) by means of analysing authentic natural language data provided in corpora

- **research questions**:
  - What are the quantitative estimates (e.g., frequency of occurrence) associated with qualitative differences between patterns of language use?
  - What discourse contexts favour one pattern of language use over another?
  - Which linguistic features show variation across speakers, language varieties/text types/genres/registers or corpora?
  - Which language varieties/text types/genres/registers show variation regarding the distribution or frequency of particular linguistic features?

see page 135 (more detailed research questions)

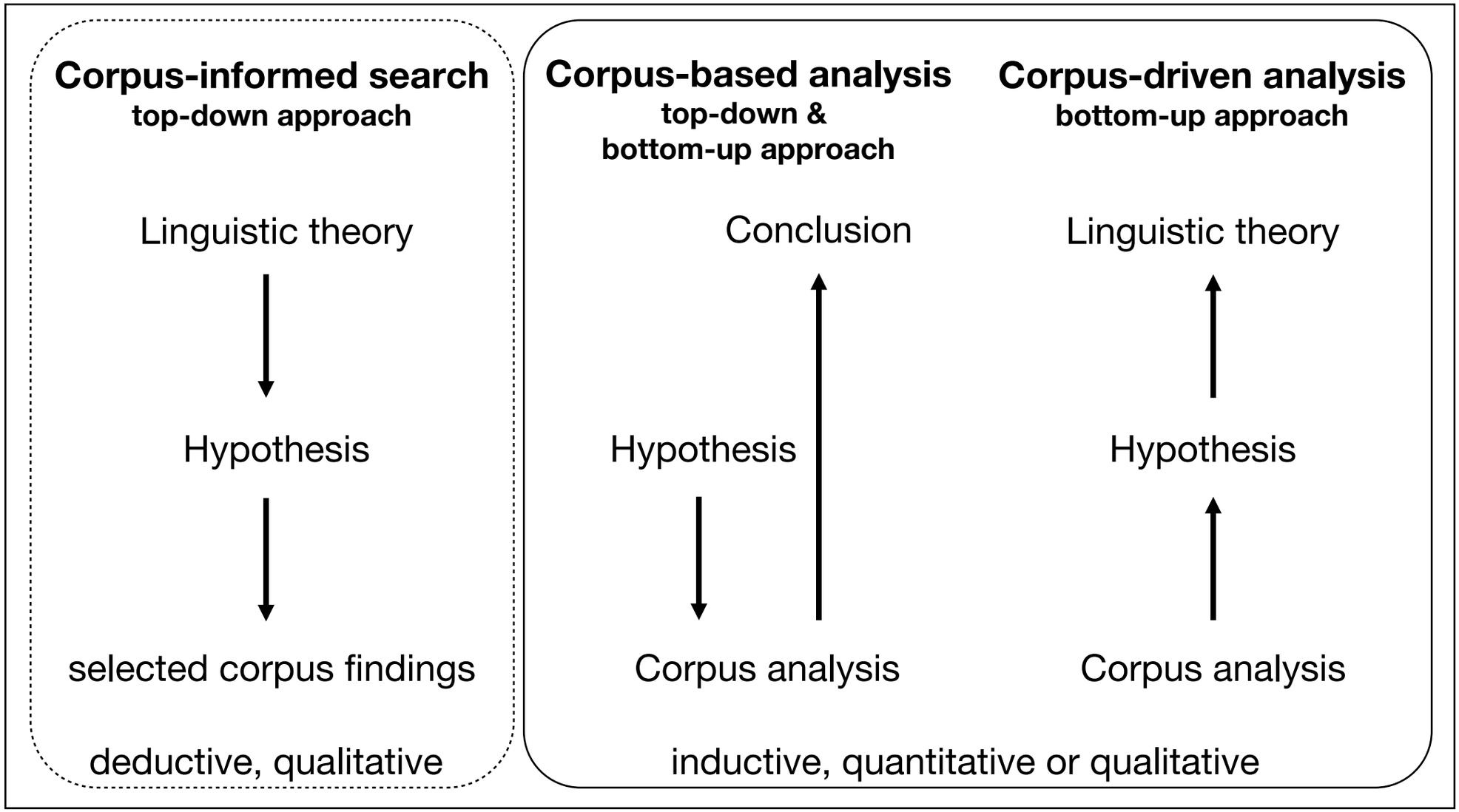


Figure 19

corpus-informed search: does not conform to analytical standards if performed with selective analysis & findings (e.g., search only for confirmative corpus examples)

corpus-based analysis: majority of research works in various fields of application

corpus-driven analysis: nowadays especially in language documentation and descriptive linguistics (cf. Section 3)

We follow proposals defining corpora along a cline of **corpus prototypicality**, i.e. what features corpora typically, but not necessarily, include (Gilquin & Gries 2009: 6; Lemnitzer & Zinsmeister 2015[3]: 13; cf. McEnery & Hardie 2012: chapter 1):

A linguistic corpus is a systematic collection of authentic texts that:

- is digitised and machine-readable (Lemnitzer & Zinsmeister 2015[3]: 13)
- is compiled to be representative and balanced as regards "a particular language variety/register/genre" (i.e., the representation of all instances of the variety, register, or genre and in a proportion that mirrors the proportion in the population) (Guilquin & Gries 2009: 6)
- may contain metadata, text mark-up and linguistic annotation (Lemnitzer & Zinsmeister 2015[3]: 13, McEnery & Hardie 2012)

Common types of corpora:

- monolingual written/spoken/signed or multimodal corpus of ‘native speakers’
- computer-mediated communication (CMC) corpus
- learner corpus
- parallel corpus
- comparable corpus
- monitor corpus
- sample corpus
- web as corpus
- treebank
- general/reference corpus
- specialised/opportunistic corpus

Read  pages 141-145. Which features define the above types of corpora? What research questions can be investigated with these corpora?

3 basic steps in corpus research (Rayson 2015):

- corpus compilation
- corpus annotation
- corpus retrieval and analysis

Depending on the research question, all or only some of them are performed in a research project.

Corpus compilation:

- data collection based on corpus design
- including preparation of data, digitisation, etc.

Corpus design:

- operationalises the research question
- specifies the function of the corpus, data selection, annotation guidelines
- specifies the sampling procedure, especially regarding representativeness, balance, and corpus size

Corpus annotation:

- can be performed at various linguistic levels (annotated data) or not at all (raw data)

- is based on 2 interconnected steps:

  – **tokenisation**: segmentation of the raw data into *tokens* as the smallest units of analysis

  – **tagging**: linking of tokens to linguistic categories (e.g., words, morphemes, semantic categories, speech acts, etc.) by means of a tag set

    → Part-of-Speech (POS) tagging for morphosyntactic categories: most commonly (and often automatically) applied to annotated corpora

    → Lemmatisation: tagged word forms/tokens associated with a superordinate base form (*lemma*), e.g. $finger_{Sg}$, $fingers_{Pl}$ -> $FINGER_{lemma}$

- is the qualitative aspect of any corpus analysis, even if the research question is quantitatively oriented (Lüdeling 2007, 2017)

1. Often, words are chosen as tokens during annotation. Discuss why this can be problematic, focusing on different linguistic definitions of a word (e.g., phonological word, orthographic word) and/or languages with different concepts of what a word is.

2. Discuss why semantic and pragmatic categories may be harder to annotate than grammatical/morphosyntactic categories. Develop an example tag set per linguistic level to illustrate your argument.

3. Many learner corpora consist of raw data without further linguistic annotation. Discuss the problems that may arise when automatic software for POS tagging or lemmatisation, which have usually been trained on the basis of native speaker material, is applied to annotate learner corpora.

Corpus retrieval (search):

- in accordance with the research question
- exhaustive to avoid false positive or false negative findings (i.e., search performed for the entire corpus or randomly chosen subsets)

Corpus analysis:

- quantitative distribution and qualitative function
  - Quantitative estimates inform qualitative analyses about preferences of contexts for linguistic features (and vice versa).
  - Inferential statistics determine significant, systematic patterns that can be generalised to the population.
- microscopic (variationist) vs. macroscopic (text-linguistic) approach (Biber 1988):
  - Units of analysis are existing variants of a linguistic feature vs. different text types/registers.
- variation in space or time:
  - synchronic variation at a single point in time (e.g., regional or sociolectal variation) vs. diachronic variation over time (temporal variation within a variety)

Concordancing:

- KWIC concordance line
- Concgram

Frequency estimates:

- absolute frequency estimates (raw or normalised) by types or tokens
- relative or comparative frequency estimates: type-token ratio, keyword list

Collocation analysis:

- collocation
- colligation
- semantic preference
- semantic prosody
- n-gram

pages 147-153

1. What is the relationship between concordance and collocation analysis?

2. The type-token ratio (TTR) is particularly suited for research on language acquisition or language complexity. Find two texts in different modalities (spoken, written, or signed) for (i) a child or (ii) an adult foreign language learner and compare the TTR values for the modalities. Discuss the implications of your results. What kind of additional analyses may also be useful?

3. Compare the frequency and distribution of the concepts love, hate and death in different music genres (e.g., pop music, heavy metal, opera), using keyword lists or collocation analysis.

Further types of analysis are often tied to specific research questions or fields of application:

Type	Field of application
Collostructional analysis: – interplay of lexeme-based and construction-based meanings, i.e. the preference of certain constructions to combine with particular lexemes and vice versa	e.g., empirical investigation of Construction Grammar, co-occurrence of morphosyntactic patterns and their semantic motivation
Corpus-assisted discourse analysis: – multi-method/mixed-methods approach using qualitative text interpretation and quantitative corpus estimates	e.g., investigation of the dominant or group-specific meanings in socio-political discourse
Multi-dimensional (MD) analysis: – multivariate statistical analysis of multiple linguistic features and their significance to define a certain language variety, genre, or register	e.g., register-specific variation in diachronic or cross-linguistic studies

Web-as-corpus approach: queries via search engines or metasearch tools to gather linguistic data

Web-for-corpus approach: systematic collection of texts from the internet to compile a corpus following criteria set by the researcher

Advantages and disadvantages of using the web for linguistic research (Gatto 2014):

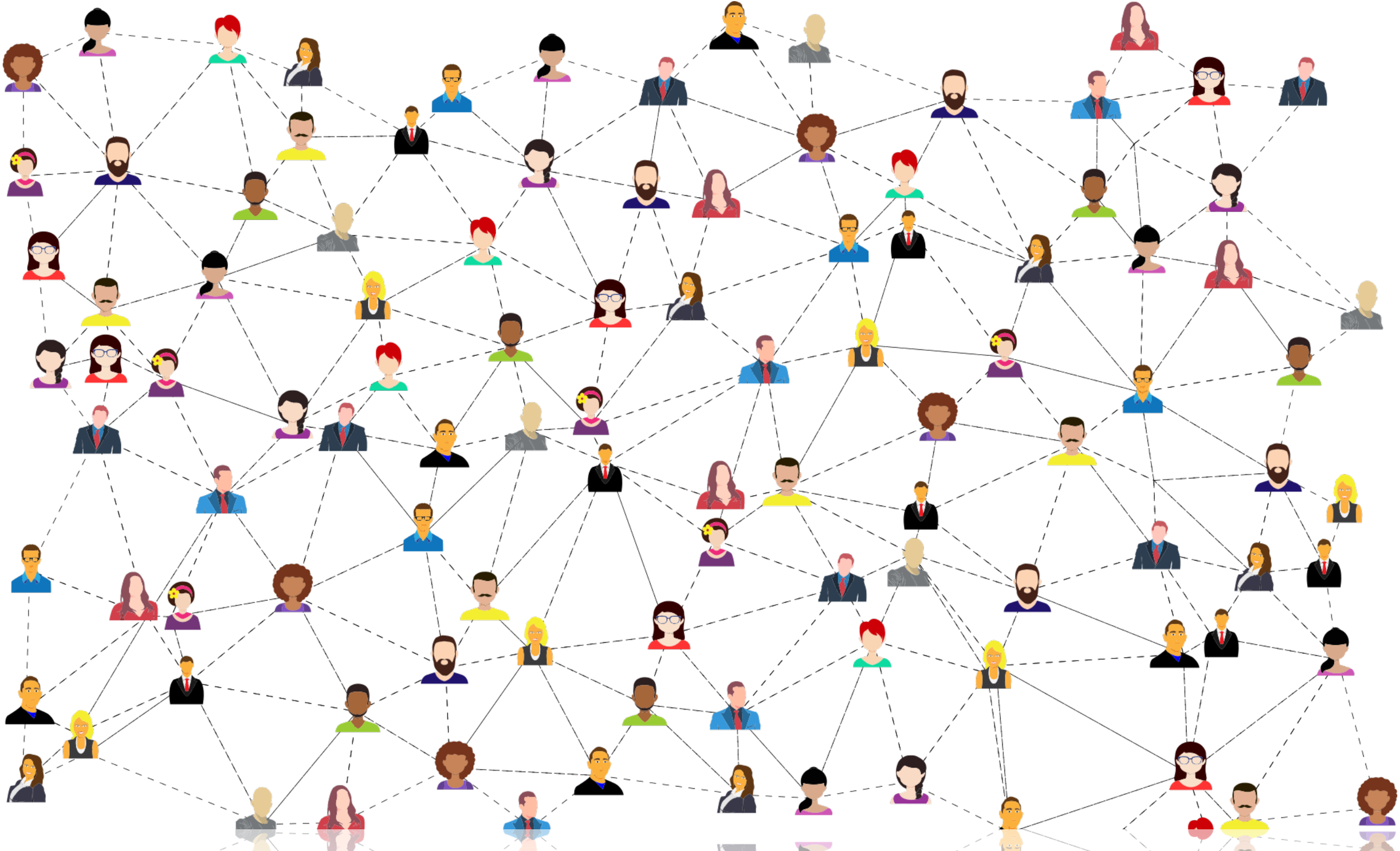
- mega corpus of infinite size
- continuous updating, reflecting language change as it happens
- free and easy access

- enormous size: problematic definition of what large or small effect sizes mean
- noisy data, medium-specific communication and lack of metadata: problematic definition of representativeness and balance
- dominant vs. underrepresented languages/ language varieties/ registers on the internet

- **quality of the analysis outcome** depends on:
  - matching the research question with the corpus design
  - the quality of corpus retrieval and (statistical) analysis
- **absence of positive evidence** for a pattern does not necessarily mean that a given pattern is not part of a language's grammar (cf. Lemnitzer & Zinsmeister 2015[3]: 51-54).
  - Cross-methodical comparison with experimental or elicitation-based methods is often beneficial in studies on mental grammar.
- **language-as-product view**: language use in real-time (e.g., hesitations, misspellings and corrections) may not be accessible to analysis in every corpus due to editing or lack of annotation
  - cognitive investigations with a focus on mental/linguistic representations ('what?') rather than temporal aspects of cognitive processing ('when?')
- **reproducibility and replicability** of research findings
  - careful documentation of corpus retrieval and analysis choices
  - comparability of different specialised corpora compiled for similar research questions

Corpus-linguistic methods are widely used across linguistic subdisciplines, with corpora representing a major source in empirical linguistic research.

Research finding	Field of application
usage-based grammars of major languages or documented languages	descriptive linguistics (cf. Section 3)
lexicography of major languages or documented languages	descriptive linguistics (cf. Section 3)
diachronic language change	historical linguistics
synchronic language-internal variation: variety-specific or speaker-specific variation	sociolinguistics (cf. Section 6)
synchronic cross-linguistic variation: analysis of parallel texts	comparative or typological linguistics (cf. Section 4)
discourse and pragmatics	discourse studies
language and cognition	cognitive and psycholinguistics (cf. Section 7)
language learning and teaching	applied linguistics, language pedagogy
natural language processing (automatic text recognition)	computational linguistics



Sociolinguistics and anthropological linguistics

Chapter 6

Chapter 6: Outline

Definitions 6.0

Research aims and questions 6.1

The sociolinguistic and the anthropological approach 6.2

Methodology
Ethnographic fieldwork, linguistic & extra-linguistic data 6.3

Basic research findings 6.4

Definitions

6.0

Anthropological linguistics is that sub-field of linguistics which is concerned with the place of language in its wider social and cultural context, its role in forging and sustaining cultural practices and social structures. [...] Anthropological linguistics views language through the prism of the core anthropological concept, culture, and, as such, seeks to uncover the meaning behind the use, misuse or non-use of language, its different forms, registers and styles. [...] **Sociolinguistics**, on the other hand, views language as a social institution, one of those institutions within which individuals and groups carry out social interaction. It seeks to discover how linguistic behavior patterns with respect to social groupings and correlates differences in linguistic behavior with the variables defining social groups, such as age, sex, class, race, etc.

(Foley 1997: 3)

[...] redefining the study of language and culture as one of the major subfields of anthropology [...] **linguistic anthropology** will be presented as the study of language as a cultural resource and speaking as a cultural practice.

(Duranti 1997: 2-3)

Sociolinguists study the relationship between language and society. They are interested in explaining why we speak differently in different social contexts, and they are concerned with identifying the social functions of language and the ways it is used to convey social meaning.

(Holmes 2013[4]: 1)

Sociolinguistics:

- **research aim**: study of language as social institution
- **research questions**:
  - What kind of language-internal variation does language of different social groups and/or in different social contexts exhibit?
  - What is the social function of language and which meaning is expressed by language in social interaction?
- **secondary or applied goals**: revealing social inequality/abuse of power in/through language (critical discourse analysis), identification of speakers via language material (forensic sociolinguistics)

Anthropological linguistics:

- **research aim**: study of language in its cultural context
- **research questions**:
  - Are there culture-specific ‘ways of speaking’ or conversational practices?
  - To what extent do patterns of cross-linguistic variation reflect cultural ideas? And what cultural meaning is encoded in linguistic forms?
- **secondary or applied goals**: reclamation/revitalisation of endangered languages and the cultural ideas and practices associated with language

More specific research questions within various subfields:

- **variationist sociolinguistics**: Does the use of linguistic variants correlate systematically with social features of the speaker? And which social parameters have an impact on language variation (e.g., gender, age, etc.)?
- **sociolinguistic gender studies**: How does the speech of women and men differ? How is gender constructed through language? But also, how is gender encoded in linguistic forms?
- **dialectology**: What are regionally divergent language features (i.e., distinctive characteristics of dialectal varieties) and how is such variation distributed geographically?
- **Interactional sociolinguistics**: What are conversational practices and discourse strategies which are characteristic of a particular language variety? How do people use language in different social contexts or settings? What is the social meaning encoded in speech?
- **historical sociolinguistics:** Which sociocultural parameters affect language change?
- **sociology/social psychology of language**: What attitudes towards a language or language variety do speakers have? What perception of a language or language variety do they have? How do these beliefs and ideologies shape (linguistic) behaviour? And, how is identity linked to language, i.e., how do sociocultural communities define themselves by language?

More specific research questions within various subfields:

- **anthropological linguistics** (closely linked to cognitive/cultural linguistics): To what extent do patterns of cross-linguistic variation reflect cultural ideas and practices? And, what is the cultural meaning encoded in linguistic forms (lexical items, semantic parameters, and formal categorisation)?
- **linguistic anthropology** (similar to interactional sociolinguistics): What are culture-specific conversational practices or 'ways of speaking'? What are emic genres? And what is the cultural meaning encoded in these linguistic practices?
- **research on language & culture contact**: Which linguistic phenomena occur in multilingual settings? What determines language choice in code-switching? How do languages change (transfer), emerge (pidgins & creoles), or vanish (language death)?
- **research on language socialisation & acquisition**: In what way does language socialisation differ across various cultural settings? And how does first language acquisition vary cross-culturally?

see page 169 for more detailed research questions

Sociolinguistic and anthropological-linguistic approach

6.2

Sociolinguistics & anthropological linguistics:

two **interdisciplinary subdisciplines** which study **language in its socio-cultural surrounding** with slightly different perspectives

Sociolinguistic approach	Anthropological-linguistic approach
main disciplines: **linguistics & sociology**	main disciplines: **linguistics & cultural anthropology**
study of **language-internal variation & social parameters** (e.g., age, gender, social class, geographic area, etc.)	study of **cross-linguistic variation & cultural aspects**
basic method: **correlation analysis** (independent variable: social parameter; dependent variable: linguistic features) → predominantly **quantitative approach**	basic method: **participant observation** → predominantly **qualitative approach**
traditional focus: **Western societies**	traditional focus: **non-Western contexts**
linguistic fieldwork (for studying language in the natural environment of its speakers)	

Ethnographic fieldwork (as compared to documentary & descriptive fieldwork):

- provides access not only to language data but also to the **socio-cultural context in which language is embedded**
- **stronger engagement/involvement** with the research participants

> The anthropologist must relinquish his comfortable position in the long chair on the veranda of the missionary compound, Government station, or planter's bungalow, where, armed with pencil and notebook and at times with a whisky and soda, he has been accustomed to collect statements from informants, write down stories, and fill out sheets of paper with savage texts.
> (Malinowski 1954: 146–147)

Participant observation:

- the researcher takes on the **role of participant** in the setting of observation.
- participation: more than mere presence but **getting involved in community interactions** and learning to become a community member without taking on an influential position
- to gain an **emic view** (inside perspective) of the research participants, their ideas and practices
- a **balancing act** to maintain an observer's 'neutral' distanced perspective while gaining an emic involved perspective

Discuss to what extent you, as an adult, can conduct participant observation in a kindergarten group. Or which difficulties a female researcher may face studying men's activities. Or what about getting access to secret societies, such as initiation schools in South Central Africa.

Stages of the researcher-community relationship in ethnographic fieldwork:

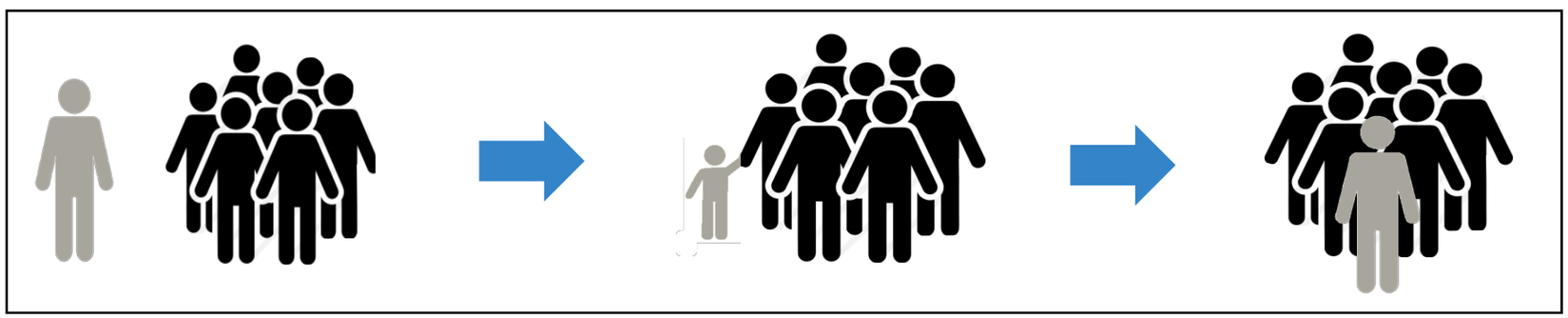


Figure 20

You may gather some practical experience in ethnographic fieldwork and participant observation in a ‘sub-culture’ of your own society with which you are unfamiliar, e.g., an allotment garden community, the banking industry, or international community groups.

During this fieldwork experience, write a field diary and focus on the following questions and how the answers change in the course of time:

- How (un)comfortable do you feel? To what missing/acquired knowledge is the feeling of (un)comfortability related?
- How are you treated by the community members? To which subgroups do/don’t you get access?
- What is your observation/perception of the community? Which aspects do/don’t you realise most/not any more?

Data types

language data (aspects of language-internal or cross-linguistic variation)

more research effort

primary data	*secondary data*	*tertiary data*
- speech recording - elicitation - experimental tasks	- ready-made corpora - grammars	- dialectal atlases, etc. - typological databases

less research effort

less data triggering

natural data	*semi-natural data*	*non-natural data*
- participant observation & recording	- sociolinguistic interviews	- picture-prompted story telling - systematic reading tasks - matched-guise task

more data triggering

extra-linguistic data (socio-cultural speaker characteristics and context parameters)

rather qualitative research

emic parameters & values (for an internal understanding)	*etic parameters & values* (for comparative purposes)
- participant observation - open interviews - cultural consensus analysis - network analysis	- systematic inquiry - use of predetermined parameters (e.g., age) - use of a comparative metalanguage

rather quantitative research

Figure 21

Research participants: investigated (sub-)group of speakers

- **speech community**: a socio-culturally significant or meaningful group of speakers with a common language foundation (i.e., people with shared linguistic and/or social parameters) which can

  - either be identified in the course of the research project, or
  - one can start from a defined group of speakers.

  → examples: *'social networks', 'communities of practice'*, etc.

- single **main informants** (specialists) vs. **a representative sample** of speakers (members of the speech community)

  - kinds of samples: - *quota sampling* according to social criteria, such as age or gender,
    - *snowball sampling* based on acquaintance, friendship, kinship, or any other kind of relationship (including shared activities)

Take specific research questions (e.g.,  page 169) and determine what kind of data and which research participants are needed to answer the respective questions.

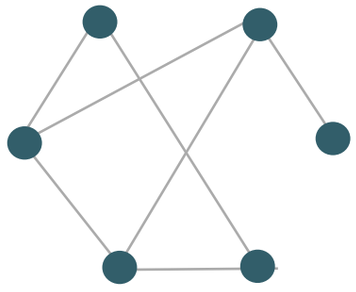
Selected types of analysis:

- **quantitative correlation analysis**: measuring the use of linguistic variants as conditioned by social parameters of the speakers (e.g., Labov 1972)
  → variable rules analysis (Tagliamonte 2006)
  - detection of linguistic variants used in the same context (independent variable)
  - statistical correlation analysis (generally corpus-based; dependent variable: social speaker characteristics → quota sampling)
  - identification of correlation patterns

- **qualitative conversation and discourse analysis**: measuring contextual and interactional patterns and their meaning in corpora of natural speech (e.g., Sidnell & Stivers 2012; Tannen, Hamilton & Schiffrin 2015)
  - compilation of natural language data (primarily oral interaction → case studies)
  - microscopic analysis of conversation components and phenomena (primarily explorative, sequence-analytical and reference-oriented)
  - identification of recurring patterns

Selected types of analysis:

- **quantitative network analysis**: measuring language change/linguistic innovation as conditioned by network density/strength of ties (e.g., Milroy & Milroy 1985)
  - detection of a network (i.e., a set of persons with linkages among each others → snowball sampling)
  - analysis of the characteristics of the linkages and their impact on the social behaviour of the persons (representation: persons as nodes, relationships as connecting lines, etc.)
  - identification of network structures
    (network density: actual connections in relation to the totality of possible connections;
    strength of ties in terms of temporal and emotional intensity, symmetry, and number of shared spheres)



Statement levels of research findings:

- **micro-level**: statements about single speakers or small-scale units
  (e.g., case studies of a particular group of speakers)
- **macro-level**: statements about large-scale units
  (e.g., a speech community)
- **a more general level**: cross-linguistic generalisations regarding language and society/culture in general
  (e.g., universal patterns)

  → Broadly generalising outcomes are relatively rare due to the great effort required to compare different languages in their diverse socio-cultural contexts.

Research outcomes:

- language-internal variation and geographic parameters:
  - **dialects**: vernaculars with regionally divergent linguistic features
  - **linguistic maps/atlases**: geographic distribution of dialects

- language-internal variation and social groups:
  - **sociolects**: vernaculars associated with a particular social group
  - **social markers** (e.g., gender markers): linguistic features statistically representative of social groups

- language-internal & cross-linguistic variation and communicative contexts:
  - **registers**: vernaculars associated with a particular social situation or purpose
  - **ethnographies of speaking/communication**: descriptions of communicative practices as characteristic of a particular group or a particular situation/setting

Research outcomes:

- cross-linguistic variation and cultural meaning as encoded in linguistic forms:
  - **cultural conceptualisations**: culture-specific ideas encoded in linguistic forms
- socio-cultural factors of language change:
  - **network density/strength of ties** and language-internal change
  - **language/culture contact** and language change (e.g., language death, emergence of pidgins/creoles, and transfer of linguistic features)

    → prestige & language change: **basilects** (languages/language varieties of low prestige) vs. **acrolects** (languages/language varieties of high prestige)

- cross-linguistic variation in language socialisation:
  - **acquisition practices** of language and cultural knowledge

Cognitive linguistics and psycholinguistics

Chapter 7

Chapter 7: Outline

Definitions

7.0

Cognitive linguistics is a modern school of linguistic thought and practice. It is concerned with investigating the relationship between human language, the mind and socio-physical experience. (Evans, Bergen & Zinken 2007: 263)

Cognitive linguistics is a theoretical and empirical programme that goes beyond the visible structure of language to investigate the complex background operations of cognition that create grammar, conceptualization, discourse, and thought itself. (Fauconnier 2006: 1)

Cognitive linguistics [emerged] as a reaction against generativism and extreme modularism. [...] It is not, strictly speaking, one single theory, but rather a group of theories that share a number of basic theoretical principles [...]. (Barcelona & Valenzuela 2006: 17, footnote 1)

Psycholinguistics is the empirical and theoretical study of the mental faculty that underpins our consummate linguistic agility. (Altmann 2001: 129)

Psycholinguistics is an interdisciplinary field of study in which the goals are to understand how people acquire language, how people use language to speak and understand one another and how language is represented and processed in the brain. Psycholinguistics is primarily a sub-discipline of psychology and linguistics, but it is also related to developmental psychology, cognitive psychology, neurolinguistics and speech science. (Fernández & Smith Cairns 2010: 1)

Cognitive linguistics

- **research aim**: explanation of linguistic and cognitive conceptualisations (the mind), and of the relationship between language, thought and culture
- **research question**: What conceptual organisation underlies the use of a certain linguistic form or practice? Is this conceptual organisation universal or influenced by culture-specific ideas?

Psycholinguistics:

- **research aim**: explanation of the mental representations and processes used in real time by speakers to comprehend, produce or acquire language across the life-span
- **research question**: What linguistic representations and processes are active during language comprehension, production or acquisition? When are they accessed or activated (especially in real-time)? And how does their influence change across the life-span?

see page 199 (more detailed research questions)

Two schools of thought characterise cognitive approaches in linguistics.

Generative approach	Cognitive approach
innate cognitive principles of language acquisition (**Universal Grammar**)	language faculty as part of or derived from **general cognitive capacities**
(core) mental representations and processes for language are strictly **language-internal**	mental representations and processes for language **can be both** particular to language and general to human cognition
deductive, top-down approach to falsify theory-specific hypotheses from **linguistic theory**	**top-down approach** to falsify theory-specific hypotheses from **cognitive theories**
search for universals in **formal linguistics**	search for mental representations and processes underlying **typological universals and linguistic diversity**
adopted in psycholinguistics (especially in studies on language comprehension and acquisition)	adopted in cognitive linguistics and psycholinguistics (in all fields of study)

Distinct participant groups to be compared in a study are selected based on:

- age
- developmental stage (e.g., children vs. adults)
- cognitive abilities (e.g., healthy vs. impaired)
- language (e.g., monolingual vs. bi-/multilingual)
- species (e.g., human vs. great apes)
- typologically distinct languages

Sampling biases regarding selection criteria and sample size:

- **age and education bias** towards local university student population (convenience sample, especially in psycholinguistics)

  → 'WEIRD' people (Henrich, Heine & Norenzayan 2010)

- **language bias** towards major languages with research infrastructure for complex experiments (especially in psycho- & neurolinguistics)

- **small sample sizes** regarding special populations due to small number of speakers or difficulty to guarantee confidentiality or anonymity (in cognitive linguistics and psycholinguistic studies on language impairments)

Discuss to what extent cognitive or psycholinguistic findings can be used as explanatory evidence in other linguistic subdisciplines (e.g., language typology) given the sampling biases.

Data types and methods of data collection and analysis depend on the basic research question.

	Mental representations as encoded in language	Language production	Language comprehension	Language and thought
data types	(1) language data produced in natural situations (2) elicited language data (3) language-related data	(1) language data produced in natural situations (2) elicited language data	(1) language-related data (2) elicited language data	(1) a. language-independent data, and b. language data
methods for data collection	(1) observational or corpus methods (2) experiments or tertiary data sources (3) experiments	(1) observational or corpus methods (2) experiments	(1) experiments (2) experiments, sometimes supported by surveys or observational and corpus methods	(1) a. experiments (non-linguistic) or ethnographic research methods, and b. all methods or typological tertiary data

Typically, experiments measure one **dependent variable** reflecting cognitive processes or mental representations in one or more of the following three domains:

- **representational access**
  - absolute access
  - relative access/processing accuracy
- **processing speed**
  - time needed to access a mental representation
- **ontological type of processing output**
  - data collected with behavioural vs. neuroscientific/neurocognitive methods
  - language data (e.g., elicited utterances) vs. language-related data (e.g., reaction time, brain activation)

Typically about two or three **independent variables** are investigated in one experiment.

- The levels of independent variables can be implemented in:
  - experimental stimuli (linguistic or non-linguistic)
  - experimental task
  - sample groups to be compared (e.g., extralinguistic variables such as age, socio-linguistic background)

→ for more general information on item types and experimental tasks see Section 2.4

Key components of the experimental design (or paradigm) in addition to sampling:

- experimental stimuli (see Section 2.4)
- experimental task
- stimulus presentation mode
- methods for data collection

All components can be used to implement levels of an independent variable.

→ for more general information on experimental design see Section 2.4

- Tasks are performed in response to stimuli.
- Different cognitive operations can be directly activated with a task.
- application of more than one task: order of tasks matters

Task type	Subtypes
decision-related tasks	- categorisation - discrimination - judgement - evaluation - verification
memory-based tasks	- recall - recognition
elicitation-based tasks	- naming - description - completion - contextualisation - association - reading aloud
comprehension tasks	- listening - reading

see page 206 (detailed definitions and examples of task subtypes)

Parameters of stimulus presentation mode

- modality
- quality
- quantity of a stimulus' information presented at once
- number of simultaneously presented stimuli
- order of stimulus presentation
- control over onset and/or duration of stimulus presentation time

Read  pages 206-207 and make yourself familiar with the definitions of these parameters. Can you find real-life examples for these parameters?

1. Section 2 contains an exercise including the gating paradigm (cf. Section 2.4). If not already done in Section 2, this exercise can be done here to reiterate the difference between natural/naturalistic stimuli and stimuli presented in isolation.

2. Find 2-3 research questions in cognitive linguistics or psycholinguistics and discuss which task (sub-)type may be appropriate for each.

3. Ask someone to (1) read aloud the following words as quickly as possible and to (2) name the ink colour:

   blue red green yellow

   Is reading or naming equally easy for the four stimuli? What does this mean for an experiment in which stimuli include several types of information?

4. Read the instructions for conducting a non-linguistic cognitive experiment on spatial reference (see page 219-221 and references cited there). Find a fellow student and perform the experiment in class.

5. Investigate communication (e.g., event description, spatial language, or conceptualisation of motion events) by use of picture-prompted story telling (the ‘frog stories’ picture book, Mayer 1969). This can be done cross-linguistically and/or in developmental research comparing children of different ages vs. adults. You can then analyse picture descriptions with respect to syntactic structure, nominal modification, number and status of referents mentioned.

Parameters of methods for data collection

Type of data	Temporal dimension	Informativity	Methods
behavioural	offline	unidimensional	– metalinguistic judgements – reaction time – self-paced reading
	online	multidimensional	– speed-accuracy trade-off (SAT) measures – eye-tracking
neuroscientific/ neurocognitive	online – temporal resolution	multidimensional	– electroencephalography (EEG) – magnetoencephalography (MEG)
	online – spatial resolution	multidimensional	– functional magnetic resonance imaging (fMRI) – functional near-infrared spectroscopy (fNIRS)

pages 207-209

type of data: behavioural (behavioural output) vs. neuroscientific/neurocognitive (measurements of brain activity with high temporal or spatial resolution; see Section 8.3)

temporal dimension: offline (reaction after processing) vs. online (measurement as cognitive processing happens)

informativity: unidimensional (one data point per measurement) vs. multidimensional (multiple data points per measurement)

Methodology

Methods for data collection and experimental designs depend on the field of study

7.3

Data collection		Psycholinguistics		Cognitive linguistics
	Language comprehension	Language production	Language acquisition	Linguistic relativity (language & thought)
methods for data collection	– offline and online behavioural – online neurocognitive (EEG)	– observation and corpus – offline behavioural (reaction time, error analysis) – online behavioural (eye-tracking) – online neurocognitive (EEG)	– observation and corpus – offline behavioural (reaction time) – online behavioural (eye-tracking) – online neurocognitive (EEG)	– observation and corpus – offline behavioural (motion, gesture, reaction time) – ethnographic research or field experiments
experimental designs or paradigms	combinations of nearly all experimental tasks and methods for data collection	– simple naming-based designs – interference designs – structural priming – SLIP paradigm	– *preverbal infants*: habituation, head-turn preference, preferential looking – *young children (production)*: truth-value judgements, picture matching/sentence-picture matching, act-out task – *young children (comprehension)*: picture naming, story narration, semi-structured elicitation	– non-linguistic tasks – picture-prompted storytelling – naming designs – judgements

Manifold and bidirectional relationship between language and thought, although maintaining independence

- linguistic influence on thinking (conceptualisations underlying linguistic expressions)
- mental focus gleanable from choice of linguistic expressions

Kinds of effects of language on thought (e.g., in Wolff & Holmes 2011; Casasanto 2017)

- no structural correspondence
- thinking for speaking
- thinking after language

- Language processing relies on mental representations roughly corresponding to basic linguistic categories (e.g., phonemes, words, phrases, ...).
- variability in processing due to interplay of language knowledge, world knowledge and other cognitive abilities (e.g., memory)
- **processing preferences** resulting from several factors such as:
  - incrementality
  - predictability
  - revision and correction
  - frequency of occurrence and language exposure/experience

Neurolinguistics

Chapter 8

Chapter 8: Outline

Definitions 8.0

Research aims and questions 8.1

Neurolinguistic approaches 8.2

Methodology 8.3
Neurolinguistic experiments, neuroscientific measures

Basic research findings 8.4

Definitions

8.0

Neurolinguistics studies the relation of language and communication to different aspects of brain function, in other words it tries to explore how the brain understands and produces language and communication. This involves attempting to combine neurological/neurophysiological theory (how the brain is structured and how it functions) with linguistic theory (how language is structured and how it functions). (Ahlsén 2006: 3)

Neurolinguistics is a highly interdisciplinary field, with influences from psycholinguistics, psychology, aphasiology, (cognitive) neuroscience, and many more. A precise definition is elusive, but often neurolinguistics is considered to cover approximately the same range of topics as psycholinguistics, that is, all aspects of language processing, but approached from various scientific perspectives and methodologies.
(de Zubicaray & Schiller 2019: ix)

[N]eurolinguistics consists of the study of language-brain relationships. Its origins are in clinical neurology of the late nineteenth century, and it continues to be a clinically related field of observation and theory construction.
(Caplan 1987: x)

Neurolinguistics is the technical term for this field, introduced into academic usage by Harry Whitaker (1971), who founded the leading journal that bears this title. As Whitaker noted at the time, it is a key assumption of neurolinguistics that ‘a proper and adequate understanding of language depends upon correlating information from a variety of fields concerned with the structure and function of both language and brain, minimally neurology and linguistics’. Today, some thirty years later, it seems necessary to add ‘cognition’ or cognitive science to the list of minimally necessary disciplines. (Ingram 2007: 3)

Psycholinguistics:

- **research aim**:
  - identification of the relationship between brain and language comprehension, production, and acquisition
- **research questions**:
  - Which brain areas and mechanisms are involved in generating language and, given the approximately 7,000 existing languages in the world, how do they relate to linguistic variability? What are the neural universals of language in the brain in terms of temporal dynamics (‘when?’) and spatial representation (‘what and where?’)?
  - Are the brain areas and mechanisms involved in generating language behaviour specific to language or do they primarily serve other, ontogenetically older cognitive functions while finding additional application in language?
  - Are the brain mechanisms that generate language particular to the human brain or are they shared with brain mechanisms in other species with different communication systems?
- **secondary goals**:
  - explanation of within-species variability in language abilities (e.g., language impairment, language loss), identification of specifically human vs. cross-species language and/or communication abilities

see page 233 (more detailed research questions)

Neurolinguistic approaches

Neurolinguistics as a collective term for several approaches

8.2

Subfields distinguishable by their research focus, participant samples, and primary methods of data collection (Small & Hickok 2016):

Neuropsychology/Aphasiology

- language impairments in patients with acquired brain lesions or with developmental neurological diseases
- open observation, post-mortem examination, neuroimaging and behavioural experiments

Cognitive neuroscience

- language processing in healthy individuals of all ages
- neuroimaging experiments, complemented by behavioural tests

Neurobiology of language

- neurobiological foundation of language and speech in context
- humans of all ages, comparison of humans and animals, neuroimaging experiments

Methods for data collection used in neurolinguistics depend on technological advancement in neuroscience.

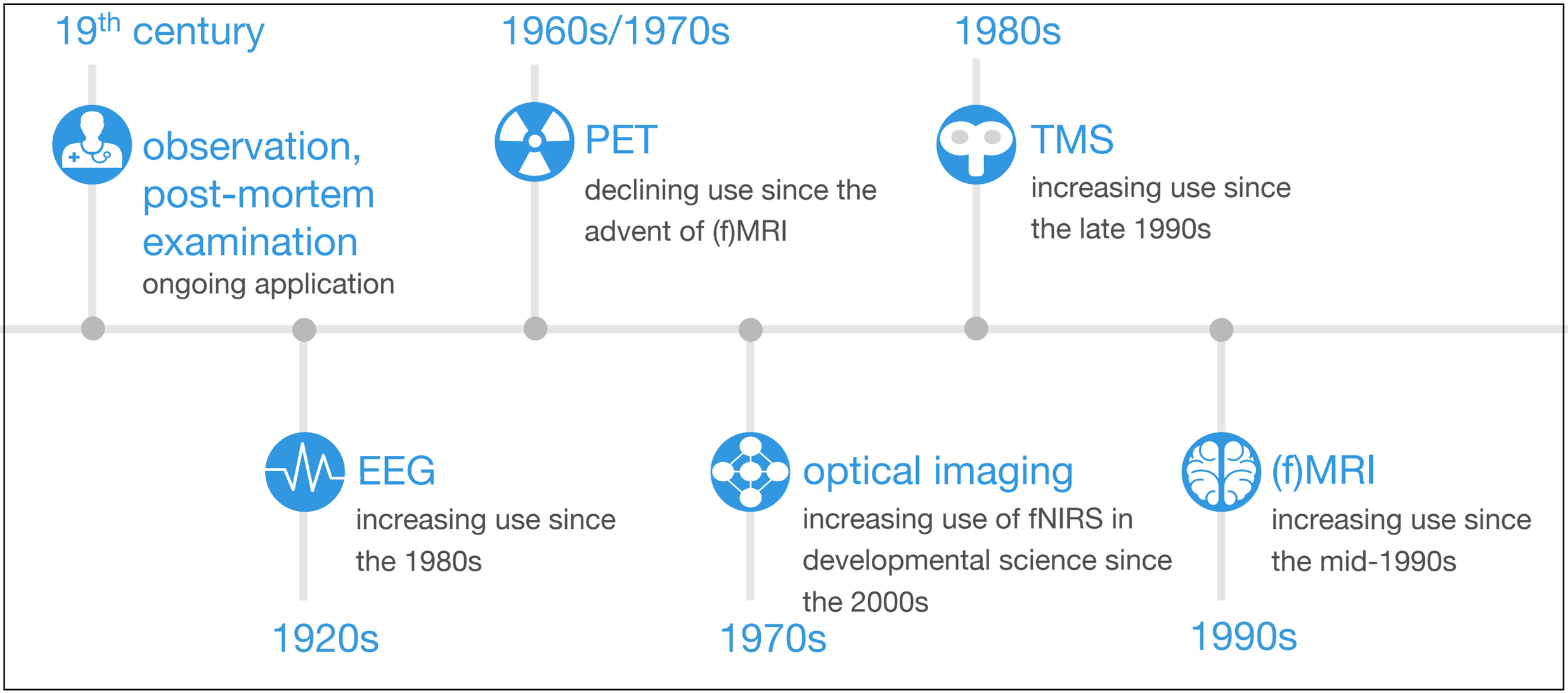


Figure 22 Dates (in blue) represent the first (significant) introduction of a method to neuroscience, subscripts roughly indicate from when a method came into use in neurolinguistics by default (cf. Gazzaniga et al. 2014[4], Horvath et al. 2010, Kemmerer 2015, Meek 2002, Mehler et al. 2008).

Dimensions to dissociate behavioural vs. neuroscientific/neurocognitive methods for data collection

temporal dimension

- neuroscientific methods: online measures of ongoing brain activity
- behavioural methods: mainly offline measures of language or cognitive processing (e.g., reaction time, elicitation), but also online measures (eye-tracking)

type of data

- neuroscientific methods: changes in brain activity in time or space (brain location)
- behavioural methods: speed and/or accuracy changes in behavioural output associated with changes in brain activity.

informativity

- neuroscientific methods: multidimensional
- behavioural methods: mainly unidimensional

pages 242-244, see also Section 7.3

Dimensions to distinguish between different types of neuroscientific/neurocognitive methods for data collection

temporal vs. spatial resolution

- temporal measures: provide high temporal resolution to capture *when* brain activity changes
- spatial measures: provide high spatial resolution to capture *where* in the brain activity changes

invasiveness

- invasive measures: application of (potentially harmful) substances to the human body or operations
- non-invasive measures: Human body and skull remain intact.

practical aspects

- high: Technical equipment is costly in acquisition and/or experiment realisation.
- low: little resources needed in equipment acquisition and/or experiment realisation

pages 242-244

Methodology

Common neuroscientific/ neurocognitive methods for data collection

8.3

Temporal dimension	Type of data	Informativity	Temporal resolution	Spatial resolution	Invasive-ness	Practical aspects	Method
online measure of brain activity	changes in brain activity	multidimen-sional data	very high	low	low	low	EEG
			very high	medium	low	high	MEG
			very high	high	high	high	ECoG
			low	high	high	high	PET
			medium-low	very high	low	high	fMRI
			medium-low	medium	low	low	fNIRS
			high	high	low	medium	TMS
offline or online	changes in brain structure		low	high	high or low	high	lesion studies: post-mortem or with (f)MRI

pages 244-251, see also Section 7.3

Find 3-5 research questions and discuss which methods for data collection may be suitable. Pay particular attention to methods with high temporal vs. high spatial resolution.

Data types

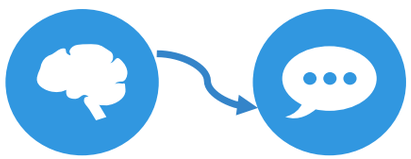
- language data: verbal utterances (e.g., picture description task with patients) or natural speech
- language-related data: neuronal or behavioural reactions in response to linguistic stimuli
- language-independent data: neuronal or behavioural reactions in response to non-linguistic stimuli

Data interpretation

- refers to the mapping of brain functions to language or cognitive functions
- functional explanation only inferred from statistically significant patterns in the data
- key assumption: **Invariance of mapping** – i.e., the relationship between function and brain structure does not change in the (healthy) brain (Rugg 1999).

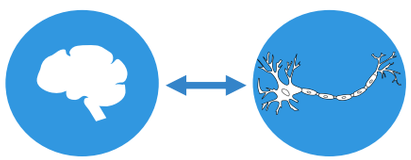
General problematic issues

(cf. Rugg 1999; Bornkessel-Schlesewsky & Schlesewsky 2009)



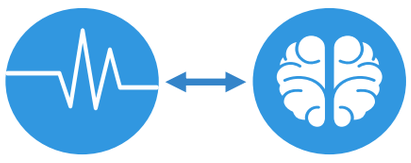
causality vs. correlation

brain *correlates* of language processing



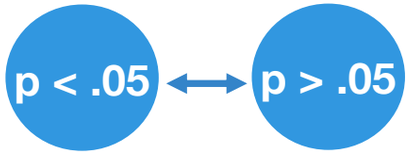
macroscopic vs. microscopic structure

units of analysis and granularity of data collection & analysis



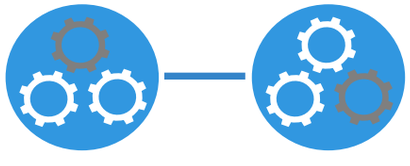
temporal vs. spatial dimension

degree of overlap between functional explanations in temporal and spatial dimensions



null effects

functional explanation constrained by statistical significance



multidimensionality and functional explanation

Multidimensional data often defy functional explanation based on only one linguistic or cognitive domain/function.

Problematic issues mostly discussed for methods with high temporal resolution (e.g., EEG, MEG).

(cf. Rugg 1999; Bornkessel-Schlesewsky & Schlesewsky 2009)

- **selectivity problem**: Not all neuronal sources engaged in a task are measured.
- **inverse problem**: Differences in brain activity across the scalp cannot be used to infer the underlying neural source.
- **mapping problem**: Are brain functions mapped to linguistic domains (and which ones, e.g., syntax, semantics) or to general cognitive domains (e.g., conflict resolution)?

1. Discuss which skills, resources and preparations are required for a neurolinguist who aims to collect data from speakers of understudied languages who live in remote areas.
2. Reflect upon the use of neurocognitive/neuroscientific methods for data collection in light of reciprocal effects (see Section 1.1), observer bias and observer's paradox (cf. Section 2.2).

Language-specific areas in the brain

- **two designated language areas** in the left hemisphere of the brain – Broca's and Wernicke's area – connected by a single fibre tract, the arcuate fasciculus

- used to describe language behaviour in patients with aphasic syndromes

- mainly based on observation and post-mortem examination

- **disputed by findings from modern neuroimaging experiments**

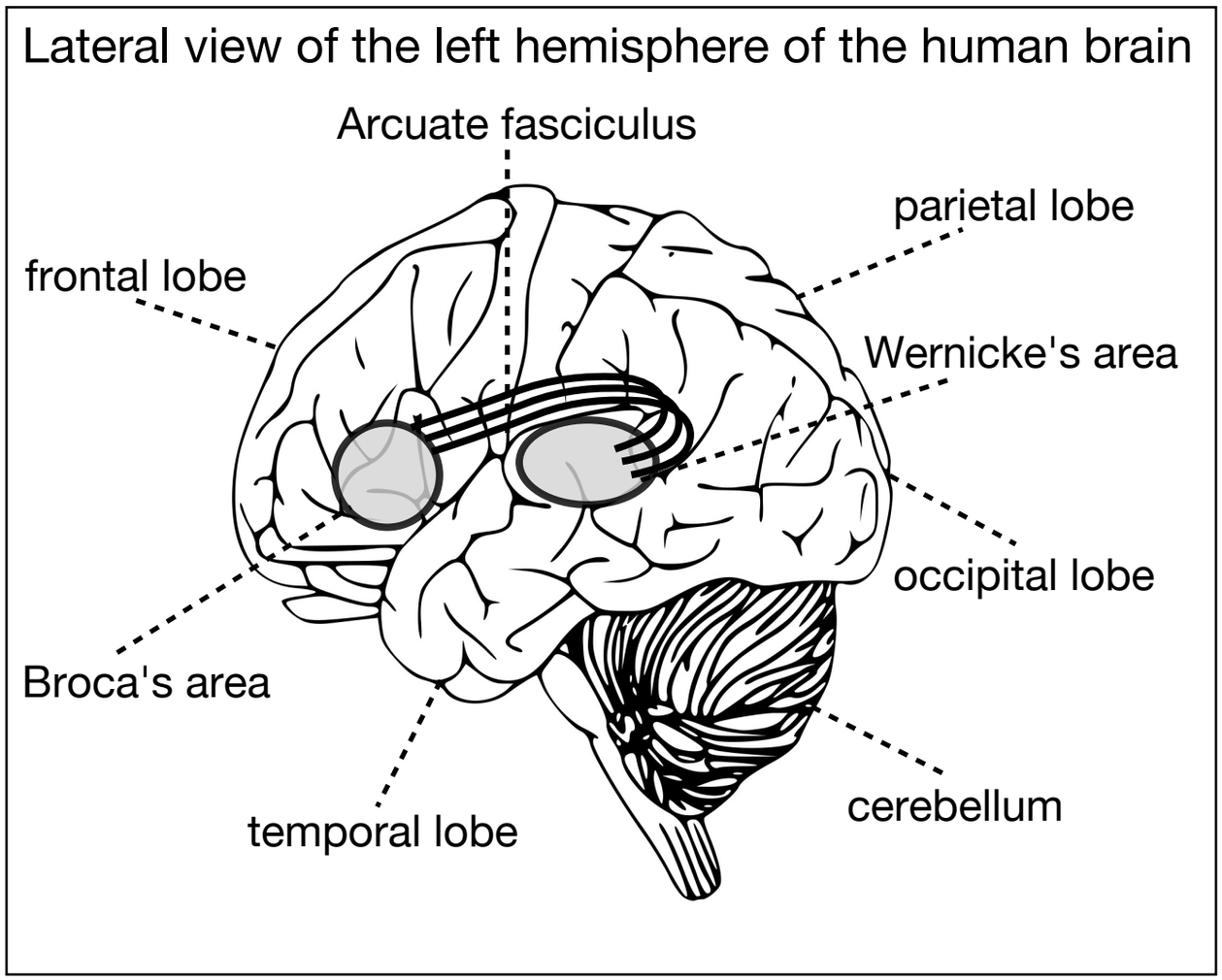


Figure 23

A network of *language-related* areas in the brain

- language processing in **several cortical and sub-cortical brain areas** that also perform other cognitive tasks, connected by several fiber tracts supporting cross-talk amongst the areas

- functional dissociation amongst areas corresponds to different linguistic domains, e.g.:
  - hemispheric division of labour in the (healthy) brain (of right-handers) for compositional and pragmatic language functions
  - semantic processing primarily in areas in the temporal lobe, phonological processing in parietal and temporal areas

- language processing is not the only or primary function of language-related brain areas

Status of animal models

- ethical and linguistic reasons for the lack of a fully fledged animal model to study language in the brain
- contribution to understanding the neural pathways for language and speech comprehension, the connection between genes and brain structure, or the evolution of language

Reductionist vs. naturalistic designs

- Current findings are based on highly reductionist experiments in the lab.
- Language must be investigated in its natural communicative situation (Andric & Small 2015).

Linguistic variability and diversity

- Individual variability is not fully accounted for in sampling procedures, except in group comparisons of impaired and unimpaired speakers.
- Linguistic diversity in the world not reflected in sampling procedures for research participants: bias towards major languages with infrastructure for lab experiments.

Discuss to what extent neurolinguistic findings can be used as explanatory evidence in other linguistic subdisciplines (e.g., language typology) given the sampling biases.

Insights from linguistic research

Chapter 9

Chapter 9: Outline

Basic research aims, methods, and findings per subdiscipline:

	Research aim/topic	Research method	Research results
Language documentation	documentary collection of data prior to language extinction	recording & editing of natural language data	language corpora
Descriptive linguistics	structural description of previously unstudied languages	collection and analysis of systematic survey (elicitation)	grammars & dictionaries
Language typology	commonalities & differences in the languages of the world	cross-linguistic comparison based on published data (grammars)	universals & typologies
Corpus linguistics	patterns of language use (primarily in major languages)	analysis of natural language data	frequencies of occurrence, concordances & collocations
Anthropological linguistics	relationship between language & culture	participant observation & survey for data collection, analysis of language data in its cultural context	cultural conceptualisations, ethnographies of speaking
Sociolinguistics	correlation between language & social features of its speakers	correlation analysis (features in natural language representative of distinct social subgroups)	social markers & sociolects, dialects & atlases
Cognitive linguistics	relationship between language & thought	language analysis (regarding mental concepts) & experimental tasks	impact of language on thought (incl. ‘thinking for speaking’)
Psycholinguistics	mental concepts & processes in language production & comprehension	language experiments (behavioural offline and online measures)	temporal and conceptual organisation of language processing
Neurolinguistics	operations of language processing in the brain	language experiments (neuroscientific online measures)	neurocognitive brain systems and temporal & spatial patterns of brain (de)activation

Summary of the subdisciplines

9.1

Basic types of research depending on the following parameters:

- **types of studied languages**:
  - well-studied major languages vs. understudied minor languages
  - single languages vs. multiple languages
  - languages vs. language varieties
- **types of research environments**:
  - field vs. office vs. laboratory research
- **types of research data**:
  - primary vs. secondary or tertiary language data
  - natural language data vs. language data that is generated explicitly by/for the research
  - written vs. spoken or signed language data
- **types of research approaches**:
  - explorative vs. hypothesis-testing approaches
  - quantitative vs. qualitative approaches

Research tendencies per subdiscipline:

	Studied language	Research environment	Research data	Research approach
Language documentation	single understudied minor languages	field research	primary data (natural, primarily oral language data)	qualitative approach
Descriptive linguistics	single understudied minor languages	field research	primary data (primarily generated/ non-natural language data)	descriptive research, qualitative approach
Language typology	multiple languages of the world (vs. artificial or constructed languages)	research in office environments	primarily secondary/tertiary data (linguistic descriptions of others)	quantitative or qualitative approach
Corpus linguistics	major languages (primarily single languages or language varieties)	research in office environments	primarily secondary data (natural language data compiled by others)	quantitative approach
Anthropological linguistics	single less studied minor languages	field research (primarily non-Western locations)	primary data (natural, primarily oral language data)	descriptive or explanatory research, qualitative approach
Sociolinguistics	single language varieties	field research (primarily Western locations) or	primary or secondary data (natural, primarily oral language data)	descriptive or explanatory research, quantitative or qualitative approach
Cognitive linguistics	single less studied minor languages	field research	primary data (natural language data or data from language-related tasks)	explanatory research
Psycholinguistics	single major languages	laboratory research	primary data (natural language data or data from language-related tasks)	explanatory research
Neurolinguistics	single major languages (vs. communicative systems of non-human species)	laboratory research (n locations with advanced socioeconomic standards)	primary data (natural language data or data from language-related tasks)	explanatory research

👍 Advantages of methodological procedures:

- **collection and editing of natural language data**: optimal to save linguistic knowledge as fast as possible prior to its extinction, as it can also be carried out by trained non-linguists
- **work with standardised language/linguistic data**: better for cross-linguistic comparability
- **work with systematic elicitation**: better suited for capturing language phenomena that rarely occur in natural language corpora
- **analysis of (representative) natural language corpora**: best or even only way to get information on unconscious language behaviour (such as slips of the tongue) and on actual occurrence patterns (such as frequency)
- **observation of native speakers in their natural sociocultural environment**: best way to study the relationship between language and society/culture and for detecting relevant (emic) parameters
- **work under controlled conditions**: allows for the systematic examination of the influence of a manipulated variable, excluding the influence of other variables

Reflect on disadvantages/points of critique regarding single methodological procedures.

👎 Points of critique regarding methodological procedures:

- **bias towards the major languages** in corpus linguistics and variationist sociolinguistics and often an overrepresentation of written data in large corpora
- **lack of (sufficiently detailed) contextual information** (e.g., socio-cultural data on the speakers/authors, data on the social context of production, etc.) on the language data and its consideration in studies that rely on secondary data (e.g., ready-made corpora)
- **comparatively (too) small corpus size** in documentary linguistics as compared to the size of ready-made corpora of the major languages and, hence, the risk of non-representativeness
- **insufficient consideration of language-internal variation** (as it occurs in natural language) in descriptive studies which are based on systematically elicited (idealised) language data
- **risk of working with bad data** in studies that rely on previous work done by others (i.e., secondary or tertiary data), particularly if the quality of this data is not reflected upon (e.g., how representative is an existing corpus) or cannot or only hardly be checked

👎 Points of critique regarding methodological procedures:

- work with an **unbalanced sample or the overrepresentation** of better described languages in typological research for reasons of convenience
- extreme **simplification of linguistic complexity** as it occurs in the individual languages for classification purposes
- rather **low generalisability** of findings in anthropological linguistics and sociolinguistics due to research in very specific situations/contexts with many interacting parameters which are hardly or not comparable across languages/varieties and language communities
- **risk of generating Western-centric findings** in psycho- and neurolinguistics (together with age and education biases) due to the overrepresentation of research participants from so-called WEIRD (Western, Educated, Industrialised, Rich, and Democratic) backgrounds and studies on the major Indo-European languages
- issue of **whether laboratory results reflect language in natural contexts**

=> **No single empirical approach** is suitable to answer all linguistic questions.

Therefore, **research of different subdisciplines** contributes and interacts to foster a deeper **understanding of language in its many facets.**

Interfaces of the subdisciplines

9.2

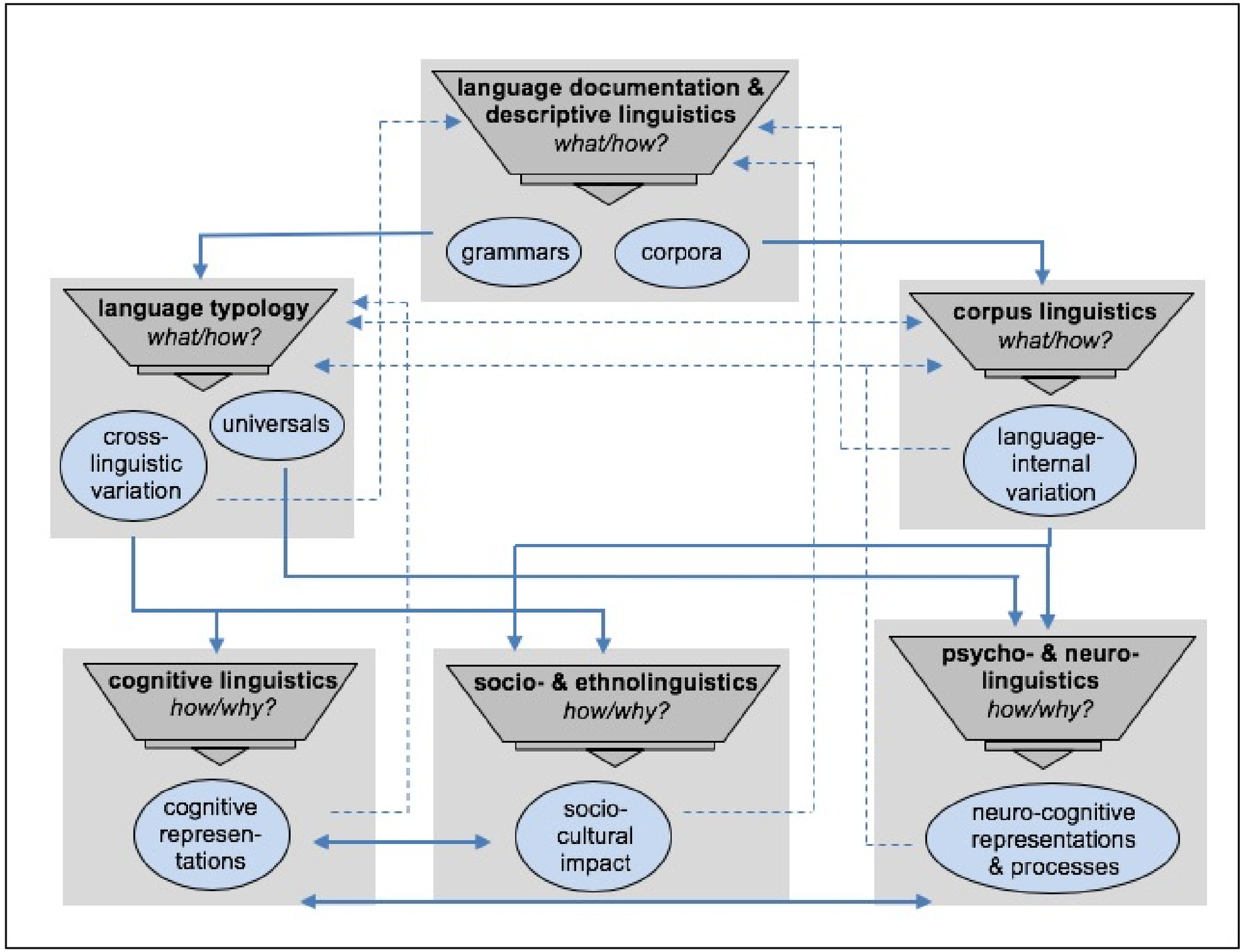


Figure 24

Interfaces of the subdisciplines

9.2

Collaboration across subdisciplines:

- **cross-disciplinary embedding**: reflections on the significance and the contribution of one's own research within the entire field of empirical linguistics
- **cross-disciplinary methodological transfer**: consideration of methodological procedures of other subdisciplines for the research in one's core area
- **cross-disciplinary genesis**: research outcomes provide data for further research in other subdisciplines or research builds upon findings of other subdisciplines
- **cross-disciplinary cooperation**: joint research projects across subdisciplines
  → *multi-methods combinations* vs. *mixed-methods designs*
  → research topics studied with multiple methods may result in *converging* or *diverging evidence*

Reflect on the aspects that different linguistic subdisciplines contribute to the understanding of first language acquisition. see pages 273-274 for some considerations

Interdisciplinarity:

a collaboration across two or more disciplines that includes **a joint research process** (i.e., researchers deal with the methods, procedures, approaches, and ideas of the different disciplines and different research approaches)

→ interdisciplinarity *in person* vs. *in teams*

Multidisciplinarity:

different disciplines only meet to **consolidate their respective findings** concerning a special topic but there is no empirical collaboration in terms of a joint research process

Interdisciplinary subdisciplines:

1 – sociolinguistics
2 – anthropological linguistics/ linguistic anthropology
3 – cognitive linguistics
4 – psycholinguistics
5 – neurolinguistics

→ 3, 4 and 5 are linguistic sub-disciplines which are part of ***cognitive science***
(an interdisciplinary network including neuroscience, artificial intelligence, psychology, philosophy, anthropology, and linguistics)

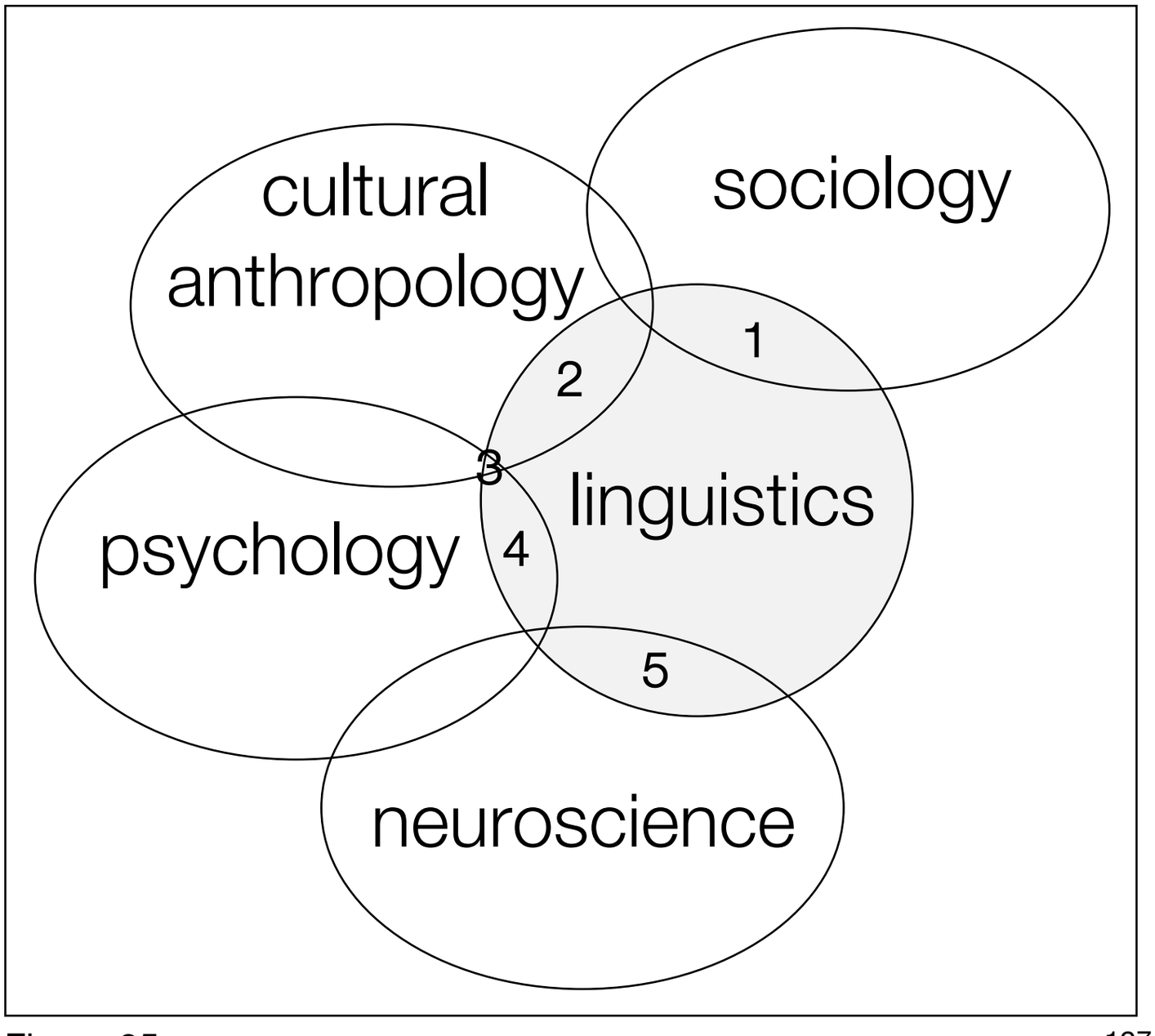


Figure 25

Technological impact:

- **measurement techniques or instruments**
  - new or better methods and possibilities in data collection (e.g., neurocognitive methods)

- IT developments overall (e.g., increased computer performance)
  - new or improved **software tools for the analysis of larger amounts of data**
    (e.g., language corpora, the Internet as data source)
  - vast **data storage capacities**
    (e.g., language corpora, linguistic databases)

    → *data management*

Data management:

- optimal handling of data as a valuable resource to ensure their maximum benefit
- aspects to consider: - data format
  - data storage and preservability
  - data governance and expandability
  - data access
  - data quality
  - ethics and data rights

Discuss what kind of data allows for data management and what kind of data is problematic and for what reason. see pages 277-279 for some considerations

Research under pandemic circumstances:

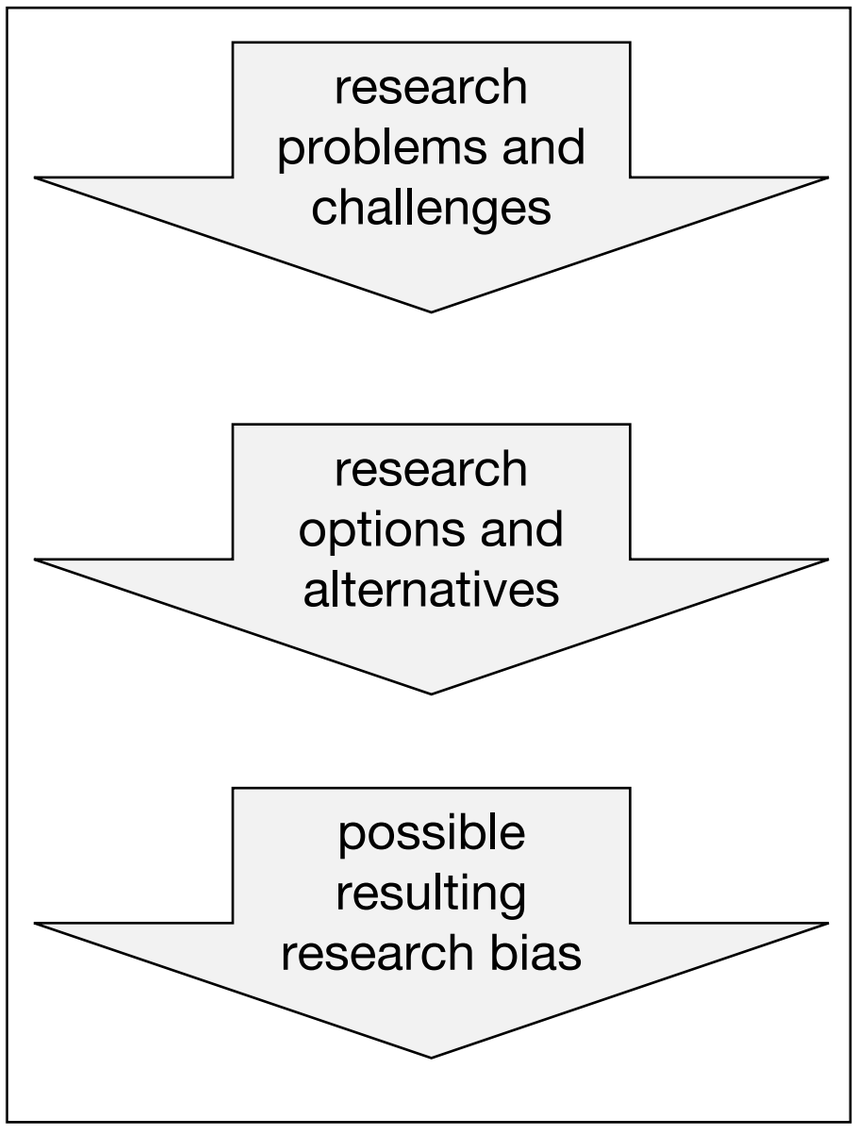


Figure 26

- **travel restrictions**: no/hardly access to distant field sites (primarily in anthropological linguistics and documentary/descriptive linguistics)
- **contact restrictions**: restricted face-to-face work with research participants in the field and in the laboratory (primarily in anthropological linguistics/sociolinguistics, documentary/descriptive linguistics, psycho-/neurolinguistics)

- **use of existing/available data** such as corpora, databases and the internet as a data source on its own
- **work with research participants via the Internet**
- face-to-face work with research participants under **appropriate safety regulations**

- lack of studies considering data in its natural socio-cultural context apart from the Internet
- overrepresentation of internet-based language varieties/ genres
- more research on better accessible languages
- predominant work with better accessible research participants (living in the researchers direct surrounding, medically less endangered groups, etc.)

Conclusion 9.5

This introduction has provided **a comprehensive cross-disciplinary overview of the broad field of linguistic research**, including planning aids for empirical projects – to identify the research components and parameters, to consider at least some potential challenges, and to evaluate pros and cons of methodological options.

However, empirical research cannot be learned theoretically.

→ It requires **practical experience** !!!

So, we strongly encourage you to design and conduct your own empirical research projects and hope this introduction provides **a helpful guideline in the process of developing your own empirical interests in linguistics or of broadening your research focus.**

Develop your own research project.
Start from your collection of research topics/questions (as collected in your research diary, cf. Section 1.2) or see pages 103-104, 130-131, 162-163, 192, 226 and 258 for some subdiscipline-specific project ideas.

Appendix

Ideas and guidelines

Appendix: Outline

Exposé/proposal A1

Research diary A2

Poster presentation A3

Feedback A4

Exposé/proposal: an overview of a research project at the end of the planning process

→ intended to inform the supervisors and/or to apply for research funding

Basic elements of research proposals:

- research topic
- description of the problem to be studied by the research project
- embedding in the current state of research (including own preliminary work)
- research question/hypothesis
- methodology to study the research question
- time frame, chronological sequence and duration of individual project phases
- required resources (e.g., participants, technical/software equipment, access to research literature) - particularly funding
- bibliography

Write a short research proposal/exposé for your own empirical project (about 3 pages).

Research diary: **personal notes** on project development & insights to individual processes (to be distinguished from scientific records, cf. ‘written notes’ in Section 1.2)

→ intended to structure & reflect on the research, to trace individual steps

Formal requirements:

- **accuracy/exactness of the notes** (the entries need to be made in a timely manner to avoid effects of forgetfulness and the distortion of events over time)
- **regularity of the entries** (usually on a daily basis)
- **systematic, clear, and transparent structure in chronological order** (by date and possibly further sub-structuring units (e.g., ‘to do list’ vs. ‘descriptions of performed activities’ or ‘facts’ vs. ‘impressions/ feelings’ – indicated by colour or by page division)

Possible contents of research diaries:

- **development of the research question** (e.g., areas of interest, topic selection, specification of research questions or hypotheses)
- **literature** (e.g., accessibility, completion status of screened literature, relevant content and quotes, bibliographical data)
- **methodological considerations** (e.g., reflections of pros and cons with regard to the research question/ hypothesis, access to needed resources (such as funding, research permissions, participants, data, technical equipment, software), need for pre-tests, plan of procedure, etc.)
- **implementation period** (e.g., functionality and effectiveness of procedures, reasons for problems or failure, necessary adaptations to unexpected occurrences, personal perceptions, self-observation and reflections)
- **teamwork** (e.g., interpersonal relationships, content of group meetings, division of labour and clarification of interfaces, understanding of the co-researchers' work, difficulties and successes of the joint project)

Write a research diary accompanying your own empirical project. Start identifying your topic and end the research diary with reflections on the feedback of the poster presentation.

Poster: presentation of an empirical study clearly arranged on one large page

→ presents the research to a broader scientific audience (usually at conferences; here: as form of examination in class).

→ serves as guideline for a short oral presentation of the research project but the poster should be meaningful on its own.

Content: **essential key information** on the research project, such as

- topic: current state of research, research question/hypothesis of the study, relevant definitions
- methodological procedure: methods of data collection and data analysis
- results including examples
- discussion
- references

Poster layout:

- **generally DIN A0** in portrait or landscape format
- **head section**: title, author(s), research affiliation, funding institution
- **a clear comprehensible structure**: sections/boxes with subheadings arranged in sequence guided by the content and taking reading direction into consideration
- **an engaging amount of text** (written in full sentences or in keywords)
- **visualisations** (illustrations, tables, etc.) give a fast overview on crucial aspects and to make the overall appearance more appealing.
- **legible font size**: Font sizes smaller than 18pt may not be legible from a distance; references however can be printed in a smaller font size.
- **intellegible respectable fonts & colours**: Different colours/fonts/font sizes indicate distinct aspects/sub-issues; similar aspects should be presented in the same way.

Develop a poster of your own completed research project. If you have the possibility to present it, prepare a short oral presentation of your project based on the poster (no more than ten minutes).

Feedback: written or oral reactions of the scientific audience to presentations/publications

→ intended to exchange thoughts & improve the quality of the research

Feedback rules

... for feedback givers:

- **be constructive & respectful** – otherwise you generate rejection
- **mention positive & negative aspects** – preferably in this order to create an atmosphere of acknowledgement
- **be precise & mention examples that illustrate your point**
- **give suggestions for improvement, but don‘t insist** – there are usually different ways of looking at things

... for feedback takers:

- **listen carefully & do not justify yourself** – it is finally up to you which feedback you take into account & how
- **be thankful** – feedback is valuable input

Include the feedback of your own poster presentation together with your reflections on it in your research diary (see Appendix A2).

References cited on the slides

R1

Chapter 1: Empirical research in linguistics

von Alemann, Heine. 19842. *Der Forschungsprozess: Eine Einführung in die Praxis der empirischen Sozialforschung*. Stuttgart: Teubner.
Barnett, Lincoln. 1948. *The universe of Dr. Einstein*. Mineola: Dover Publications.
Miner, Horace. 1956. 'Body ritual among the Nacirema'. *American Anthropologist*, 58. 503–507.
Popper, Karl. 1973. *Objektive Erkenntnis: Ein evolutionärer Entwurf*. Hamburg: Hoffmann und Campe.

Chapter 2: Basic methods for data collection

Biemer, Paul; Lars Lyberg. 2003. *Introduction to Survey Quality.* Hoboken, NJ: John Wiley & Sons.
Cordaro, Lucian; James Ison. 1963. 'Psychology of the scientist: X. Observer bias in classical conditioning of the planarian'. *Psychological Reports*, 13. 787–789.
Grosjean, Francois. 1980. 'Spoken word recognition processes and the gating paradigm.' *Perception and Psychophysics*, 28. 267-283.
Ivankova, Nataliya; John Creswell. 2009. 'Mixed methods'. In Juanita Heigham & Robert Croker (eds). *Qualitative Research in Applied Linguistics: A Practical Introduction.* Houndmill: Palgrave Macmillan. 135–163.
Kirk, Roger E. 2003. 'Experimental design'. In John Schinka & Waye Velicer (eds). *Handbook of Psychology, vol. 2 (Research methods in psychology)*. Hoboken, NJ: John Wiley & Sons. 3–32.
Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
Pike, Kenneth. 1967[2]. *Language in Relation to a Unified Theory of Structure of Human Behavior*. Den Haag: Mouton.
Wagner, Elvis. 2010. 'Survey research'. In Brian Paltridge & Aek Phakiti (eds). *Continuum Companion to Research Methods in Applied Linguistics*. London: Continuum. 22–38.
Zwitserlood, Pienie. 1989. 'The locus of the effects of sentential-semantic context in spoken word processing'. *Cognition,* 32. 25-64.

Chapter 3: Language documentation and descriptive linguistics

Chelliah, Shobhana; Willem de Reuse. 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht: Springer.
Dixon, Robert. 2007. 'Field linguistics. A minor manual'. *Sprachtypologie und Universalienforschung*, 60 (1). 12–31.
Dixon, Robert. 2010 & 2012. *Basic Linguistic Theory*, vol. 1–3. Oxford: Oxford University Press.
Evans, Nicholas ; Alan Dench. 2006. . 'Introduction. Catching language'. In Felix Ameka, Alan Dench & Nicholas Evans (eds). *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter. 1-40.
Everett, Daniel. 2001. 'Monolingual field research'. In Paul Newman & Martha Ratliff (eds). *Linguistic Fieldwork*. Cambridge: Cambridge University Press. 166–188.
Everett, Daniel. 2013. *Monolingual Fieldwork Demonstration.* [www.youtube.com/watch?v=sYpWp7g7XWU].
Foley, William. 2002. 'Field methods'. In Kirsten Malmkjær (ed.). *The Linguistics Encyclopedia*. 131–137.
Gippert, Jost; Nikolaus Himmelmann; Ulrike Mosel (eds). 2006. *Essentials of Language Documentation*. Berlin: Mouton de Gruyter.
Harrison, David; Greg Anderson. 2008. *The Linguists*. Garrison, NY: Ironbound Films.
Himmelmann, Nikolaus. 1998. 'Documentary and descriptive linguistics'. *Linguistics*, 36. 161–195.
Lehmann, Christian. 1999. *Documentation of Endangered Languages: A Priority Task for Linguistics*. (Arbeitspapier, 1). Erfurt: Institut für Sprachwissenschaft.

References cited on the slides

Mosel, Ulrike. 2006. 'Grammaticography. The art and craft of writing grammars'. In Felix Ameka, Alan Dench & Nicholas Evans (eds). *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter. 41–68.

Noonan, Michael. 2006. 'Grammar writing for a grammar-reading audience'. *Studies in Language*, 30 (2). 351–365.

Payne, Thomas. 1997. *Describing Morphosyntax. A Guide for Field Linguists*. Cambridge: Cambridge University Press.

Payne, Thomas; David Weber (eds). 2007. *Perspectives on Grammar Writing*. Amsterdam: Benjamins.

Woodbury, Anthony. 2011. 'Language documentation'. In Peter Austin & Julia Sallabank (eds). *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press. 159–186.

Chapter 4: Language typology

Croft, William. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.

Daniel, Michael. 2011. 'Linguistic typology and the study of language'. In Jae Jung Song (ed.). *The Oxford Handbook of Language Typology*. Oxford: Oxford University Press.

Dryer, Matthew. 1989. 'Large linguistic areas and large sampling'. *Studies in Language*, 13. 257–292.

Dryer, Matthew. 1992. 'The Greenbergian word order correlations'. *Language*, 68. 81–138.

Dryer, Matthew; Martin Haspelmath (eds). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [http://wals.info].

Greenberg, Joseph. 1963 (1966). 'Some universals of grammar with particular reference to the order of meaningful elements'. In Joseph Greenberg (ed.). *Universals of Language*. Cambridge, MA: MIT Press. 73–113.

Moravcsik, Edith. 1993. 'Government'. In J. Jacobs et al. (eds). *Syntax. An International Handbook of Contemporary Research.* Berlin: de Gruyter. 705-721.

Hawkins, John. 1983. *Word Order Universals*. New York: Academic Press.

Pesetsky, David. 1995. *Zero Syntax. Experiencers and Cascasdes.* Cambridge, MA: MIT Press.

Rijkhoff, Jan; Dik Bakker. 1998. 'Language sampling'. *Linguistic Typology*, 2. 263–314.

Rijkhoff, Jan; Dik Bakker; Kees Hengeveld; Peter Kahrel. 1993. 'A method of language sampling'. *Studies in Language*, 17 (1). 169–203.

Ruhlen, Merrit. 1987. *A Guide to the World's Languages*, vol. 1. (*Classification*). London: Edward Arnold.

Simons, Gary; Charles Fennig (eds). 201720. *Ethnologue. Languages of the World*. Dallas, TE: SIL International. [http://ethnologue.com].

Tomlin, Russell. 1986. *Basic Word Order: Functional Principles*. London: Croom Helm.

Velupillai, Viveka. 2012. *An Introduction to Linguistic Typology*. Amsterdam: Benjamins. 43-68.

Whaley, Lindsay. 1997. *Introduction to Typology. The Unity and Diversity Of Language*. Thousand Oaks, CA: Sage Publications.

References cited on the slides

R1

Chapter 5: Corpus linguistics

Biber, Douglas. 1988. *Variation Across Speech and Writing.* Cambridge: Cambridge University Press.

Gatto, Maristella. 2014. *Web as Corpus: Theory and Practice.* London: Bloomsbury.

Gilquin, Gaëtanelle; Stefan Gries. 2009. 'Corpora and experimental methods. A state-of- the-art review'. *Corpus Linguistics and Linguistic Theory*, 5(1). 1–26.

Lemnitzer, Lothar; Heike Zinsmeister. 2015[3]. *Korpuslinguistik. Eine Einführung.* Tübingen: Narr.

Lüdeling, Anke. 2007. 'Das Zusammenspiel von qualitativen und quantitativen Methoden in der Korpuslinguistik'. In Gisela Zifonun & Werner Kallmeyer (eds). *IDS-Jahrbuch 2006*. Berlin: de Gruyter. 28–48.

Lüdeling, Anke. 2017. 'Grammatische Variation. Empirische Zugänge und theoretische Modellierung'. In Marek Konopka & Angelika Wöllstein (eds). *Jahrbuch des Instituts für Deutsche Sprache 2016.* Berlin: de Gruyter. 129–144.

McEnery, Tony; Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press.

Meyer, Charles. 2002. *English Corpus Linguistics: An Introduction.* Cambridge:Cambridge University Press.

Rayson, Paul. 2015. 'Computational tools and methods for corpus compilation and analysis'. In Douglas Biber & Randi Reppen (eds). *The Cambridge Handbook of English Corpus Linguistics.* Cambridge: Cambridge University Press. 32–49.

Teubert, Wolfgang. 2005. 'My version of corpus linguistics'. *International Journal of Corpus Linguistics*, 10(1). 1–13.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: Benjamins.

Chapter 6: Sociolinguistics and anthropological linguistics

Duranti, Alessandro. 1997. *Linguistic Anthropology*. Cambridge: Cambridge University Press.

Foley, William. 1997. *Anthropological linguistics*. Malden, MA: Blackwell.

Holmes, Janet. 20134. *An Introduction to Sociolinguistics*. London: Routledge.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Malinowski, Bronislaw. 1954. *Magic, Science and Religion and Other Essays*. New York: Doubleday.

Milroy, James; Lesley Milroy. 1985. 'Linguistic change, social network and speaker innovation'. *Journal of Linguistics*, 21. 339–384.

Sidnell, Jack; Tanya Stivers (eds). 2012. *The Handbook of Conversation Analysis*. Malden, MA: Wiley-Blackwell.

Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Tannen, Deborah; Heidi Hamilton; Deborah Schiffrin (eds). 20152. *The handbook of discourse analysis*. Malden, MA: Wiley-Blackwell.

Chapter 7: Cognitive linguistics and psycholinguistics

Altmann, Gerry. 2001. 'The language machine. Psycholinguistics in review'. *British Journal of Psychology*, 92. 129–170.

Barcelona, Antonio; Javier Valenzuela. 2011. An overview of cognitive linguistics. In Mario Brdar, Stefan Th. Gries & Milena Zic Fuchs (Eds.) *Cognitive linguistics: convergence and expansion*. Amsterdam: John Benjamins. pp17-46

References cited on the slides

R1

Casasanto, Daniel. 2017. ‘Relationships between language and cognition’. In Barbara Dancygier (ed.). *Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press. 19–37.

Evans, Vyvyan; Benjamin Bergen; Jörg Zinken (eds). 2007. *The Cognitive Linguistics Reader*. London: Equinox.

Fauconnier, Gilles. 2006. ‘Cognitive Linguistics’. In Lynn Nadel (ed.). *Encyclopedia of Cognitive Science*. Malden, MA: John Wiley. https://doi.org/10.1002/0470018860.s00214

Fernández, Eva; Helen Smith Cairns. 2010. *Fundamentals of Psycholinguistics*. Malden, MA: Wiley-Blackwell.

Henrich, Joseph; Steven Heine; Ara Norenzayan. 2010. ‘The Weirdest People in the World?’. *Behavioral and Brain Sciences*, 33 (2–3). 61–83.

Wolff, Phillip; Kevin Holmes. 2011. ‘Linguistic relativity’. *Cognitive Science*, 2(3). 253–265.

Chapter 8: Neurolinguistics

Ahlsén, Elisabeth. 2006. *Introduction to Neurolinguistics.* Amsterdam: Benjamins.

Andric, Michael; Steven Small. 2015. ‘fMRI methods for studying the neurobiology of language under naturalistic conditions’. In Roel Willems (ed.). *Cognitive Neuroscience of Natural Language Use.* Cambridge: Cambridge University Press. 8–28.

Bornkessel-Schlesewsky, Ina; Matthias Schlesewsky.2009. Processing Syntax and Morphology: A Neurocognitive Perspective. Oxford: Oxford University Press.

Caplan, David. 1987. *Neurolinguistics and Linguistic Aphasiology*. (Cambridge studies in speech science and communication). Cambridge: Cambridge University Press.

Gazzaniga, Michael; Richard Ivry; George Mangun. 2014[4] . *Cognitive Neuroscience: The Biology of the Mind*. New York: Norton.

Horvath, Jared C.; Jennifer M. Perez; Lachlan Forrow; Felipe Fregni. 2010. ‘Transcranial magnetic stimulation: a historical evaluation and future prognosis of therapeutically relevant ethical concerns.’ *Journal of Medical Ethics*, 37, 137-143.

Ingram, John. 2007. *Neurolinguistics. An Introduction to Spoken Language Processing and Its Disorders*. New York: Cambridge University Press.

Kemmerer, David. 2015. *Cognitive Neuroscience of Language*. New York: Psychology Press.

Meek, Judith. 2002. ‘Basic principles of optical imaging and application to the study of infant development.’ *Developmental Science*, 5(3), 371-380.

Mehler, Jacques; Marina Nespor; Judith Gervain; Ansgar Endress; Mohinish Shukla. 2008. Mechanisms of Language Acquisition: Imaging and Behavioral Evidence. In Charles A. Nelson & Monica Luciana (eds.). *Handbook of developmental cognitive neuroscience*, 325-335.

Rugg, Michael. 1999. ‘Functional neuroimaging in cognitive neuroscience’. In Colin Brown & Peter Hagoort (eds). 1999. The Neurocognition of Language. Oxford: Oxford University Press. 15–36.

Small, Steven L.; Gregory Hickok. 2016. 'The neurobiology of language' In Gregory Hickok & Steven L. Small (eds.) *Neurobiology of Language*. Amsterdam: Elsevier. 3-9.

de Zubicaray, Greig; Niels Schiller (eds). 2019.*The Oxford Handbook of Neurolinguistics.* New York: Oxford University Press.

see pages 281-308 for a more comprehensive list of references

Image credits

slide 1 (title page): https://pixabay.com/de/photos/laptop-mac-computer-browser-2557615/

slide 10 (ch. 1 title page): https://pixabay.com/de/illustrations/fragen-schrift-wer-was-wie-warum-2245264/

slide 46 (ch. 2 title page): https://pixabay.com/de/illustrations/erfolg-strategie-geschäft-lösung-2081168/

slide 85 (ch. 3 title page): https://pixabay.com/de/photos/übersetzung-ferien-besuchen-sie-4620781/

slide 100 (ch. 4 title page): https://pixabay.com/de/photos/hände-welt-karte-global-erde-600497/

slide 117 (ch. 5 title page): https://pixabay.com/de/illustrations/nachrichten-zeitung-globus-1074604/

slide 135 (ch. 6 title page): https://pixabay.com/de/vectors/social-media-verbindungen-vernetzung-3846597/

slide 152 (ch. 7 title page): https://pixabay.com/de/illustrations/kommunikation-kopf-sprechblasen-1991851/ and https://pixabay.com/de/illustrations/mindmap-karte-rechtecke-brainstorm-6238107/

slide 170 (ch. 8 title page): https://pixabay.com/de/illustrations/binär-code-datenschutz-gehirn-5137349/

slide 185 (ch. 9 title page): https://pixabay.com/de/vectors/integration-datenintegration-2031395/

slide 202 (appendix title page): https://pixabay.com/de/illustrations/und-sprache-kommunikation-ordnung-99375/

Image credits

Figure 1: modified from Figure 1.1. in Völkel & Kretzschmar (2021: 4)

Figure 2: modified from Figure 1.2 in Völkel & Kretzschmar (2021: 5)

Figure 3: Figure 1.3 in Völkel & Kretzschmar (2021: 6)

Figure 4: Figure 1.4 in Völkel & Kretzschmar (2021: 7)

Figure 10: Figure 2.1 in Völkel & Kretzschmar (2021: 44)

Figure 11: Figure 2.2 in Völkel & Kretzschmar (2021: 48)

Figure 12: Figure 2.3 in Völkel & Kretzschmar (2021: 54)

Figure 14: Figure 2.4 in Völkel & Kretzschmar (2021: 61)

Figure 15: Figure 2.5 in Völkel & Kretzschmar (2021: 70)

Figure 16: modified from Figure 3.1 in Völkel & Kretzschmar (2021: 95)

Figure 19: Figure 5.1 in Völkel & Kretzschmar (2021: 136)

Figure 20: modified from Figure 6.1 in Völkel & Kretzschmar (2021: 175)

Figure 23: Figure 8.1 in Völkel & Kretzschmar (2021: 235)

Figure 24: Figure 9.1 in Völkel & Kretzschmar (2021: 271)