

1 Empirical Research in Linguistics

In this introductory chapter, we address some basics of empirical research (1.1) and illustrate some aspects of the research process (1.2). Our focus is on key concepts and classifications relevant for linguistic studies in distinct subdisciplines. After covering these foundations, the chapter contains suggestions for smaller exercises (to gain practical experience with individual research procedures) and tasks (to support readers in developing their own empirical project). Finally, we provide a summary of the contents of the chapter (1.3), followed by suggestions for further in-depth readings on diverse issues in linguistic research (1.4).

1.1 Basics of Empirical Research

Section 1.1 provides an overview of empirical research. Starting from considering what research is in the first place (1.1.1) and also looking at the interaction of empiricism and theory (1.1.2), we focus on the research process and its stages (1.1.3), as well as research components and basic classifications of types of research (1.1.4). Finally, in Sections 1.1.5 and 1.1.6, we address two important basic requirements of empirical research, namely quality criteria and research ethics.

1.1.1 Research

As many well-known researchers have aptly noted, an impression, or even a strong feeling that something is or is not the case, does not mean that this is actually true. According to Barnett (1948: 58), Einstein once said that ‘common sense is actually nothing more than a deposit of prejudices laid down in the mind prior to the age of eighteen’. Durkheim (2006 [1897]) similarly points out that the common sense is vague and unreliable and that we can only learn about the social world by thorough research.

But What Exactly is ‘Research’?

Research is the **systematic search for new knowledge**. It is generally pursued in academic disciplines – for which gaining new insights and contributing to the growth of knowledge is a crucial aim and systematic research

a crucial approach. This means that existing knowledge about the phenomena in the world (such as language) always constitutes the foundation for further research. In this way, knowledge is constantly being questioned and built upon.

In detail, research questions or even concrete hypotheses are formulated and then systematically examined on the basis of concrete experienced facts **by using an adequate methodological procedure**. The research taking place in this framework is required to fulfil certain **scientific standards** (i.e., quality criteria and research ethics, cf. Sections 1.1.5 and 1.1.6).

Curiosity and a fundamental interest in finding answers to as-yet-unknown phenomena as well as common sense; an analytical understanding of complex issues are basic requirements for successful research. **Common sense** and **intuition** are vague, imprecise and unreliable – as stated above – and cannot replace systematic data-based research. Nevertheless, they remain important key tools for research. On the part of the researcher, common sense and intuition can serve as general guidance. They help to identify interesting issues, questions and hypotheses, which are the starting point for every research project, to find appropriate methods for investigation, and to interpret and reflect on the findings. However, it is crucial to ensure that the researcher’s intuition does not result in personal bias (cf. Section 1.1.5). On the part of research participants, intuition can provide valuable data, such as linguistic judgements or self-evaluation.

1.1.2 Empiricism and Theory

Empiricism and theory are often perceived to be a pair of terms with opposing meaning. However, as Popper (1963, 1973) has pointed out, research is an evolutionary process, consisting of the following three components: problems/issues, theoretical considerations, and empirical examination (see Figure 1.1).

This means that empiricism and theory actually go hand in hand. A theory is a mental system describing and explaining a phenomenon based on existing scientifically produced knowledge. This abstract system or construct serves not only to explain regularities regarding the phenomenon but also to make

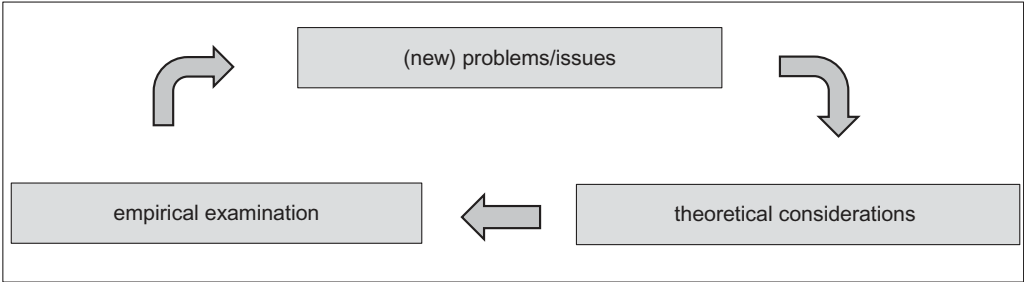


Figure 1.1 *The process of research*

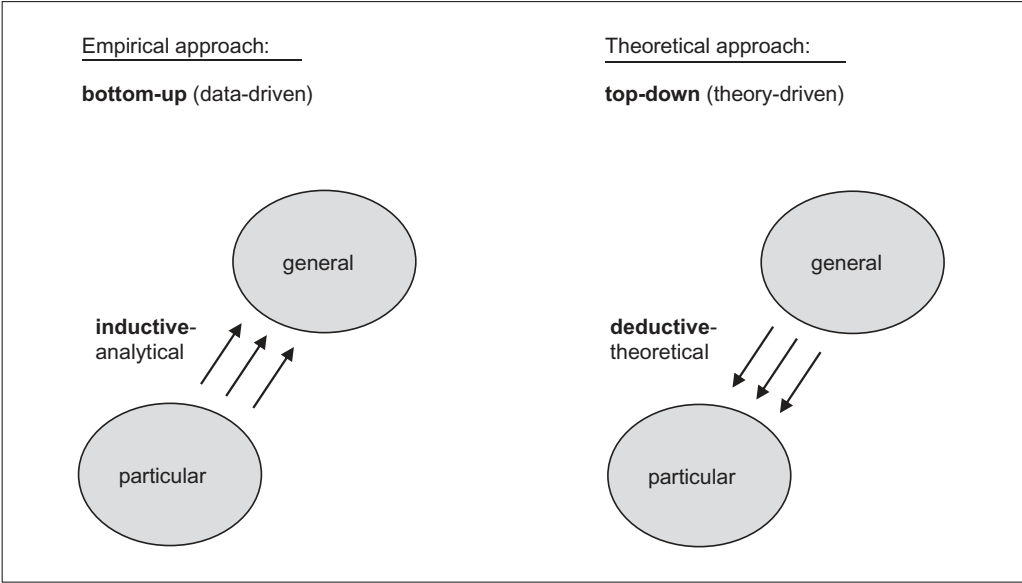


Figure 1.2 *The empirical and theoretical approach*

predictions. In the further process, these theoretical assumptions need to be verified by systematic data-based research. Based on new empirical findings, new issues arise. Either the empirical study confirms the theory, it necessitates modifications, or it disproves the theory completely and a new one must be formulated. In any case, it is always an improvement of (academic) knowledge in a continuous evolutionary process.

Although scientific progress is based on theoretical considerations as well as empirical investigations, at base, empiricists and theorists take different approaches (see Figure 1.2). While empiricists start with data (i.e., actually occurring facts), which is analysed in order to identify general underlying structures or practices, theorists operate in the opposite direction, starting with a theoretical framework and deriving individual predictions that can subsequently be tested. More precisely, the theorist searches for evidential data in support of theoretical considerations.

A radical empiricist would conduct purely explorative studies and abstain from any kind of anticipation regarding the data to be analysed. In contrast, a radical theorist would disregard empirical data – either for the development of a theory or even for its verification. These two opposing positions, however, rarely occur in their extreme form. Just as empirical research generally includes theoretical assumptions of some kind (i.e., a specific research focus or procedure based on theoretical premises, such as hypotheses or criteria for data classification), a theory needs to be tested against empirical findings. Thus, ideally, there is a dialogue between theorists and empiricists.

1.1.3 The Research Process and its Main Stages

The illustration of Alemann (1984²: 152–153, redrawn by Angelika Morgh, with modifications and translations by the authors) in Figure 1.3 offers a good depiction of the empirical research process in its overall complexity.

A clear takeaway from this illustration is that the research process is by no means a straight path but rather a windy road with detours, forking roads, and backtracking, all of which a researcher cannot entirely anticipate in advance. The researcher has to evaluate options, make decisions, cope with challenges, work simultaneously on different aspects, recognise interrelationships, and so on and so forth. All in all, it is crucial to **plan a research project thoroughly** but also to **remain flexible regarding potential replanning for practical reasons** (e.g., the sudden unavailability of data sources, uncooperative behaviour of research participants, or misunderstandings resulting in unusable data). Thus, **research practice entails, by definition, the unforeseeable**, the best laid plans notwithstanding.

Therefore, even researchers with extensive research experience will still encounter challenging situations during their projects. Nevertheless, previous **research experience** does provide helpful guidance when navigating this sometimes-difficult path. Ultimately, empirical expertise is gained primarily through practice. Therefore, this book can only raise awareness regarding

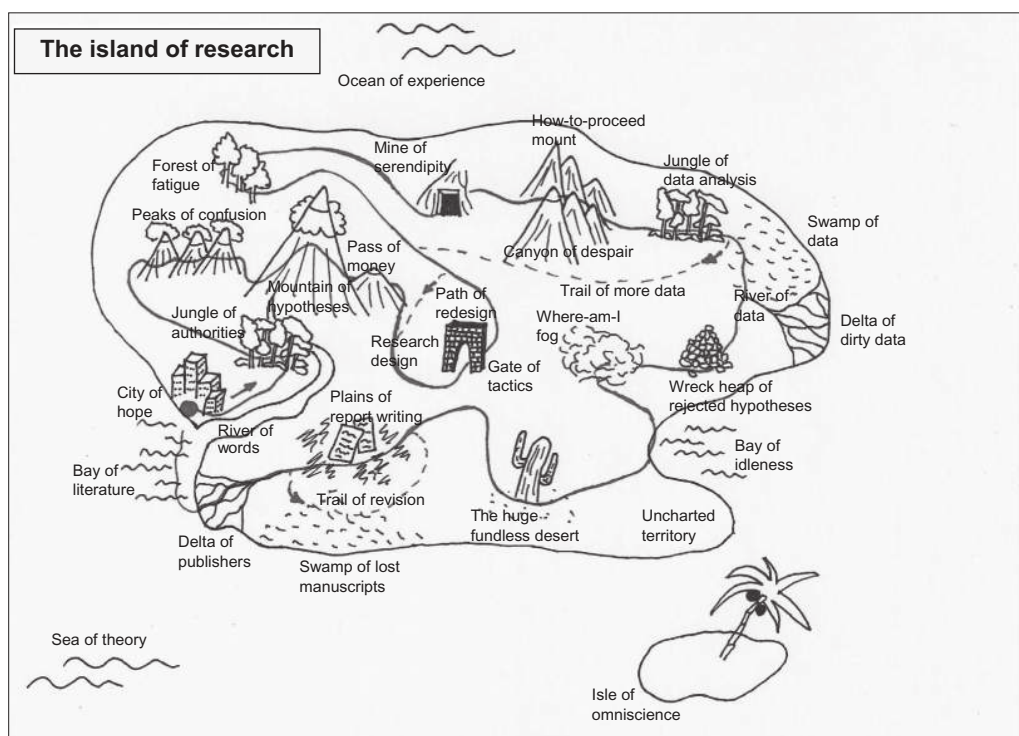


Figure 1.3 *The island of research*

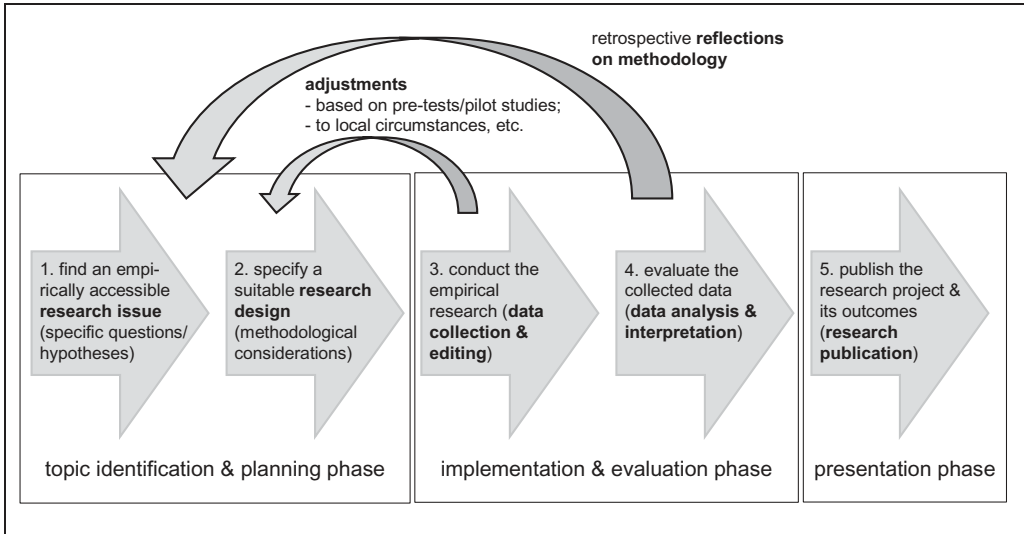


Figure 1.4 *The main stages of a research process*

numerous fundamental aspects of research. Consequently, we strongly encourage any empirically interested beginners to conduct studies of their own as soon as possible to complement the insights available here.

The main stages of empirical research projects are depicted in Figure 1.4. These stages basically follow on from one another as indicated by the arrows. Nevertheless, information acquired at later stages may lead to reconsidering aspects of a prior stage (e.g., a specification, a modification, or even a complete rescheduling of a research issue or an empirical procedure). At a certain stage (of data collection and/or analysis), however, replanning the research project may no longer be feasible – generally for practical reasons, such as limited time or financial resources. In that case, such considerations can only be discussed in the post-study reflections and appropriate recommendations can be made for subsequent studies.

1.1.4 Research Components and Basic Classifications of Research ■

The basic components of empirical research are:

- the **researcher** (who): This is the person who plans and conducts the empirical investigation. It might be a single person or a group of multiple researchers working together on a project. In the case of linguistic studies, the researchers are generally linguists. Depending on the performed activity, they can be called *observers*, *interviewers*, *experimenters*, *analysts*, and so on. An unpopular moniker in linguistics is ‘armchair linguist’ – a term that is used by corpus linguists (who study natural language data) to denote linguists working with conveniently generated language data. Field linguists studying a language by direct participation in the life of its speakers, in turn,

can label all linguists who study a language outside of this context ‘armchair linguists’.

- the **research issue** (what): This is the research topic, or more precisely the research question or hypothesis to be answered or tested by the empirical investigation. In the case of linguistic studies, this pertains to a linguistic topic of any area (e.g., phonology, syntax, pragmatics, or language contact). Alongside the linguistic area, we can distinguish between research on language production or use (speaking/signing and writing) and research on language perception (listening and reading). Another basic distinction is between topics in applied linguistics (the study of language in relation to practical application or real-life issues, such as language teaching, forensic linguistics, translation, and language software) and topics in academic core disciplines of linguistics (as addressed in Part II). Thus, the **object of research** can be past and present languages and their varieties (see temporal research framework: diachronic vs. synchronic language/dialect studies) – either single languages (studies on individual languages/varieties) or more than one language, as in comparative/cross-linguistic studies (cf. Section 1.2.8), contrastive studies (cf. Section 1.2.8), or contact studies. A further distinction is made between languages and language varieties (dialects, registers, sociolects, etc.) with different numbers of speakers, degrees of diffusion, and levels of research (e.g., better researched or major vs. minor or less studied/unstudied languages or varieties).
- the **research aim** (why): The general aim is to answer the research question and thus to contribute to knowledge growth, and, in the specific case of linguistic research, to the understanding of language in all its aspects.
- the **research design** (how): This is the empirical approach and the methodological procedure of how the research issue is to be investigated. In linguistics, this differs fundamentally in the distinct sub-disciplines, as we will make clear in Part II.
- the **research participants/subjects** (whom/who): As language is produced by humans, its speakers, writers, hearers, or readers can be regarded as data sources, or, more specifically, data providers or suppliers. Depending on the method of data collection, they are also called *informants*, *interlocutors*, or *respondents* in survey studies. In linguistic studies, these are people with the respective language skills as relevant for the research topic – that is, mainly native speakers but also certain subgroups (e.g., children – for the study of first language acquisition; certain bilinguals – for the study of contact phenomena, such as code-switching) or learner groups (for the study of foreign language acquisition). The researcher is generally not a research participant, but if so, it is then in the context of *introspective research* (cf. Section 1.2.5). The number of research participants can vary

between one person (*single case studies*) and much larger numbers (cf. Section 1.2.3).

- the **research data**: This is the needed data that is to be analysed. In linguistics, this is *language data* and/or *data from language-related tasks* (such as acceptability judgements or reaction time measures), depending on the research topic. It can be collected or compiled within the empirical study or it can already be available from previous investigations. Thus, we distinguish between *primary data* (i.e., data collected by the researcher for the specific research issue), *secondary data* (i.e., data that was collected by someone else for another purpose but is usable for the researcher's own analysis), and *tertiary data* (i.e., convenient data that is already edited/processed by someone else and can be used for further analysis). Furthermore, *natural language data* can be distinguished from *language data, which can be generated explicitly by/for linguistic research*. The transitions between these types of data are fluid (e.g., the compilation of primary data that already exists or quasi-natural language data resulting from conversations initiated by the researcher) and different combinations are possible (e.g., natural speech vs. elicited primary data). A final distinction pertains to the ethical treatment of data: *confidential data* (i.e., the research participants are guaranteed that nobody other than the researcher(s) can know their identity) versus *anonymous data* (i.e., not even the researcher knows the identity of the research participants).
- the **research location** (where): This is the environment in which the research or more precisely the data collection takes place. A basic distinction in linguistics is made between *field research* (i.e., research in a natural surrounding with numerous interacting variables) and *laboratory studies* (i.e., research in an artificially controlled environment with a reduced number of variables). Depending on the object of research, field studies can take place in different locations, while most laboratories are in industrialised Western cities with the corresponding infrastructure. However, mobile laboratory equipment allows for a combined research environment. Secondary or tertiary data-based research takes place in an *office environment*; this may include the consultation of archives and libraries.
- the **temporal research framework** (when): This is the time in which the research is carried out. Altogether, it encompasses the time frame starting from the research-planning phase up until the publication of the empirical findings and, in particular, the time devoted to data collection. The data can either be collected at a certain point in time such as in *synchronic research* (i.e., studies on language at a specific point in time as is, e.g., generally the case in language typology) or in *cross-sectional studies* (i.e., investigations of a population at a specific point in time – e.g., comparing different age groups) or repeatedly over a longer period such as in *diachronic research*

(i.e., studies on language as it develops over time, e.g., in historical linguistics) or in *longitudinal studies* (i.e., investigations of the same individual(s) over time – e.g., observing their change).

Other than the mentioned **kinds of research** resulting from basic distinctions regarding a basic component (field vs. laboratory research, diachronic vs. synchronic research, monolingual vs. comparative or multilingual research, etc.), there is a further fundamental classification of empirical research relating to the research question/hypothesis (cf. Section 1.2.1), as well as to the data and the research design:

- **qualitative studies:** Typically, this kind of research aims at textual descriptions of a complex research issue (multiple interacting variables in natural situations) from a personal perspective (subject-related). In order to capture the multiple interacting parameters, qualitative studies are generally based on natural language data obtained from small samples (see Section 1.2.3) or even single cases, which is then analysed interpretatively (in-depth analysis). A potential critique is to what extent the data is representative and generalizable, but this is generally not the aim of case studies.
- **quantitative studies:** Typically, this kind of research aims at numeric presentations (object-related) of a restricted research issue (a limited number of focused measurable variables under standardised conditions). Other than measurable/countable units, quantitative studies require certain methodological procedures (large samples of standardised data) and analytical tools (statistical evaluation). A potential critique is whether the collected data is realistic and corresponds to natural data.

Table 1.1 gives an overview of the linguistic subdisciplines as presented in Section II and indicates the kinds of research this book deals with.

Table 1.1 *Basic kinds of research per linguistic subdiscipline*

	Basic kinds of research
Language documentation & descriptive linguistics	• Field research, collection and analysis of primary language data
Language typology	• Comparative/cross-linguistic research, usually by use of secondary 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 natural language data or typological findings
Neurolinguistics	• Laboratory research

1.1.5 Quality Criteria

Empirical research is based on general quality criteria that need to be met in order to fulfil scientific standards and ensure that one's research findings will be taken seriously. The three fundamental quality criteria are:

1. **objectivity** (or neutrality):
 This criterion strives for the independence of the research results from the researcher or any person involved in conducting the research (analysists, data compilers, etc.). This means that the same research conducted by other researchers should deliver the same results. However, absolute objectivity is never a fully achievable ideal. Observations, for instance, are generally more or less shaped by the observers' perspective and even just by their presence – regardless of whether they are aware of their influence or not (cf. Section 2.2). Therefore, *intersubjectivity* can be considered a more realistic aim. It implies that a described fact is equally recognisable and *transparent* (or comprehensible) not only for the researcher but also for various other people. In this way, intersubjectivity is contrasted with subjectivity.
2. **reliability** (or dependability):
 This criterion aims at the exactness of the collected data, i.e., that the research method produces consistent results that are reproducible with the same methods under the same conditions (replicability). The exact repetition of research processes, however, is more realistic in quantitative and/or laboratory studies than in qualitative and/or field research in which it is only possible to find similar/comparable conditions but not identical ones. Reliability can be tested by:
 - the exact repetition of the research (where possible) – by the same or even different researchers (e.g., *inter-observer reliability*);
 - parallel studies, i.e., the same group of research participants (sample, cf. Section 1.2.3) is investigated twice (if the first investigation does not affect the second one) or two or more comparable groups are investigated by use of the same methodological procedure;
 - split half of the research results, i.e., the data is divided in halves, then each half is evaluated/analysed separately, and the two sets of results are compared.
3. **validity** (or trustworthiness, credibility):
 This criterion means that the research procedure actually needs to measure what is intended and that the research successfully answers the research question. A distinction can be made between *internal validity*, i.e., the controllability of parameters to rule out alternative influencing factors, and *external validity*, i.e., the generalisability of research outcomes. The latter means applying the conclusions of an empirical study outside its proper context/scope of investigation such as to other research locations (*ecological validity*: applicability of laboratory findings to natural settings), to other people (*population*

validity), or over time (*historical validity*). Validity has to be tested regarding a multitude of research components, such as the representativeness of the selected research population, the appropriateness of the used method of data collection and of the applied evaluation procedure, the plausibility of the conclusions (generalisations/interpretations), and so on. First of all, we can ask ourselves, if the research results are plausible. And then we can rule out any kind of research bias. Furthermore, the methodology has to be sensitive enough to show robust results. The *scope of application* (or applicability) describes under which circumstances research results are valid. The data may be representative for a certain subpopulation or only under particular circumstances. Thus, careful consideration has to be given to the extent to which research outcomes are generalizable.

Conversely, several fundamental issues can have a negative impact on research quality. Good scientific practice includes the **researcher's awareness of their own role in the research process**. Therefore, we will point out some effects that the researcher may cause unintentionally:

- **personal bias:** These are distorted research outcomes that result from the researchers' subjective viewpoint (selective perception) – their personal beliefs, thoughts, expectations, feelings, or attitudes (see, for instance, Bargh, Chen & Burrows 1996). As people are often not aware of themselves and take their own perspective for granted (cf. exercise in Section 2.2.4), researchers need to be conscious and critical of their self-perception via self-reflection. They should ask themselves questions like:
 - Do I have expectations or a certain attitude (such as personal preference or animosity) regarding the research topic? Does this become apparent in my behaviour or empirical design (e.g., questions)?
 - Do the research parameters reflect only my own speaker-, language-, or culture-specific categories that may be inappropriate for the research participants? Do I interpret the data in a way that is also appropriate given the categories of the research participants?
- **reciprocal effects:** These are (unwanted) research effects that result purely from the researcher's presence, i.e., the researcher's presence leads to certain behaviours on the part of the research participants, which would not have occurred or would have occurred differently in the case of the researcher's absence (e.g., actions of hospitality and politeness vis-à-vis the researchers or restrained and sceptical behaviour in their presence). The *observer's paradox* (cf. Section 2.2.3), for instance, acknowledges response effects of the observer's presence. However, this holds true for any kind of research interaction, including observations, surveys, and experiments (cf. Sections 2.3.3 and 2.4.3). The use of media, in particular, may cause or intensify