Palaver
Palaver 10 (2021), n. 1, 47-74
e-ISSN 2280-4250
DOI 10.1285/i22804250v10i1p47

Lumnije Jusufi
HU Berlin

A critical view on a classical approach: Empirical data collection in linguistics

Abstract
This paper takes both a diachronic and a synchronic perspective on the empirical approach while exploring the emerging field of Albanology. While still applying methods elaborated during the 1960s, Albanian dialectology is considered widely researched. Other linguistic areas have focused on idiomatic written language, mainly that of print media (e.g., belletristic literature). Precisely because of this, in my view, spoken forms of Albanian have received only little attention. Empiricism in linguistics is as old as the field itself, but with that tradition comes an almost stubborn limitation to only a couple of methods applied. This needs to change, as empiricism is now present in all areas of science and has been revolutionized by digitalization.

Keywords: *Fieldwork; linguistics; transcription; archiving; albanology.*

0. Introduction

Empirical research in the humanities and social sciences is positioned between two poles. While it represents a "primary method" in anthropology and sociology (Beer 2001: 11), it does not have any greater importance for mor traditional linguistic branches focusing on the language system, sometimes even being characterized as nothing but a data collection method (Vaux, Cooper 2003: 5). However, the empirical approach in linguistics is as old as the discipline itself. While dialectology is probably the most famous linguistic branch employing empirical fieldwork methods, structuralist and comparative linguistics need field work and data collection, too. Yet, since they concern the language system, all of these linguistic subfields have a tradition of standing close to the natural sciences, as they are likewise involved in establishing “laws” (Berger et al. 2009: 12) and searching for linguistic regularities. This is also observable in the frequent use of questionnaires, which were intended to meet scholars’ demands for comparability and verifiability, but “without an effort to reconstruct cultural practice in natural contexts and without developing research in interaction with the field” (Spranz-Fogasy, Deppermann 2001: 1007). Moreover, dialectology has long been concerned only with collecting “relatively small sections of lexical or morphological information” (Labov 1978: 200). The other extreme in linguistics, research on the language system and comparative linguistics, with their main goal of making even completely unknown languages accessible to scholars, also adopted field research methods from the very beginning or made them the core of data collection methods (Nida 1947). However, both subdisciplines are mainly interested in language as a system of rules, and less in actual language use.

The sociologist Basil Bernstein and the linguist William Labov laid the groundwork for research in sociolinguistics and empirical analysis during the 1960s and 1970s. Subsequently, linguistics had a subfield, prone to use empirical methods like ethnographic participant observation and sociolinguistic interviews, challenging other fields, including anthropology and sociology. William Labov is considered the founder of empirical research in this sociolinguistic sense. He brought linguistics closer to ethnology and thus removed from it scientific, lawseeking methods, although he still regarded the natural sciences as a model (Labov 1978: 185).

Accordingly, the definition or description of empirical research is very diverse and is based on various criteria. In honor of Gianni Belluscio, who put empiricism at the core of his scientific work, this article aims at taking a closer look at this field while focussing on (albanological) linguistics and pointing out difficulties as well as possible solutions.

1. The Fieldwork

Field research comprises three phases: Planning, data collection and data evaluation. The planning phase includes the following steps: Formulation of the research objective and the research question(s), selection of the field research area, theoretical references, literature research and language acquisition. The data collection phase, i.e., the actual field research, comprises two phases: the exploratory or discovery phase and the problem-oriented phase (Beer 2001: 23). The final phase, data evaluation, begins after the fieldwork with the analysis of the collected data material.

The main characteristic of empirical research is the *field* (< *field* research) in which the data is collected, also known as *lifeworlds*. With changing scientific interests, the conception of the field has also significantly changed. While in the early days a field used to be characterized by, for example, a remote, barely known region of Latin America, it could just as well be a big city or company today. The selection of the field research region is closely linked to the formulation of the objective and the research question. Empirical studies are always case studies and do not aim at representativeness (Werlen 1996), as the researcher takes an inductive approach. These consideration determine the according preparation steps.

Acquisition of the target region language is sometimes seen as part of the preparation phase, but it could also become part of the implementation phase. As a rule, the acquisition of a dialect or vernacular can only take place in the second phase. Mastering neighboring dialects could significantly ease access to dialectal fieldwork. Although such linguistic knowledge is central to empiricism, reports on field research methods, as well as some isolated studies, often contain references to translators, which implies that the field researcher does not have, and presumably doesn not need, an adequate (or even any) level of proficiency in the language spoken in the given field. Although such reported translation activity mainly relates to matters outside of systematic linguistics, there are also some references to translation being employed within the domains of general linguistics concerned with the language system itself. Furthermore, having translators as part of the field team is an integral part of semantic field research, i.e., in cases where the researcher does not aim at collecting data on lexicological issues or the like, but rather searches for the meaning of words and expressions. In these cases, operating with the concept of the so-called meta-language (for example, English) and according translations in the narrowest sense (Matthewson 2004: 378-393), is a regular. The field researcher is in such cases, however, entirely depending on the translator, with often problematic effects, as we will see below.

Empirical research can be carried out quantitatively as well as qualitatively. Qualitative research covers all types of data where the content of the information is the main focus, whereas quantitative research focuses on the volume of information. The two methodological directions are not mutually exclusive, but complement each other (Röbken ²2016: 12). Both directions comprise quality criteria to be considered when planning the field research. These criteria include reliability, objectivity and validity (Albert, Marx 2010: 28ff.).

Disturbance factors to be considered may either relate to the situation or to individual characteristics of the informant. The researcher needs to ensure that these disturbances do not occur during data collection. Possible factors include limitations of the informant due to health (such as deficient articulation) or mental factors (such as stressful situations). The researcher can try to minimize these obstacles in the selection of informants. Situation-related disturbances can include very long questionnaires, which could lead to fatigue and lack of concentration in the informant. Group conversation may also lead to a disrupted data flow. Although this survey method allows for the collection of natural conversation sequences, it can also disturb the flow if, for example, speaker changes are not well maintened and too much parallel conversation takes place (Vaux, Cooper 2003: 8f.). The researcher also has to pay attention to the correct timing of the survey (regarding diurnal as well as week and annual rhythms). Loud noise could be another situation-related disturbance to consider. Closed spaces, such as rooms in private houses, offices or (switched off) cars[1], however, are ideal interview locations. The observation itself may also be a disturbance factor. This phenomenon, known as the *Hawthorne effect* (Albert, Marx 2010: 36), results from the observation situation and its potential effect on the informant's

[1] In the summer of 2003, as a student in a field research internship at the Universitá della Calabria in Italy, the author received this tip from the mentor of the internship, Gianni Belluscio, to whom the author is still grateful today. Her best audio recordings were taken in a car.

behavior and thus on the naturalness of their speech production. This effect may range from minor to severe influence. The researcher can apply various techniques to minimize, if not completely eliminate, such effects. Labov's well-known concept of the *observer's paradox* (1978: 192, 200), is worth mentioning in that context, which led to efforts of observing informants as unobtrusively as possible.

Five functional persons may be involved in field research: The field researcher, an accompanying guide, the informant, and possibly a translator and a field research assistant.

The researcher is the initiator, organizer or moderator of the field research. As we saw above, he or she has a responsibility to ensure favorable conditions for the survey and the sense of well-being of everyone involved in the process. The researcher is both a guest and a host in the field. In his "10 commandments of field research", the sociologist and cultural anthropologist Roland Girtler describes the field researcher as a "witness of foreign life", who should be neither judge, social worker, nor missionary (Girtler 2001: 183-197). On the one hand, the researcher should observe and experience their object of investigation – language and/or culture – and the related region "not from the outside, but from the inner perspective". (Girtler 2001: 183). Yet, on the other hand, they must maintain an objective distance from their object of investigation. When the researcher has such a deep involvement during their data collection, objectivity might not be guaranteed. Nor can the influence of the field researcher be completely eliminated (Fischer 1992). Researchers have used the ethnological concept of *self-reflection* in methodical attempts to control these influences and compensate for their proximity to their research object, and to place themselves in the relevant situation or informant's position (Linska 2012). As part of the discussion about the internal and the external perspective, the influence of the researcher's place of origin is also discussed, searching for answers to the question of how beneficial it is to conduct research on one's own language and culture. This approach originates from the fields of ethnology and general linguistics (Samarin 1967: 20). Even today, both fields still aspire to unlock foreign and unknown cultures and languages. This question has reached linguistics with the emergence of sociolinguistics. It is, quite rightly, an attempt at controlling the linguist's subjectivity both in the field and in the scientific interpretation of the data. This approach, if carried to its logical end, makes all national philologies, which are run mainly by "internal" experts, superfluous. Yet, here too, science has gained from method development. Just as "foreign" researchers have to work out their view from the inside and scientifically make up for the resulting proximity through self-reflection, "internal" researchers have to work out the view from the outside and prove their objectivity and sufficient distance through self-reflection. The latter is always "suspected" of being subjective from the beginning (Hirschauer, Amann 1997). However, the scope of the "internal" are not clearly defined in the relevant research. There is much debate over identity affiliations, as is the case with migrant groups. The further a researcher works and lives from the field, the more they may be attributed with their region and culture of origin. In this regard, the field researcher has yet another bridge to cross – between the suspicion of the supposedly own and the actually foreign, even if this is very narrowly defined.

The second main actor, recruited by the field researcher, is the informant. This person is the source of data or information

(Samarin 1967: 1). Careful consideration needs to be given to the characteristics of the informants when deciding on the field research methods. Especially in quantitative research, one has to consider the number of informants, their gender, age and level of education, as well as the question of origin and migration, or, in linguistics, the aspect of their mother tongue. There are numerous definitions of the optimal informant, but, unfortunately, the required attributes are unlikely to be written on their foreheads. It often takes at least one long conversation to determine whether a person is suitable for research or not. It requires good people skills to smoothly get out of such situations again[2]. The informant's proficiency in the language and dialect is well-discussed in the previously mentioned considerations on empirical data collection methods. It is important that the informant is a native speaker of the language that will be examined (Samarin 1967: 35). Another requirement is that they are monolingual (Vaux, Cooper 2003: 8). The first requirement is obvious and easily achievable. The second, on the other hand, faces many difficulties in the age of digitization and globalization, not just in the field of contact linguistics and migration research. When it comes to the study of dialects, the requirement of monolingualism is almost impossible to meet. In times of mandatory schooling, people, particularly those in Europe, speak at least their own standard language in parallel with their dialect. There is no general consensus on the ideal number of informants for effective data collection – practice ranges from one, over several, to as many as possible. However, researchers tend to use as many informants as possible for quantitative research purposes, to guarantee representativeness,

[2] Thede Kahl reports about such "human bottlenecks" of the researcher's access to the field (Kahl 2008).

while, for qualitative purposes, they prefer as few informants as possible, even only one[3]. Recommendations to use one informant only are often found in methodologies of general and comparative linguistics that focus on drawing up the complete grammar of a little-known language. In such cases, however, only a few informants are usually available anyway. Sparsely populated regions also pose informant-related problems (Samarin 1967: 27ff.). When selecting informants, both sexes and as many age groups as possible should be included (Samarin 1967: 22). The gender of an informant could lead to various problems. On the one hand, depending on the researcher's gender, access to one of the two sexes might be culturally denied. On the other hand, personal sensitivities could be a cause. Previous field research methodologies have recognized this issue and proposed to start with an informant of the same sex (Dimmendaal 2001: 60).

Furthermore, field research by one person alone, especially in patriarchal societies and rural areas and particularly if undertaken by a female researcher, is not only difficult in human terms, but also professionally often problematic if not impossible. For this reason, it is advisable to be accompanied by someone of the opposite sex. Ideally, this companion should be a colleague, or else, a person from one's private environment. The companion must only play the role of an alibi and should not interfere in the empirical events. If the person is also a specialist on the relevant theme, they should rather carry out their own survey separately. The companion role is not mentioned in any previously existing empirical research theories.

[3] Sometimes there is also talk of three main informants (Dinmmendaal 2001: 60).

A further companion can or must be involved when there is a lack of language and/or dialectal knowledge. This person is the translator, as mentioned at the beginning. There are many known cases from the beginning of the 19th century in which researchers used translators to study completely unknown cultures and languages. Initially, the researchers did not speak the studied language. They learned it, albeit only passively, during the course of their research. Therefore, working with a translator, next to gesturing, was unavoidable (Reichertz 1992). The early days of (foreign language) Albanian studies were also characterized by the field research role of the translator. It is much more common to find a "translator" in the area of dialectal studies. This person could "translate" from the standard language into the dialect. By including the translator, the originally linear face-to-face communication between researcher and informant transforms into a more complex matter, consisting of three directions: researcher-translator, translator-informant, and researcher-informant. Error checking or contextualization of errors also requires multidimensional communication directions. The translator represents a kind of *black box* to both the researcher and the informant. This is why this person's role in field research is deemed highly questionable today and should be avoided if possible.

The final field research actor – the field research assistant – is introduced here based on the author's own experience. Literature on field research methodologies contains approaches of this kind, but not assigned to one person. This role involves a mixture of translator, companion and trained informant. Informant training is often mentioned in field research methodologies for general and comparative linguistics. This training is for those employed as informants over a more extended period (Nida 1947: 140; Samarin 1967: 41f.; Dimmendaal 2001: 62). Trained informants participate in qualitative surveys, especially concerning the language system and semantics. Here, however, the informant's influence needs to be critically reviewed. From an ethnological point of view, such a method is highly questionable and inappropriate. The training creates a lot of intervention into the informant's language behavior. Elsewhere, when searching for informants, it is recommended to give preference to well-educated people with linguistic knowledge, if possible. These, too, may not be entirely appropriate when it comes to language ideologies and dialectal data collection (Vaux, Cooper 2003: 10). All these needs and experiences have given rise to a new participant in field research. This person is the companion and field research assistant, who could be an educated person, on a linguistic level (most likely a teacher) (Vaux, Cooper 2003: 9). This assistant can be made familiar with survey techniques initially, and can help with interviews and selection, as well as finding informants or, at the researcher's request, independently carry out smaller supplementary and control surveys. It is advisable to select more than one assistant, at least representing both sexes.

Empirical language data collection methods can be divided into four groups: Observation, interviews, experiments and translation. In linguistics, Labov was the first to introduce a strict separation between these methods in the 1970s. In earlier times, there was a firm reliance on using questionnaires in empirically oriented linguistics, especially in dialectology. These questionnaires were also used in combination with interviews, such as when answering the questionnaire orally and recording what was said. This method later became known as the targeted interview.

Observation is one of the most important qualitative empirical analysis methods, while at the same time one of the most difficult. Regional research cannot avoid this procedure. Merely staying in the region that is to be examined implies observation as an empirical survey procedure. The generic term *observation* comprises various processes, which can be divided into three levels: participant vs. non-participant, structured vs. unstructured, and open vs. hidden (Werlen 1996: 751). The first level refers to the presence or absence of the researching person, whereby in the humanities, the researcher's presence and subjective perception form an integral part of the empirical data collection. Participant observation, also called ethnography (Fischer 1992) in English and American discussions, represents, as repeatedly mentioned before, the core of ethnological and anthropological empiricism, and it was not considered a typical linguistic method before Labov (1978: 195ff). In dialectology, especially in traditional and nationalized approaches, this form of observation found no resonance for a very long time (Werlen 1984). However, for sociolinguistic investigations, especially into language settings and ideologies, this method is of central importance, at least as a control mechanism in checking for the Hawthorne effect. Meanwhile, participant observation has become the core method used in empirical data collection in some linguistic areas, such as contact linguistics (Werlen 1996) and conversational linguistics (Sprance-Fogasy, Deppermann 2001). This type of observation entails researchers collecting data while participating in the daily practices of the observed group as long as possible (permanently or temporarily) to become familiarised with their way of life (Lüders 2003: 384f.). It forms the core of qualitative research, in that it allows to find data and reflect on interpretation processes. Criticism of the method pertains to the subjective forms of experience, which should, nevertheless, be absorbed by the concept of self-reflection. Participant observation is not suitable for large amounts of data, large language communities, precise investigations, accurate measurements or "rapid" research (Werlen 1996: 751ff.). The second level relates to the criteria for and framework of data collection. Unstructured observation contains a rough framework and only few criteria. The researcher is more flexible and open to their subject. Such a procedure is well suited for the early stages of field research. Structured or systematic (focused or selective) observation is applied to a specific situation or a special event, e.g. in a ritual that follows a fixed scheme (Labov 1978: 200). This observation method is well suited for advanced data collection. For linguistic purposes, this may involve certain moments of interaction, such as during the hustle and bustle in the schoolyard, a lesson, or a sporting or political event. Systematic observation can also involve watching a television program or comprehensively monitoring radio and television broadcasts. At both levels, the researcher can participate actively or passively. The active role demands more open observation. The passive role, on the other hand, strongly favors covert observation. This refers to the third and final level of observation procedures, namely openness vs. masking of the observation (the observing personnel or equipment). Open observation is the most common method in which all involved participants are informed in advance, and their consent is obtained. In doing so, one has to accept that the informants will change their behavior as the observation influences them. In contrast, during a covert observation, the informants are not informed. This process promises more authentic data, but is critically discussed in terms of ethical standards. However, a foreign researcher can never really work in a completely concealed manner. The mere presence of a stranger makes locals change their behavior. Covert or concealed observation also means holding the recording device in a way that does not constantly remind the informant of the observation. But that can lead informants to reveal confidential information that they would not like to make public. One should, therefore, distinguish between a completely hidden and a partially hidden observation. However, strictly speaking, no observation can or may remain completely hidden, solely for moral, legal and temporal reasons (Gehrau ^{2}2017: 33f.). Participant observation should never be hidden, even if, over time, it might gain elements of a hidden survey. Systematic observation may be hidden, but the participants should be informed later. However, it remains a controversial practice. Unmediated vs. mediated observation takes place on both levels, whether the observed person is recorded by a fixation medium (recording device) or not (Gehrau ^{2}2017: 46ff.). In the case of unmediated observation, at most, notes can be taken, preferably after the observation. Ethnologists need to keep an empirical diary, which will form an integral part of their field research. This practice can also benefit linguists (Beer 2001: 23).

There are four types of survey procedures: Interviews, questionnaires, experiments and translation.

Although Labov (1978: 203) revolutionized linguistic elicitation with the introduction of participant observation, he still sees the (individual) interview as the core of linguistic data collection. There are two distinct interview types: the open interview or free conversation and the closed or targeted interview. The free conversation is particularly fitting for the exploratory phase. Suitable for this are topics that are nonlinguistic or private information relating to the informants. The researcher is as reticent as possible, and the interview process is left to the informant. Löffler distinguishes between three possible constellations: 1) free history or a narrative about an experience (so-called *narrative interviews*, Schütze 1983) or about something in their life (so-called *biographical interviews*, Schütze 1983), 2) conversation with acquaintances or relatives and 3) interview with the researcher. A free conversation is usually very short and should always be recorded. The Hawthorne effect, due to the microphone, is only there initially and later compensated for by a relaxed conversation atmosphere (Löffler 2003: 49f.). The targeted interview attempts to collect language data "without regard to conventional and situational conditions" (Löffler 2003: 48). This can be achieved through a list of questions that are run over individually in the course of the conversation. The researcher's way of expression is particularly important here, as it needs to avoid "echoing forms" by the interviewee (Löffler 2003: 48). Open questions are to be preferred in order to avoid monosyllabic yes- and no-answers. Disadvantages of this type of interview are its artificiality and a heightened risk of the Hawthorne effect. These interviews can take a long time. Therefore, written notes are advisable (Löffler 2003: 48f.). This interview-type is a mixture between an actual interview and a questionnaire.

The questionnaire is the oldest and most well-known survey method in linguistics in general and dialectology in particular. This process always takes place in written form, and therefore requires readers to be literate. In South-Eastern Europe, this causes a problem when including older people, as they may not be able to read and write. Therefore, this procedure could evolve into a targeted interview. Two types of survey are distinguished:

the direct type in which the researcher is personally involved, and the indirect type – probably the most well-known type in dialectology [think of the famous Wenker-phrases (Ger. Wenker-Sätze)[4]], in which researchers send their questionnaires by post or commission someone to do so (Löffler 2003: 47f.). The researcher's presence allows informants to ask counter questions or ask for clarification, while also forcing them to complete the form up to the end and leave out nothing. In the case of longer questionnaires, there is always a risk that informants will skip some questions and stop reading towards the end.

The experiment as a linguistic survey procedure is becoming more common with the advent of modern technology, although it was already used in Labov's research (Labov 1978: 201). This method is often used in psycholinguistics, but it can also be useful in other areas of linguistics. Methodical approaches may include reading, reading from a computer screen, adding, imitating, working with pictures (a well-established method) and videos (naming pictures, telling stories by means of pictures, naming differences between pictures, describing pictures and videos), naming or writing down word associations, recounting, etc. (Sprance-Fogasy, Deppermann 2001: 1009; Albert, Marx 2010: 96ff.).

The last survey procedure to be mentioned here is translation, which occurs in both general linguistics and dialectology. The disadvantages of this process have been noted several times. It is still reasonable to look at the process itself: Here, informants are asked to "translate" something – either from one language to another (Matthewson 2004: 378ff.; Harris, Voegelin 1953: 70), or from a dialect to the standard language or vice versa. The

[4] Cf.: https://www.regionalsprache.de/wenkerbogen.aspx.

process does, admittedly, help with language production and acquisition, but faces the problem of translation technology. The researcher should, nevertheless, be able to understand both languages or varieties to compensate for possible errors.

Empirical survey procedures are aimed at spoken language. Their main task is to record, fix or preserve fleeting spoken statements, whether natural or provoked by the researcher, for later analysis (Bergmann 1985: 306) or archiving (Newman 1954). This step is an indispensable prerequisite in interpretative research approaches (Bergmann 1985: 300). The data collection and recording (fixation) can be carried out through different media. Both questionnaires and translated interviews produce written fixations, whereas interviews and experiments require technical fixation media. Participant observation only allows for limited fixation, for example, in systematic observation by means of audiovisual fixation. Auditive fixation is more important for linguistics than visual, unless non-verbal communication or language-induced emotionality is being investigated. Audiovisual fixation is regarded as a "privileged form of documentation of verbal interaction" (Sprance-Fogasy, Deppermann 2001: 1009).

The type of method used depends heavily on the phase of field research and the required data. If one is interested in socio-cultural data, such as language use, participant observation is advantageous. If more information about a particular situation is anticipated, systematic observation is suitable. If one needs to collect words, lists (Löffler 2003: 45) or pictures that allow paying particular attention to semantic fields (Dimmendaal 2001: 66) may be a fitting choice. For concrete nouns, gesturing can be used (2003: 46), but to collect verb forms, one should use verbal paradigms, on the basis of filling in incomplete, predefined tables (Dimmendaal 2001: 69). If one is interested in sentences, there is the questionnaire, in which pre-prepared sentences are submitted for translation (Löffler 2003: 45). If one is interested in texts, it becomes more difficult, especially if one is dealing with longer texts. Videos[5] or photo series are useful for longer texts (Harris, Voegelin 1953: 77). While translation has been touted as the best method for longer texts (Harris, Voegelin 1953: 60), it is considered problematical here. Traditional folk heritage, such as fairy tales and songs, are no longer used nor considered a suitable method for language production, as they are seen to be normative in themselves (Labov 1978: 190).

Overall, applying a variety of methods is recommended (Beer 2003: 12). There is no limit to the imaginitive ways in which data can be collected[6]. To transition to the next part of this chapter with the classic Labov: it is precisely the abundance and variety of methods and sources employed that enables to receive "correct answers to difficult questions" (Labov 1978: 207).

2. *Data processing and data archiving*

Data obtained through the use of eliciting can be processed in a variety of scientific ways. These include transcripts for

[5] This method was successfully applied by the author in the project about the border region Dibra. These videos were silent films.

[6] Yet, it is questionable whether stirring up the market women by damaging their sales goods to make them swear, is a morally sound way of getting linguistic data. In this way, it is said that Konstandin Kristoforidhi, linguist and one of the most important representatives of the former Albanian national movement, elicited curses at the market by breaking the eggs of the saleswomen and paying them after they started cursing. Although the story could not be proven, it is too interesting in this context to leave it unmentioned. The author thanks Dr. Pandeli Pani for the reference.

qualitative data (recordings or interviews), statistics for quantitative data and mapping, especially in field studies and dialect geography. Data processing also includes archiving, which is hardly noticed, especially in linguistics. Four different variants are mentioned concerning qualitative data evaluation, namely memory-based, protocol-based, tape-based and transcription-based analysis. For linguistic purposes, however, the transcription-based analysis, which we will focus on in the following, is the most central variant. Because of the technology involved, the scientific study of transcription methods is not very old. However, these methods have now become an integral part of empirical research (Dittmar [3]2009).

Before any transcriptional work proceeds, the researcher oughts to decide what actually needs to be transcribed. This decision is made according to the objective of the analysis, which determines not only the extent of the transcripts, but also the type of transcription: what exactly should the transcription look like, and which transcription system should be applied? In terms of accuracy, Pompino-Marshal distinguishes between two types of transcription, a broad transcription and a narrow transcription (Pompino-Marshal 2000: 746). Dittmar speaks of coarse and fine transcription. The degree of accuracy is determined not only by the phonetic and verbal side of the data, but also, again, by the objective of the analysis. To this end, there are now a wealth of transcription systems available, based on one of two approaches – phonetic or pragmatic. The phonetic approach includes the *International Phonetic Alphabet* (IPA), the Heidelberg *Pidgin Deutsch Lautschrift* (PDL), and the *Speech Assessment Methods Phonetic Alphabet* (SAMPA) (Dittmar [3]2009: 63ff.). The pragmatic approach includes five other systems (Dittmar [3]2009: 81ff.) Which system one decides upon, depends on the research question.

Transcription work is considered very labor-intensive and expensive, but it should not be left to others. On the one hand, it enables the researcher to test their own fieldwork, allowing them to extend, supplement or improve it afterward. On the other hand, a large part of the analysis is already covered and thereby facilitated. If the transcription work is still to be outsourced, the researcher should preferably proofread it in the second round to gain the best possible advantage from the transcription work. At the least, a proofreader is absolutely necessary for linguistic purposes. Dittmar calls the person who improves the first version in consultation with the "initial transcriber" the "transcription proofreader" (Dittmar [3]2009: 217). Such an arrangement provides the best possible reliability process. The intensity of the work also depends on the research question, i.e., how narrow or broad a transcription should be and how much it should be polished. The sociologist Kuckartz ([2]2007) calculates the transcription to take five to ten times the interview time. For linguistic analyses, this time may extent by far. The transcription time specified in the transcription header gives an approximate degree of accuracy. In addition to the transcription time, the transcription header or transcription mask also contains a lot of other information embedding both the recordings and the transcription into different contexts. These information may include the techniques used, the recording and transcription dates, the location of the recording, the length of the conversation, the participants, age and gender of the informants, and, if necessary, the names of the informant, the transcriber and the proofreader. Depending on the research question, further biographical information about the informant can be supplied, for example, information about their profession, origin, etc.[7]

Statistics are useful for processing and describing quantitative data. One should distinguish between descriptive and validation statistics. Here, it might be useful that the linguist uses the support of a statistician. Thanks to advanced technology, data can be displayed very easily via statistical software programs such as SPSS and Excel. The most well-known representations are tables, followed by graphs such as histograms and polygons (Albert, Marx 2010: 122).

The most well-known language cartography is the dialect atlas, representing one of the oldest and most traditional forms of data processing. It was used long before computer technology existed. Traditional dialects form the core of dialect geography. There are two distinct types of language mapping: word and sound geography, the two core areas of dialectology. The advent of digital technology has made it possible for this method to progress into digital language mapping. Efforts to digitalize old language maps and dialects are taken in many languages.

The management and archiving of empirically collected data and the associated transcriptions are not a significant problem in this digital era. Yet, this is merely true for private archives, whose suitability for the scientific handling of these data is limited. Public and institutional archives are more challenging. Large databases offer a possible solution. However, they require institutional deposition and large technical equipment. Smaller languages, such as Albanian, face great difficulties in that regard. Because of such problems, the Academy of Sciences and Arts of Kosovo currently faces the imminent failure of a major

---

[7] Dittmar ([3]2009: 212) summarizes them under “biographic data of the speaker(s)” with much more information than given above.

project to create written language corpora. Corpora of the spoken language, on the other side, are by far more challenging than those of written language, both technically and scientific-ideologically. In Albanian studies, mainly only dialects are understood as spoken language and such corpora are therefore left to the (somewhat out-of-fashion) dialectology.

Now, what kind of publishing and accessibility opportunities are available to the linguist? Apart from public archives, there remain only transcripts, published as attachments to corresponding books, together with the associated recordings on an attached CD or online archiving in databases. However, this option is still very rarely practiced despite extensive digitalization. This is due, on the one hand, to the enormous expense and, on the other, to linguists still being more prone to support printed media. What is more, with more access to one's data comes also more target surface for criticism. Therefore, Labov's critical review of linguists from the 1970s unfortunately still holds, namely:

> The historian tries to present his data to us as directly as possible, while the descriptive linguist keeps us away from theirs. (Labov 1978: 187)

One such initiative for both archiving and accessibility is the LaZAR (*Langzeitarchivierung Regionalforschung* 'Long-term archiving regional research') project[8], led by Thede Kahl and funded by the German Research Foundation (DFG). In addition to accessibility and research options, the project also aims at the long-term archiving of research data. Regarding Albanian, the privately funded online project of the late albanologist Robert Elsie should be mentioned here. Since his death, it has been run

[8] https://lazar.gbv.de/.

by the research department on Balkan Research of the Austrian Academy of Sciences[9].

3. Gianni Belluscio's empirical philosophy

Gianni Belluscio was, first and foremost, a fascinating scholar doing empirical research. I was able to encounter this as he was teaching at Università della Calabria in the early 2000s. Fieldwork in almost all Italo-Albanian villages, e.g., played a central role in his seminars on dialectology in 2003. His philosophy, which he not only taught but rather radiated with his whole being, was that linguistic studies had their beginning in the field and that they stand or fall with the informants. Therefore, he regarded them, the informants, more as colleagues and cooperation partners than as mere sources for his studies. I observed encounters at eye-level with a more camaraderie-style approach in his practice days, which he integrated into his seminars. The two photos below show such a day in southern Calabria in summer 2003.

Moreover, he was among the absolute exceptions in the field of Albanology, as he published his (IPA) transcribed corpus of language. With this practice, he drove many a publisher mad as the transcriptions took up a lot of space and thus caused additional costs. However, what remained unnoticed was that he thus made an outstanding contribution to archiving the languages and varieties threatened by extinction and had the sovereignty to lay everything open to the reader and thereby bind them to himself, as Labov used to say. He lived empiricism.

[9] http://dialects.albanianlanguage.net/index_de.htm; http://www.albanianhistory.net/.

fig. 1 - Gianni Belluscio on field research talking with informants in Vena di Maida in southern Calabria[10]

fig. 2 - Gianni Belluscio on field research with students and informants in Vena di Maida in southern Calabria[11]

[10] The photo is from the author's private archive from summer 2003.

[11] The photo is from the author's private archive from summer 2003.

4. Conclusion

There are no limits to empirical methods because they are both very old and very diverse. Which approach one chooses in the end, depends on the goal of one's work. However, the chosen empirical approaches and methods should be clarified and explained in advance and, above all, used. Therefore, it is a great pity that research methods as such hardly find access to university curricula. The *jump into cold water* may be instructive, but it certainly is not professional. Gianni Belluscio thought differently, as so often. For him, too, it was costly and time-consuming, but all the more valuable for his students.

Bibliografia

1. Albert, Ruth / Marx, Nicole (2010): *Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht*. Tübingen: Narr Verlag.
2. Beer, Bettina (2003): *Einleitung: Feldforschungsmethoden*. In: Beer, Bettina (ed.): *Methoden und Techniken der Feldforschung*. Berlin. 9–32.
3. Berger, Peter / Berrenberg, Jeanne / Fuhrmann, Berit / Seebode, Jochen / Strümpell, Christian (eds.) (2009): *Feldforschung*. Ethnologische Zugänge zu sozialen Wirklichkeiten. Berlin.
4. Bergmann, Jörg R. (1985): *Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit: Aufzeichnungen als Daten der interpretativen Soziologie*. In: Bonß, Wolfgang / Hartmann, Heinz (eds.): *Entzauberte Wissenschaft: Zur Realität und Geltung soziologischer Forschung*. Göttingen: Schwarz. 299–320.

5. Shobhana, Chelliah L. (2001): *The role of text collection and elicitation in linguistic fieldwork*. In: Newman, Paul / Ratliff, Martha (eds.): *Linguistic fieldwork*. Cambridge. 152–165.

6. Dimmendaal, Gerrit J. (2001): *Places and people: field sites and informants*. In: Newman, Paul / Ratliff, Martha (eds.): *Linguistic fieldwork*. Cambridge. 55–75.

7. Dittmar, Norbert (^{3}2009): *Transkription. Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien*. Wiesbaden: VS Verlag für Sozialwissenschaften.

8. Fischer, Hans (1992): *Feldforschung*. In: Fischer, Hans (ed.): *Ethnologie. Einführung und Überblick*. Berlin: Reimer. 79–89.

9. Gehrau, Volker (^{2}2017): *Die Beobachtung als Methode in der Kommunikations- und Medienwissenschaft*. München: UVK Verlagsgesellschaft.

10. Girtler, Roland (2001): *Methoden der Feldforschung*. Wien u.a.: Böhlau.

11. Harris, Zellig S. / Voegelin, Charles F. (1953): *Eliciting in Linguistics*. In: Southwestern Journal of Anthropology Bd. 9, Nr. 1. 59–75.

12. Hirschauer, Stefan / Amann, Klaus (eds.) (1997): *Die Befremdung der eigenen Kultur. Zur ethnographischen Herausforderung soziologischer Empirie*. Frankfurt a. M.: Suhrkamp.

13. Kahl, Thede (2008): *Pleiten, Pech und Pannen auf der Feldforschung*. In: Lindenbauer, Petrea / Sartingen, Kathrin / Šoltys, Jaroslav / Thir, Margit (eds.): *Omaggio a Michael Metzeltin. Von Bukarest bis Santiago de Chile*. Bratislava. 97–110.

14. Kuckartz, Udo (^{2}2007): *Einführung in die computergestützte Analyse quantitativer Daten*. Wiesbaden: Verlag für Sozialwissenschaften.

15. Labov, William (1978): *Einige Prinzipien linguistischer Methodologie*. In: Dittmar, Norbert / Rieck, Bert-Olaf (eds.):

William Labov: Sprache im sozialen Kontext. Bd. 2. Königstein: Athenäum FAT. 184–207.

16. Linska, Marion (2012): *Selbst-/Reflexion in der Kultur- & Sozialanthropologie*. Norderstedt: Books on Demand.
17. Löffler, Heinrich (2003): *Dialektologie: Eine Einführung*. Tübingen: Gunter Narr Verlag.
18. Lüders, Christian (2003): *Beobachten im Feld und Ethnographie*. In: Flick, Uwe / Kardorff, Ernst von / Steinke, Ines (eds.): *Qualitative Forschung. Ein Handbuch*. Berlin: Rowohlt Taschenbuch Verlag. 384–401.
19. Matthewson, Lisa (2004): *On the Methodology of semantic Fieldwork*. In: International Journal of American Linguistics 70, Nr. 4. 369–415. Genauer die Seiten 378–393.
20. Newman, Stanley (1954): *Suggestions on the archiving of linguistic material*. In: International Journal of American Linguistics Bd. 20, Nr. 2. 111–115.
21. Nida, Eugene (1947): *Field techniques in descriptive linguistics*. In: International Journal of American Linguistics Bd. 13, Nr. 3. 138–146.
22. Pompino-Marschall, Bernd (2000): *Transkription*. In: *Metzler Lexikon Sprache*. Glück, Helmut (ed.). Stuttgart, Weimar.
23. Reichertz, Jo (1992): *Beschreiben oder Zeigen – über das Verfahren Ethnographischer Berichte*. In: Soziale Welt 43, (3). 331–350.
24. Röbken, Heinke ([2]2016): *Qualitative und quantitative Forschungsmethoden. Oldenburg*. In: https://www.uni-oldenburg.de/fileadmin/user_upload/c3l/Studiengaenge/BABusinessAdmin/Download/Leseproben/bba_leseprobe_quli_quanti_forschungsmethoden.pdf.
25. Samarin, William J. (1967): *Field Linguistics. A Guide to Linguistic Field Work*. New York u.a.: Holt, Rinehart and Winson.

26. Schütze, Fritz (1983): *Biographieforschung und narratives Interview*. In: Neue Praxis Bd. 13. 283–293.

27. Spranz-Fogasy, Thomas / Deppermann, Arnulf (2001): *Teilnehmende Beobachtung in der Gesprächsanalyse*. In: Brinker, Klaus / Antos, Gerd / Heinemann, Wolfgang / Sager, Sven F.: *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. 2. Halbband. Berlin, New York. 1007–1013.

28. Vaux, Bert / Cooper, Justin (2003): *Introduction to Linguistic Field Methods*. München.

29. Werlen, Erika (1984): *Studien zur Datenerhebung in der Dialektologie*. Wiesbaden: Franz Steiner.

30. Werlen, Erika (1996): *Teilnehmende Beobachtung*. In: Goebl, Hans / Nelde, Peter H. / Stary, Zdenek / Wölck, Wolfgang: *Kontaktlinguistik*. Berlin / New York: Walter de Gruyter. 750–764.

Websites

https://lazar.gbv.de/.

http://dialects.albanianlanguage.net/index_de.htm.

http://www.albanianhistory.net/.