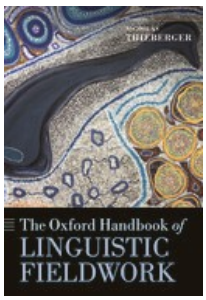

The Oxford Handbook of Linguistic Fieldwork

Nicholas Thieberger (ed.)

https://doi.org/10.1093/oxfordhb/9780199571888.001.0001

Published: 2011 **Online ISBN:** 9780191744112 **Print ISBN:** 9780199571888

CHAPTER

3 Morphosyntactic Analysis in the Field: A Guide to the Guides

Ulrike Mosel

https://doi.org/10.1093/oxfordhb/9780199571888.013.0004 Pages 72–89

Published: 18 September 2012

Abstract

This article focuses on existing field guides for morphosyntactic analysis of previously undescribed languages. The first comprehensive modern linguistic field guide was written by Samarin (1967), followed by Bouquiaux and Thomas, Healey (1975), and Kibrik (1977). Apart from Burling's (1984) small but very useful book, *Learning a Field Language*, there were no fieldwork guides published in the 1980s. Since the early 1990s the rising awareness of the imminent global loss of language diversity, the recognition of linguistic typology as an important discipline of linguistics, and the advances in language recording and processing technology have led to an increasing interest in the documentation of endangered languages and fieldwork methods. These guides have to be distinguished from publications that inform students on the diversity of language structures, or train them through exercises in the analysis of linguistic data. This article briefly explains selected areas of the typology of morphosyntactic structures and gives useful recommendations for further study along with her presentation of data gathering methods. In each section, the morphosyntactic characteristics of Indian languages are explained followed by suggestions about how they can be elicited. The sections in this article recommend a number of books that provide basic and specialized information on languages and language structures and then address anthropologists and other non-linguistic researchers who are interested in collecting language data in the course of fieldwork. Finally, the article summarizes the state of the art.

Keywords: morphosyntactic analysis, language diversity, linguistic typology, endangered languages, language structure

Subject: Linguistic Anthropology, Grammar, Syntax and Morphology, Linguistics

Series: Oxford Handbooks

Collection: Oxford Handbooks Online

3.1 Introduction

This chapter will focus on existing field guides for morphosyntactic analysis of previously undescribed languages. The first comprehensive modern linguistic field guide was written by Samarin (1967), followed by Bouquiaux and Thomas (1976; 1992), Healey (1975), and Kibrik (1977). Apart from Burling's (1984) small but very useful book, *Learning a Field Language*, there were no fieldwork guides published in the 1980s. Since the early 1990s the rising awareness of the imminent global loss of language diversity, the recognition of linguistic typology as an important discipline of linguistics, and the advances in language recording and processing technology have led to an increasing interest in the documentation of endangered languages and fieldwork methods (Himmelmann 1998; Austin 2003–10; Gippert, Himmelmann, and Mosel 2006). In addition to new fieldwork guides (Abbi 2001; Bowern 2008; Crowley 2007; Newman and Ratliff 2001; Vaux, Cooper, and Tucker 2007), fieldwork manuals and questionnaires are now also published as on-line resources in the internet (see the end of this chapter for links).

p. 73 These guides have to be distinguished from publications that inform students on the diversity of language structures, or train them through exercises in the analysis of linguistic data (e.g. Kroeger 2005 and Payne 2006). The borderline between these two kinds of text book, however, is not sharp. Bouquiaux and Thomas (1992: 95–173) devote a whole section of their book to 'concepts of linguistic analysis' before they present their questionnaires, while Bowern (2008: 73–106) briefly explains selected areas of the typology of morphosyntactic structures and gives useful recommendations for further study along with her presentation of data gathering methods. Abbi (2001: 115–220), who focuses on fieldwork in India, structures her chapters on morphology and syntax in a similar way. In each section of these chapters she first explains the morphosyntactic characteristics of Indian languages and then makes suggestions about how they can be elicited. Since the Indian languages belong to different, typologically diverse language families (Indo-European, Dravidian, Sino-Tibetan, Austric, and Andamanese), the book is also useful for fieldworkers in other areas of the world.

The following sections will first recommend a number of books that provide basic and specialized information on languages and language structures (§3.2) and then address anthropologists and other non-linguistic researchers who are interested in collecting language data (§3.3) in the course of fieldwork. The next section (§3.4) discusses the interaction of researchers and indigenous consultants, while §3.5 presents an overview of the various methods of collecting data, i.e. language learning and participant observation (§3.5.1), elicitation (§3.5.2), and the collection of texts (§3.5.3). The last section (§3.6) gives a summary of the state of the art. The term morphosyntactic analysis refers in this chapter to the analysis of the structure and meaning of linguistic units from sentence to word level. It does not presuppose any specialized theoretical approach.

Since the collection of data for the morphosyntactic analysis of a previously undescribed language implies that the researcher spends some time in the field, this chapter will not address language surveys and short-term exploratory field trips.

3.2 Obtaining Information on Languages and Their Morphosyntactic Structures

More than any other linguistic discipline, fieldwork requires intuition and emotional intelligence, but this does not mean that the prospective fieldworker can neglect the study of general linguistic theories or the findings of linguistic typology, which explores the diversity of language structure. All fieldwork guides
p. 74 emphasize the importance of careful preparation. Here we will distinguish between the essential ↳ prerequisites of any kind of linguistic fieldwork (§3.2.1), sources that help fieldworkers to inform themselves on the languages of a particular field site (§3.2.2), and sources that are useful to consult while working on the morphosyntactic analysis of the target language and planning further fieldtrips for supplementary elicitation (§3.2.3).

3.2.1 Basic information on diverse language structures

Many linguistics departments all over the world offer courses on the phonological, grammatical, and semantic analysis of languages from a typological perspective, but there are also departments that focus on linguistic theories and only use secondary sources for examples from 'exotic' (i.e. other than standard European) languages, or departments of regional studies that offer courses on particular languages and cultures without giving the students some background in general linguistics or linguistic typology. It is the latter two groups for whom this section is written. Anthropologists and researchers who are not primarily interested in morphosyntax, but nevertheless want to collect data that are useful for linguists, are referred to §3.3.

Assuming that they have some education in the grammar of English (if not, Biber, Conrad, and Leech 2002, Blake 2008, and McGregor 2009 are recommended for a start), prospective linguistic fieldworkers must be fluent in transcription and skilled in analysing the phonological and grammatical structure of words and sentences of structurally different languages. The more researchers know about the structural diversity of the world's languages, the better they will understand the structure of the target language and the less they will be unconsciously influenced by the contact language or their native language. Good introductions into the theoretical background and the methods of morphosyntactic analysis are found in various students' textbooks. The list below recommends books which also offer ample exercises, and selects those chapters that in my experience are absolutely necessary for a basic training in morphosyntactic analysis.

Morphology:

Aronoff and Fudeman 2005 (chs 1–6; suitable for beginners)

Bauer 2003 (ch. 1–4, 6; suitable for beginners)

Booij 2007 (chs 1–5; difficult for beginners)

Haspelmath 2002 (chs 1, 2, 4, 5; difficult for beginners)

Morphology and syntax:

Kroeger 2005 (chs 1–8, 10–16)

Payne 2006 (chs 1, 2, 4, 5, 7, 8, 10)

3.2.2 Information on the languages of a particular field site

Rice, who has done extensive fieldwork on North American Indian languages recommends: 'As a
p. 75 fieldworker, it is necessary to embrace all sources of material ↳ and learn from them, but at the same time to treat them with necessary scepticism' (Rice 2001: 248). This scepticism is especially warranted when the sources you consult are written with a Eurocentric bias and do not recognize the presence of exotic, unexpected features or the absence of familiar standard-European properties that have been believed to be universal, as Gil (2001) demonstrates.

Preliminary information on about 7,000 particular languages and their distribution in the countries of the world is provided by Ethnologue (http: //www.ethnologue.com), while information on the grammatical structure of languages can be obtained from the *World Atlas of Language Structures* (http: //wals.info). This atlas allows you to search for 141 particular language features and 2,650 languages by language name, language family, region, and country. (Haspelmath et al. 2008; reviewed by several authors in the journal *Linguistic Typology* vol. 13, 2009).

Books in which the basic structures of a representative sample of the languages of the world are described are Comrie (1990a–d), Garry and Rubino (2001), and Brown and Ogilvie (2009), surveys are also found in the Cambridge Language Surveys, the Routledge Language Family, and the Curzon Language Family series.

3.2.3 Information on particular morphosyntactic phenomena

Students who plan to write a reference grammar are recommended to consult:

- encyclopedias and handbooks (Booij, Lehmann, and Mugdan 2000 and 2004; Haspelmath et al. 2001; Spencer and Zwicky 2001);
- textbooks on linguistic typology (Comrie 1989; Givón 2001; Shopen 2007; Song 2001; Whaley 1997);
- articles about individual languages in volumes on specific typological topics such as argument structure (Plank 1985), case (Malchukov and Spencer 2008; Plank 1995), complementation (Dixon and Aikhenvald 2006), coordination (Haspelmath 2004), nominal classification (Senft 2000), or valency (Dixon and Aikhenvald 2000);
- the articles in *The World Atlas of Language Structures Online* (Haspelmath et al. 2008).

3.3 Recommendations for Anthropologists and Other Non-Linguistic Fieldworkers

Researchers who do not have a background in linguistics can contribute to the study of endangered languages when they learn to understand the language and record, transcribe, and translate samples of
p. 76 spoken language with the help of bilingual native ↳ speakers. Burling's small book *Learning a Field Language* (1984) is especially written for researchers who want to learn the language in the field but are not primarily interested in linguistic analysis. He emphasizes 'techniques for learning to understand' and devotes half of his book to this 'neglected aspect of language pedagogy' (p. 6). For both comprehension and production, he recommends that the earliest stages of learning focus on vocabulary and word order without worrying too much about grammatical detail. But the book also offers an excellent, concise description of the essentials of grammar and gives some advice on how learners can advance their grammatical competence (Burling 1984: 55–64, 86–91).

3.4 Linguists and Indigenous Consultants

The quality of the data for the morphosyntactic analysis of a previously unresearched language very much depends on the interaction between the researcher and the native speakers the researcher works with in the fieldwork project. In many publications these people are called 'informants' (Abbi 2001: 57; Samarin 1967: 20), but as 'informant' has the connotation of 'informant to the police', other linguists prefer the term 'language helper' (Crowley 2007: 85f.) or 'consultant' (Bowern 2008: 10; Burling 1984). In the context of linguistic fieldwork, it seems useful to distinguish between consultants and local experts. While the consultants directly help the researcher with the collection and processing of field data, the local experts are people like storytellers, fishermen, healers, or architects who are interviewed on their specialized knowledge.

3.4.1 Selection of consultants

Most fieldwork guides contain a section on 'Selection of informants', 'Choosing language helpers', and the like, which lists the qualifications of the ideal consultant (Samarin 1967: 20–44; Kibrik 1977: 54–6; Vaux, Cooper, and Tucker 2007: 6). However, in practice matters can be quite different. The researchers coming as guests to the speech community are neither in a social position to choose people by themselves nor do they know the people well enough to identify their talents. Consequently they have to ask their hosts or the elders of the community to find the right people. These will have different interests and qualifications, so that it is the researchers' task to adapt to their various talents and accordingly train them in tasks they enjoy and can cope with. (Dimmendaal 2001: 58–66; Grinevald 2003: 67–8; Rice 2001: 245–7).

p. 77 Recommendations like the following are not practicable and may be counterproductive. 'Before hiring the reference speaker, we must test him. One day of work with the candidate will be enough for this purpose' (Bouquiaux and Thomas 1992: 33). Grinevald (2003: 67) warns, 'one should try never to turn away any member of the language community that expresses interest in working on the project…one never knows how things will evolve, and what contribution any particular person can make.' Even semi-speakers or non-speakers may be helpful (Evans 2001).

3.4.2 Training of consultants

The training of consultants has several components and very much depends on their educational background, their standing within the language community, the personal relationship between researcher and consultant, and, of course, their talents and interests. As Healey emphasizes, '[a]lthough there are differences of aptitude from helper to helper, it is nevertheless true that good research assistants are not born, they are trained. And giving this training is one of the major responsibilities of the fieldworker' (1975: 347) With respect to collecting data for a grammatical analysis of the language, this means that linguists should explain what needs to be done for which purpose and train the consultants on the job without any kind of patronizing attitude. Rather, the fieldwork project has to be understood as a joint enterprise in which the researchers from outside and the local experts and consultants share their knowledge and treat each other with the utmost respect.

Typically, linguistic field guides mention the possibility of training native speakers as consultants only in passing (Bowern 2008: 200–201; Bouquiaux and Thomas 1992: 34–5; Healey 1975: 347–9; Samarin 1967: 41–4; Vaux et al. 2007: 29) or completely ignore this important aspect of research (Abbi 2001). There is a series of articles on capacity building in various endangered speech communities in Austin (2004). However, guidelines on how to train consultants in the field do not exist yet. Samarin's remark that 'the

ultimate goal is to get the informant to think about language as the investigator does' and to answer questions in 'the way he should respond' (Samarin 1967: 41) is misleading (see §3.5.2).

In order to get recordings that are as natural as possible (see Himmelmann 1998 on the 'naturalness' of recordings and Samarin 1967: 56–7 on the notion of 'natural speech'), researchers are recommended to train members of the community to do the recordings themselves without the researchers being present at a place where the speaker feels comfortable.

Literate consultants can be trained in doing preliminary transcriptions (Dixon 2010: 322), and literate and bilingual consultants can also be trained in doing translations. Such transcriptions can be done in any kind of orthography—even inconsistent orthographies may be helpful. However, the transcribers must learn to
p. 78 only write down what the speakers have said and not what, in their opinion, they ↳ should have said. Thus all hesitation phenomena, wrong names, code switching, etc. need to be rendered in this preliminary transcription. Transcriptions done by native speakers can be very helpful, especially when they render phonetically reduced and fused forms by their corresponding full forms. They may also reveal the transcriber's metalinguistic intuitions about the morphophonology of the language and the boundaries of linguistic units (Himmelmann 2006b: 254).

If the speakers want the transcriptions to be edited before presenting them to other people, the researchers may give them some advice on how to transform a transcription into a readable text without completely changing its style and adapting it to that of the contact or the written dominant language. Editing texts does not mean rewriting them. Rather, the editors should respect the speaker's way of expression and, for example, remove only those repetitions that are caused by stuttering, or add words and phrases that are absolutely necessary to understand the written text. Nevertheless, some editors will not refrain from changing expressions and will, for example, simplify constructions or make them more complex. Such changes should not be criticized by the researchers but considered as an interesting source for the morphosyntactic analysis of the language, because they show how the same content is expressed by formally different constructions. A parallel corpus of the transcriptions of spoken texts and the corresponding edited versions also provides a new kind of data for the study of the differences between spoken and written language in general. (Bowern 2008: 120–21; Mosel 2006a: 80; 2008; Murray and Rice 1999).

3.5 Collecting Data

The following section reviews a number of fieldwork guides with respect to the methods they recommend for gathering data on the morphology and syntax of the target language. Instead of discussing the strengths and weaknesses of each book in turn, we will scrutinize what they say with respect to the following four kinds of method:

language learning and participant observation;

translational elicitation;

non-translational elicitation;

collection of texts (in the widest sense).

See the chapters in this volume dealing with technology of recording (Margetts and Margetts, Chapter 1), archiving and language data management (Thieberger and Berez, Chapter 4), and experimental elicitation (Majid, Chapter 2).

p. 79

3.5.1 Language learning

Speech communities differ in their attitudes towards outsiders trying to learn their language. Some might not appreciate the researcher's ambitions to speak their language (Crowley 2007: 157; Hill 2006; Mosel 2006a: 73–4). But if they do, the researcher should try to learn the language because being able to communicate with the people

- is a pleasure and helps to cope with the fieldwork situation;
- contributes towards acceptance by the host community;
- brings the researcher into contact with more people in different situations and consequently allows the researcher to gather a more diversified corpus;
- allows the researcher to collect data by participant observation;
- raises the researcher's awareness of certain constructions, when native speakers correct his or her mistakes;
- allows the researcher to use the target language as the mediator language, which will reduce the danger of interference from the lingua franca. (Abbi 2001: 146; Bowern 2008: 9–10; Crowley 2007: 155; Dimmendaal 2001: 72–3; Everett 2001; Hale 2001: 81–2; Kibrik 1977: 52; Samarin 1967: 49–55).

However, with the exception of Healey (1975) and Burling (1984), language learning methods are not discussed in the literature on linguistic fieldwork.

3.5.2 Elicitation

Elicitation means collecting linguistic data by asking native speakers to produce words, phrases, or sentences that can serve as data for the analysis of a particular linguistic phenomenon. Some authors also speak of 'eliciting' texts, but here the term 'elicitation' will exclusively be used in the narrower sense defined above.

3.5.2.1 Questionnaires

For each elicitation session the researcher should have prepared a list of specific questions that he or she wants to ask the consultant in order to obtain data for hitherto unexplored areas of grammar or to clarify problems that have come up when analysing the results of preceding sessions. The collection of data with the help of questionnaires is not a 'mechanical process', as Kibrik (1977: 51) remarks, but should be guided by hypotheses that are based on the findings of linguistic typology and research into language universals and on the analysis of previously collected data. If the new data contradict a hypothesis, this hypothesis needs to be revised and checked against new data, until 'it predicts the construction of new data the investigator has not yet encountered' (p. 51). This routine of making and testing hypotheses can be equally
p. 80 applied to elicited and textual data, but the ↳ literature on linguistic fieldwork and typology only provides questionnaires for elicitation and for analysing and describing the typological profile of the target language:

1. Translational questionnaires consist of lists of words, phrases, and sentences whose translation into the target language is supposed to reveal some grammatical properties of the target language, e.g. '*a chief, the chief, some chiefs, the chiefs, both chiefs, the two chiefs,…*' (Tersis 1992: 277).
2. In scenario questionnaires, the questions first describe a particular scenario and then ask for an expression of a particular content that would be grammatically appropriate in the given context. Dahl's 'TMA questionnaire' is of this kind (1985: 198–206). The first question '[Standing in front of a

house] The house BE BIG,' means that the interviewee should imagine that he/she is standing in front of a house and makes the statement that the house is big.

3. Grammatical structure questionnaires contain questions about the existence of particular grammatical structures, e.g. 'Does the language make any distinction between direct speech and quoted speech?' or 'Are there adjectives that take arguments? optionally/obligatorily?' (Comrie and Smith 1977). A similar checklist is given by Aikhenvald (2007: 63–4) for the analysis of 'word-formation', while other articles in Shopen (2007) unfortunately lack such lists. For some critical remarks on the questionnaire compiled by Comrie and Smith (1977), see Mosel (2006b).

In comparison to translational questionnaires, scenario questionnaires are less likely to produce data that are influenced by the contact language, especially when they are used in the manner and tone of a casual conversation. The grammatical construction questionnaires cannot directly be used for the elicitation of data, but only serve as a checklist for the design of translational or scenario questionnaires or for the analysis of data that have already been gathered from texts and elicitations. Bowern (2008: 214–18) presents 'a basic morphology/syntax checklist' that is 'loosely based on the *Lingua* Questionnaire by Comrie and Smith 1977,' but it looks more like the table of contents of a particular grammar because it lists the terms of grammatical categories, e.g. 'causatives', 'passives', 'copular clauses', 'auxiliary verbs', without any further comments. This kind of list creates the impression that these categories are universal and can be elicited in any language. But passive constructions are, for example, far from being universal; in her survey of 373 languages, Siewierska (2008) found that 211 languages lack a passive construction.

3.5.2.2 Kinds of elicitation techniques

None of the linguistic fieldwork guides systematically describes:

- what kind of questions would trigger what kind of answers;
- what kind of questions would be useful for which area of grammar;

p. 81

- ↳ which areas of grammar would be most efficiently investigated by what kind of questions;
- how different kinds of questions could possibly complement each other.

Samarin (1967: 77) distinguishes between 'translational' and 'non-translational' elicitation, while Bouquiaux and Thomas (1992: 186–397) exclusively rely on translations from the contact language into the target language. Others mix both methods, but often do not explicitly state how the questions are actually to be worded in an interview on a particular grammatical topic, although the way you ask is crucial for the kind of answer you get. In her section on data manipulation, for example, Bowern (2008: 81) simply lists a number of transformations without further comment:

- Turn sentences into questions (and vice versa).
- Manipulate voice and valency possibilities; e.g., active–passive/antipassive... (Bowern 2008: 81)

In order to test if the sentences produced by translations, scenario descriptions or sentence manipulations express the intended meaning, the consultant should always be asked for a translation from the target language into the contact language. Ideally such back-translations are not done on the same day.

A third type of elicitation mentioned in fieldwork textbooks is that the researcher constructs sentences in the target language by him- or herself and then asks native speakers for a so-called 'grammaticality judgement', i.e. telling him or her if the sentences sound right (Abbi 2001: 108, 118; Bowern 2008: 76, 78) This method cannot be recommended because the acceptance or rejection of a sentence created by a non-

native speaker can be based on various, but not necessarily grammatical, reasons (see Bowern 2008: 78–80; Chelliah 2001: 158–61; Mithun 2001: 48; Samarin 1967: 57; Vaux et al. 2007: 278).

3.5.2.3 Translational elicitation

Linguistic fieldwork on a previously unresearched language starts with the compiling of wordlists in order to investigate the sound system of the language and develop a practical orthography. Most fieldwork manuals recommend the translation of wordlists in the contact language and present a sample in their appendices. These samples contain language specific function words like 'at', 'if', or 'in', or words like 'freeze', 'ice', and 'snow' which are only appropriate for certain regions of the world (Abbi 2001: 244–5; Bowern 2008: 223–4; Kibrik 1977: 99–124; Samarin 1967: 220–23). They therefore need to be modified. Function words like prepositions and conjunctions must be removed, and the list of content words should be adapted to the natural environment and culture of the speech community, as done by Abbi in her wordlist 'for Indian concepts' (2001: 246–7).

The manifold problems of translational elicitation which have been identified by several authors, include:

p. 82

- ↳ The consultants may be ashamed when they do not understand the meaning of a word or a construction in the contact language or when they do not remember the translation equivalent in the target language or the concept in question is not expressed by a single word in their language (Bowern 2008: 89; Mosel 2006a: 75).
- The consultant's knowledge of the contact language may not be sufficient to understand all nuances of meaning the linguist wants to have reflected in the translation (Bowern 2008: 89; Chelliah 2001: 157; Samarin 1967: 141).
- The consultant and the researcher speak different varieties of the contact language like Indian and American English (Chelliah 2001: 157).
- The form of the translation may be not idiomatic because of interference from the contact language (Bowern 2008: 85–7; Chelliah 2001: 155).
- If the contact language lacks a grammatical category that exists in the target language, it can hardly be systematically uncovered by translational eliciting (Chelliah 2001: 157).
- In different languages words, grammatical categories and constructions are polysemous in different ways so that there is always the risk that an expression in the contact language or its translation equivalent in the target language are not interpreted in the same way by the researcher and the consultant, especially when the researcher does not describe the context in which the target sentence might be used (Abbi 2001: 88–91; Bowern 2008: 86).

3.5.2.4 Non-translational elicitation

Non-translational elicitation does not generally exclude the use of the contact language, but only avoids the direct translation of single sentences from the contact language into the target language. In order to avoid all the flaws of translational elicitation mentioned above, I recommend a non-translational approach even in the initial stage of fieldwork: first explain what the word list is used for, and then ask the consultants to list any words denoting persons, things, actions, and properties of a particular semantic domain that come to their mind or that you suggest (Mosel 2006a: 75–6). Once the words of a particular domain like food and food preparation have been collected, e.g. 'wash', 'peel', 'cook', 'boil', 'potatoes', 'pot', 'water', 'dirty', 'hot', the 'word-to-text technique' (Samarin 1967: 83) can be applied. This technique requires the native speakers to select a few words from the list and compose short meaningful utterances like 'boil the water!', 'dirty potatoes'. These sentences are then translated by the consultant into the contact language, and can later serve as the basis for the following non-translational techniques, each of which is discussed further below.

1. substitution elicitation (Kibrik 1977: 58; Samarin 1987: 115–17);
2. paraphrasing (Kibrik 1977: 58; Samarin 1967: 119);
3. sentence completion (Samarin 1987: 83);

p. 83

4. ↳ eliciting examples (Kibrik 1977: 58);
5. transformational elicitation (Bowern 2008: 81–2; Kibrik 1977: 60);
6. paradigmatic elicitation (Kibrik 1977: 57–8).

1. The **substitution technique** uses phrases and clauses already elicited in the target language as a frame in which a word or a phrase is substituted for another one. If, for example, you want to know how singular and plural are distinguished or whether there is agreement between certain constituents with respect to number, you may take a simple clause with a singular argument and ask, 'What would you say if it is not only one but several X?' In a similar way, a wide range of morphosyntactic phenomena can be investigated, for example:
    1. the argument structure of clauses by substituting a verb with various other kinds of verbs that presumably require a different argument structure;
    2. gender by substituting nouns denoting males by nouns denoting females and classifiers by using nouns referring to objects of different shapes, sizes, and substances, human beings, and animals as the head of subject and object noun phrases (Abbi 2001: 118, 123);
    3. the tense, aspect, and mood marking by adding or substituting temporal or modal adverbs ('today', 'yesterday', 'never', 'always', etc.) or adverbial phrases ('a long time ago', 'in the future', 'for a week', 'in a week', etc.);
    4. the person marking on verbs or in the verb complex by substituting nominal subjects and objects with 1st and 2nd person pronouns;
    5. complement clauses by substituting a complement taking predicate, for example verbs meaning 'know', 'believe', 'see', 'say', 'want' (Noonan 2007: 149–50).
2. **Paraphrasing** means that the consultants are asked to say 'the same thing in a different way', which may reveal new types of clause structure, if the sentences are 'related by a common meaning and by sharing the fundamental lexical items' (Samarin 1967: 119).

3. With the **sentence completion** technique, the investigator chooses a sentence from the existing corpus, removes parts of it, and asks the consultant to complete the sentence by adding anything that seems appropriate. This technique can, for example, be employed to elicit different types of complement or adverbial clauses, or to investigate the use of tense, aspect, and mood categories in complex sentences (Samarin 1967: 83).

4. The greatest variety of expressions can be elicited by asking the consultants to **create example sentences** for a particular word. Since there is always the danger that created examples sound unnatural, the best results are achieved when two or three native speakers work together. The disadvantage of this method is that the researcher has no control over the kinds of grammatical constructions the native speakers might use. On the other hand, these freely created utterances may reveal unexpected constructions that the researcher would never have thought of.

p. 84

5. ↳ **Transformational elicitation** means that the interviewee is asked to transform one type of construction into another one, e.g. affirmative clauses into negative ones, statements into questions by asking 'what would you say if this was not true?' or 'what would you ask if you did not know that it was X?'

6. The most difficult kind of elicitation is the **elicitation of paradigms**, but it is indispensable for inflecting languages, as even very large corpora do not supply all forms needed for a comprehensive presentation of inflectional paradigms in a reference grammar. I would recommend applying the substitution method first to parts of the paradigm and then, on the basis of the elicited data, explaining to the consultants what a paradigm is. One or the other consultant might then understand the nature of paradigms and be able to construct them by themselves. For example, one could first take a simple clause in the past tense with a 3rd person singular subject from the data, e.g. 'the woman cooked the potatoes', and ask the consultants to substitute 1st and 2nd person singular subjects for 'the woman' to make them aware of the grammatical category of person, then transform the singular arguments into plural arguments to understand how the category of number is formally expressed, and eventually substitute a few other verbs for the verb 'cook'. After several of these past tense paradigms have been completed, one could ask what people would say to express that the woman always, now, or tomorrow would cook potatoes to elicit other tense/aspect categories. Bowern (2008: 89f.) observes that consultants react very differently to paradigm elicitations, and that it might be necessary to do them 'in small batches on different days, combined with other topics'.

3.5.2.5 Shortcomings of elicitation

All fieldworkers agree that elicitation is a useful tool to quickly gather data in a controlled way, but that it cannot serve as the sole empirical basis for the grammatical analysis and description of a language. There are two reasons.

First, by definition, elicitation only provides examples of decontextualized isolated sentences, whereas natural speech is always embedded in the context of a particular speech situation. Consequently, elicited examples cannot show how the meaning of linguistic units is shaped by their context. Furthermore, elicitation fails to uncover any structures that only occur in contexts larger than sentences.

Second, elicitation focuses on certain linguistic phenomena the researchers are particularly interested in, so that they run the risk of missing those grammatical categories and constructions they have been unaware of (Chelliah 2001: 156; Gil 2001: 115). Mithun concludes:

> But if the research is limited to eliciting translations of English vocabulary and syntactic constructions, collecting grammaticalitys and checking off known typological diagnostics, we may

p. 85

> miss what is unexpected about the language under study. In so doing, ↳ we risk depriving the speakers' descendants of what is special about their heritage, and we lose opportunities to expand our own theoretical horizons. (2001: 53)

3.5.3 Text collections

Since 'linguistic elicitation is artificial even under the best circumstances' (Samarin 1967: 59), any grammatical analysis and description should in the first place be based on a good text collection, while the elicitation of data should only be used at the very beginning of the project or as a means of filling gaps in the data as they usually occur in inflectional paradigms (Dixon 2010: 321–2; see §3.5.2.4 above). Correspondingly, the examples illustrating grammatical categories and constructions in a grammar should as much as possible be quotations of naturally occurring utterances (Bright 2007: 16; Mithun 2007a: 59–60, 62–4; Weber 2007: 200). If the corpus does not provide a simple example to illustrate the grammatical phenomenon in question, the grammarian can resort to an elicited example and supplement it by a quotation from the text corpus.

In view of this central role of text collections, the fieldwork manuals contain surprisingly little information on what constitutes a good text collection and how it can be gathered, why it should cover various genres, and what kind of linguistic data can be found in texts of different genres. The fullest accounts of what constitutes a good corpus and what kind of texts it might contain are given by Samarin (1967: 55–68) and Rivierre (1992: 56–63). In the following I will only discuss the content of conventional text collections and how it relates to the morphological and syntactic analysis of the target language; for the technical aspects of recording see Margetts and Margetts (Chapter 1 above), Austin 2006, Schultze-Berndt 2006, and Seifart 2006; for experimental and stimuli-based techniques of recording connected discourse see Majid (Chapter 2 above).

3.5.3.1 Features of a good corpus

Samarin discusses 'six of the outstanding features of a good body of data' (Samarin 1967: 55–68). A good corpus is:

1. 'dialectally uniform';
2. 'natural', i.e. produced and accepted by native speakers as 'appropriate under a given set of circumstances';
3. 'varied', i.e. it would ideally cover all varieties of language that can be attributed to (a) the age, (b) sex, and (c) social class or occupation of the speaker, (d) the emotion at the time of speaking, (e) the speed of utterance, and (f) the topic, (g) type, and (h) style of discourse;
4. 'complete' in that 'all the closed classes of linguistic elements are fully accounted for';

p. 86

5. ↳ 'repetitious' in order to facilitate the identification of the distribution and function of particular grammatical elements;
6. 'interesting', i.e. containing authentic genres and telling something about the culture of the speech community.

The native speakers' use of particular grammatical categories and constructions is determined not only by the structural properties of the language but also by the nature of the particular communicative event, because all languages provide for alternative ways of expression and rules for their contextually appropriate selection. For an introduction to the ethnography of speaking, see Hill (2006), Franchetto (2006), Trudgill

(2000: 81–104). More detailed accounts are given in specialized textbooks on ethnolinguistics, register, and genre such as Saville-Troike (2003) and Biber and Conrad (2009).

With regard to morphosyntactic analysis, Samarin's requirements for a good corpus imply the following recommendations for gathering textual data. First, for a varied corpus one needs to make recordings of several types of text spoken by different kinds of people, such as traditional narratives (epics, legends, etc.), spontaneous narratives (anecdotes, personal histories, etc.), descriptions of activities, descriptions of objects, conversations, etc. (see §3.5.3.2), and as the corpus should also be repetitious, each genre should be represented several times. Since this ideal text corpus, which also has to include transcriptions and translations of all recordings, cannot be gathered in just a few years, the researchers and the speech community have to be selective and set priorities which are not determined by linguistic criteria, but by the practical necessity of recording first what the community considers as most important. If this, for example, is the traditional oral literature or the description of traditional rituals, the corpus may in the end lack casual conversations. On the other hand, the community may want to prioritize the recording of typical everyday communication as the basis for language revitalization measures. Accordingly, a grammar based on such corpora would not cover the full spectrum of verbal interaction, but nevertheless it can be an excellent grammar as long as it is made clear on what kind of data it is based.

Since the speakers' selection of linguistic forms depends on the kind of speech situation, the sources of all texts have to be described by metadata (Austin 2006: 92–4; Bowern 2008: 56–8; Caprile, Rivierre, and Thomas 1992; Himmelmann 2006a: 11–15; Samarin 1967: 102–4; and Thieberger and Berez, Chapter 4 below). Furthermore, in order to allow future researchers to scrutinize the morphosyntactic analysis, all text examples given in the grammar should be retrievable in the text corpus, and the text corpus itself should be accessible. To date these requirements are only met by a very few grammars, for example Thieberger (2006) and Wegener (2008).

Second, to 'get people talking', some fieldworkers use picture prompts like the frog stories as recommended
p. 87 by Bowern (2008: 116); others are more critical, ↳ because what people say when looking at such picture books is neither 'natural' nor 'interesting'. As Foley (2003) demonstrates, it may also differ structurally quite considerably from authentic narratives, and consequently may lead to false generalizations about the grammar of the language.

Third, translated texts of any kind should not be used, unless the researcher is fully aware of translational problems and wants to conduct a specialized investigation on translational interferences from the contact language. Recommendations to use translated material otherwise should not be taken seriously, although they are found in the fieldwork guides. Vaux et al. (2007: 105–7) even suggest inventing texts for an 'informant' to translate and 'tailor the text to fit your own interests as an investigator'. For a critical assessment of the use of Bible translations in morphosyntactic research, see De Vries (2007).

3.5.3.2 Text types

Rivierre (1992) presents a classification of text types that are relevant for linguistic fieldwork. He starts with the distinction between texts of the oral tradition 'with their careful, affected and often even archaic style', including the major traditional genres of historical narratives, myths, poems, etc., and 'more spontaneous texts, such as explanations of techniques, biographical accounts and anecdotes, or conversations in quite a different style which is often neglected' (Rivierre 1992: 56). He recommends collecting texts 'while developing the lexicon, proceeding by means of categories' and distinguishes six thematic categories (pp. 59–61):

1. locations and geographical and social environment;

2. plants and animals;
3. social organization;
4. seasonal and non-seasonal activities;
5. life of the individual;
6. technology.

Each of these categories is further subdivided and accompanied by comments on how these themes and subthemes relate to various genres and how the various kinds of texts can be collected.

Rivierre's categorization is also useful for the creation of a corpus for the grammatical analysis of a previously undocumented language because different themes are talked about in different ways and different genres stimulate the use of different grammatical constructions. In other words, the text type—here defined in terms of themes and genres—determines the frequency of certain grammatical constructions. Consequently, the choice of a particular text type can help to avoid artificial elicitation. Here are a few examples from my own research:

- In order to investigate how the use of temporal expressions differs in accounts of habitual activities and narratives of past events, I recorded and analysed the ↳ description of how to butcher a rooster, and an anecdote telling how someone butchered a particular rooster.

p. 88

- Instrumental constructions such as prepositional phrases or applicatives occur quite frequently in procedural texts describing how something is done with certain tools or made from certain materials.
- Narratives focusing on the actions of people usually favour clause structures in which an agentive human participant features as the subject of an active verb. In order to test what kind of constructions the Teop language has to background human agents, I recorded descriptions of plants and what they are used for.
- A convenient text type for the analysis of syntactic behaviour of property words denoting size, colour, and shape are descriptions of fishes, flowers, and fruit.

Text types like procedural texts or the description of plants and fishes may not belong to the traditional genres of the speech community or other kinds of conventionalized language use, but as they reflect the native speakers' linguistic competence and show the expressive power of the language, they can be considered a reliable source for the syntactic analysis of the language.

3.6 Summary and Concluding Remarks

Linguistic fieldwork on a previously unresearched language presupposes a good knowledge of linguistic typology or at least of the grammar of a closely related language (§3.2). Without this background knowledge, the researcher will not be able to analyse the data he or she collects, develop new hypotheses, and accordingly prepare new questionnaires while still in the field. Researchers without linguistic training can contribute to the documentation of a language by compiling lists of words and sentences or doing recordings, transcriptions, and translations with the help of bilingual speakers (§3.3).

The success of a fieldwork trip very much depends on how the researchers interact with their indigenous consultants. The more the consultants understand the aims and the methods of the research project, the

more they will be interested in cooperation and contribute good data. Ideally they can be trained in some research activities like recording, transcription, and translation (§3.4).

Data on morphosyntactic phenomena can be gathered by various methods of elicitation and by the recording of texts. Elicitation methods can be classified into translational and non-translational elicitation. Since translational elicitation may, as many fieldworkers have observed, lead to unreliable data, the non-translational strategies like substitution, paraphrasing, and transformation are preferred (§3.5.2). For a
p. 89 thorough understanding of the morphosyntax of the target language, ↳ fieldworkers record, transcribe, and translate texts as early as possible. This corpus of texts should, if possible, comprise texts of a variety of genres and deal with diverse topics, because the selection of certain grammatical constructions depends on the linguistic and extralinguistic context of the speech event (§3.5.3).

The fieldwork guides discussed in this chapter provide useful information on how to collect various types of data, but they also make evident that the methodology of gathering data for morphosyntactic analyses in the field is still in its infancy.

Websites for further information include:

http://www.ethnologue.com

http://www.ling.udel.edu/pcole/Linguistic_Questionnaires/LinguaQ.htm

http://wals.info

http://projects.chass.utoronto.ca/lingfieldwork