Documentary and descriptive linguistics*

NIKOLAUS P. HIMMELMANN

Abstract

Much of the work that is labeled "descriptive" within linguistics comprises two activities, the collection of primary data and a (low-level) analysis of these data. These are indeed two separate activities as shown by the fact that the methods employed in each activity differ substantially. To date, the field concerned with the first activity — called "documentary linguistics" here — has received very little attention from linguists. It is proposed that documentary linguistics be conceived of as a fairly independent field of linguistic inquiry and practice that is no longer linked exclusively to the descriptive framework. A format for language documentations (in contrast to language descriptions) is presented (section 2), and various practical and theoretical issues connected with this format are discussed. These include the rights of the individuals and communities contributing to a language documentation (section 3.1), the parameters for the selection of the data to be included in a documentation (section 3.2), and the assessment of the quality of such data (section 3.3).

1. Distinguishing description and documentation

This article presents some reflections on the framework of descriptive linguistics. My concern is the application of this framework for recording little-known or previously unrecorded languages. Most of these languages are endangered, and the present reflections have been occasioned in part by the recent surge of interest in endangered languages (cf., for example, Robins and Uhlenbeck 1991; Hale et al. 1992) and the concomitant call for descriptive work on these languages.

The task of recording a little-known language comprises two activities, the first being the collection, transcription, and translation of primary data and the second a low-level (i.e. descriptive) analysis of these data.

The two activities differ substantially with respect to the methods employed as well as to their immediate results. The collection of primary data may, for example, result in a sample of some 50 utterances, in many of which a segment *lu* occurs. The methods used in putting together this sample may include participant observation (jotting down overheard utterances), various forms of elicitation, or recording, transcribing, and translating a text. Methodological issues arise with respect to the reliability, naturalness, and representativeness of the data.

The second activity — the analysis of these primary data — leads to statements such as that in language L an ergative case exists, formally expressed by a suffix *lu*, which is part of a case paradigm and has such and such further formal and semantic properties. The procedures used in arriving at such statements involve distributional tests (commutation, substitution, etc.), the analysis of the semantic properties of the utterances containing *lu*, etc. Methodological issues arise with respect to the definition of the notions 'ergative', 'suffix', etc., and the kind of evidence adduced for analyzing a certain segment as an ergative suffix.

Table 1 provides a synopsis of some of the conspicuous differences between collecting primary data and providing a descriptive analysis of these data.

Despite the differences listed in Table 1, the two activities are also closely interrelated and partially overlap for various epistemological, methodological, and practical reasons. The most important area of overlap pertains to the transcription of primary data. Any transcription requires some kind of orthographic representation. This representation will be informed by at least a preliminary phonetic and phonological analysis. Furthermore, any transcription involves decisions about segmentation (words, intonation units, turns, clauses, sentences/paragraphs). All of these decisions presuppose a certain amount of analysis on various levels. Some analysis is also involved in translation. In particular, an interlinear translation obviously presupposes some kind of morphological analysis.

Because of these interrelations and overlaps, a strong tendency exists within descriptive linguistics to blur the differences between the two activities and to consider them part of a uniform project called "describing a language." There are, however, various reasons to keep the two activities clearly separate or, more generally, to distinguish between the documentation and the description of a language. The major reason pertains to the methodological differences between the two activities listed in Table 1. The methodological differences are mirrored by the substantial differences between the products of the two activities, a language documentation and a language description, respectively, which will be highlighted in section 2.1 below. Other reasons include the following three arguments.

First, no automatic, infallible procedure exists for deriving descriptive statements from a corpus of primary data; that is, any collection of primary data allows for various kinds of analyses even within the framework of descriptive linguistics.[1] Therefore, a data collection and its analysis are not just simply two different ways of presenting the same information. This is obviously nothing new to linguistics but, rather, belongs to the few general assumptions shared by most, if not all, linguists since the failure of the post-Bloomfieldian discovery procedures project.

Second, a comprehensive descriptive analysis is not the only kind of analysis possible for a given set of primary data. A set of primary data may be of interest to various other (sub-)disciplines, including sociolinguistics, anthropology, discourse analysis, oral history, etc. This, of course, presupposes that the data set contains data and information amenable to the research methodologies of these disciplines. The chance that this kind of data and information is found in a language description is practically nil. Language descriptions are, in general, useful only to grammatically oriented and comparative linguists. Collections of primary data have at least the potential of being of use to a larger group of interested parties. These include the speech community itself, which might be interested in a record of its linguistic practices and traditions.

Third, as long as collection and analysis are considered part of a single, uniform project, the collection activity is likely to be (relatively) neglected. Historically speaking at least, it has been the case that the collection activity has never received the same attention within descriptive linguistics

Table 1. *Some conspicuous differences between collecting and analyzing primary linguistic data*

	Collection	Analysis
Result	Corpus of utterances; notes on observations and comments by speaker and compiler on a particular form or construction	Descriptive statements, illustrated by one or two examples
Procedures	Participant observation, elicitation, recording; transcription and translation of primary data	Phonetic, phonological, morphosyntactic, and semantic analyses (instrumental measurings, distributional tests, etc.)
Methodological issues	Sampling, reliability, naturalness	Definition of terms and levels, justification (adequacy) of analysis

as the analytic activity. Descriptive theory has almost exclusively been occupied with the procedures for analyzing primary data and presenting this analysis (in the format of a grammar and a lexicon). Methodological issues with respect to obtaining and presenting primary data have never been dealt with in depth within descriptive linguistics.[2] The presentation (publication) of the primary data has generally been considered a secondary task. In recent decades, hardly any comprehensive collections of primary data have been published.[3] A clear separation between documentation and description will ensure that the collection and presentation of primary data receive the theoretical and practical attention they deserve.

Much more fundamental objections could be raised against the idea that language documentation and language description are part of a single, uniform project. The essence of such objections pertains to the fact that any close link between these two activities has the consequence of the descriptive concept of language determining the kind of data considered relevant in language documentation. Consequently, any objections raised against the descriptive concept of language as a system of units and regularities will also apply to a language documentation done within the descriptive framework. As is well known, the descriptive concept of language has been criticized from various points of view, with a notable increase of criticism in recent times.[4] Among the targets of such criticism are its abstract and ahistoric conception of the speech community as a homogeneous body, its neglect truly to confront the complexities of spoken language (rather than reducing spoken language to "language as it may be written down")[5] and the concept of a language as an overall coherent system.

I will not, however, pursue this series of objections in detail. In my opinion, the arguments given above suffice in establishing the need for distinguishing language documentation from language description. Further, indirect support for such a distinction will be found while spelling out the details of a framework for language documentation in the following sections. Needless to say, any discussion of a framework for language documentation should be informed by the objections leveled against the descriptive concept of language.[6] The following sections outline such a framework. Throughout these sections, I will continue to contrast language documentation with language description in order to profile major characteristics of documentary linguistics.

2.   Language documentation and documentary linguistics

The way the argument has been presented in the preceding section does not aim at any specific framework for language documentation. In this line of argument, any format for language documentation will do as long as the documentary activity is kept separate from the analytic activity. Thus, any linguistic research that involves the collection of primary language data may, in principle, contribute to a language documentation, irrespective of the specific goal of the research. The only requirement is that the primary data be made available to other interested parties, which will always involve some editing in order to make the data accessible to the uninitiated. In this view, (a contribution to) a language documentation is nothing else than an EDITED VERSION OF THE FIELDNOTES. Particularly in those instances where further data collection will not be possible in the future, for example in the case of endangered languages, this is, I hold, a viable and useful concept for language documentation. Furthermore, it is simply a feature of a scientific enterprise to make one's primary data accessible to further scrutiny. In the case of a descriptive analysis based on fieldwork, this means making one's fieldnotes and recordings accessible.

A somewhat more radical proposal is that language documentation be conceived of as a fairly independent field of linguistic inquiry that is no longer linked exclusively to the descriptive framework. In this view, language documentation may be characterized as RADICALLY EXPANDED TEXT COLLECTION. Such a proposal does not imply that it is possible to make a "pure" documentation without any descriptive analysis (this is precluded by the interrelations and overlaps between the documentary and descriptive activities noted in section 1 above). However, the interrelation between the two activities is no longer seen as one of unilateral dependency, with the documentary activity being ancillary to the descriptive activity (i.e. primary data are collected IN ORDER TO make a descriptive statement of the language). Instead, it is assumed that the interrelation between the two activities is one of bilateral mutual dependency. Thus, conceiving of documentary linguistics as a fairly independent field of linguistic inquiry focuses on the converse perspective in which the descriptive activity is ancillary to the documentary activity (i.e. descriptive techniques are part of a broad set of techniques applied in compiling and presenting a useful and representative corpus of primary documents of the linguistic practices found in a given speech community).

Making the collection and presentation of primary data its central concern does not mean that theory has no role to play in documentary linguistics. Instead, research in this field is informed by a broad variety of language-related theories, and unifying methods and insights of various frameworks that are often treated as unrelated. The present section as well as the following one (section 3) may serve as an example of the kind

of theorizing and synthesizing a broad range of research traditions that is constitutive for the field of documentary linguistics.

Section 2.1 makes explicit some of the basic assumptions characterizing this new field of linguistic inquiry. Section 2.2 presents a proposal regarding the general format of language documentations.

2.1.   *Basic assumptions*

The concept of language documentation as a field of linguistic research and activity in its own right proceeds based on the assumption that the linguistic practices and traditions in a given speech community are worthy of documentation in the same way as material aspects of its culture (arts and crafts) are generally deemed worthy of documentation. The aim of a language documentation, then, is to provide a comprehensive record of the linguistic practices characteristic of a given speech community. Linguistic practices and traditions are manifest in two ways: (1) the observable *linguistic behavior*, manifest in everyday interaction between members of the speech community, and (2) the native speakers' *metalinguistic knowledge*, manifest in their ability to provide interpretations and systematizations for linguistic units and events.

This definition of the aim of a language documentation differs fundamentally from the aim of language descriptions: a language description aims at the record of A LANGUAGE, with "language" being understood as a system of abstract elements, constructions, and rules that constitute the invariant underlying structure of the utterances observable in a speech community. A language documentation, on the other hand, aims at the record of THE LINGUISTIC PRACTICES AND TRADITIONS OF A SPEECH COMMUNITY.[7] Such a record may include a description of the language system to the extent that this notion is found useful for collecting and presenting characteristic documents of linguistic behavior and metalinguistic knowledge. The record of the linguistic practices and traditions of a speech community, however, is much more comprehensive than the record of a language system since it includes many aspects commonly not addressed in language descriptions (cf. sections 3.2 and 3.3 below).

Related to the first assumption is the further assumption that a corpus of (extensively annotated) primary data documenting linguistic practices and traditions is of use for a variety of purposes. These include further analysis in the framework of a language-related discipline as well as projects concerning the cultivation and maintenance of its linguistic practices administered by the speech community. Conversely, no single one of the possible specific uses of a language documentation provides the

major guidelines for data collection in this framework. Instead, the makeup and contents of a language documentation are determined and influenced by a broad variety of language related (sub-)disciplines, including the following:

–sociological and anthropological approaches to language (variationist sociolinguistics, conversation analysis, linguistic and cognitive anthropology, language contact, etc.);

–"hardcore" linguistics (theoretical, comparative, descriptive);

–discourse analysis, spoken language research, rhetoric;

–language acquisition;

–phonetics;

–ethics, language rights, and language planning;

–field methods;

–oral literature and oral history;

–corpus linguistics;

–educational linguistics.

The importance of these analytic frameworks to language documentation certainly differs. The major theoretical challenge for documentary linguistics is the task of synthesizing a coherent framework for language documentation from all of these disciplines. This includes the task of determining which purposes a language documentation may realistically be expected to serve and how they can best be served by a single multipurpose documentation. Furthermore, these approaches influence language documentation procedures on two counts: first and foremost, they influence the collection process inasmuch as they contribute to the compilers' understanding of linguistic practices and traditions (and hence, influence the choice of data to be recorded). Second, they influence the recording and presentation of the data inasmuch as certain kinds of information are indispensable for a given analytic procedure (no phonetic analysis is possible without some high-quality sound recording, no analysis of gestures is possible without videotaping, etc.). These issues are taken up in sections 2.2 and 3, which contain some more concrete proposals for such a coherent framework for documentary linguistics.

The important point to repeat here is that language documentation is not to be conceived of as an ancillary procedure to any single one of these research frameworks. Instead, the interrelation between language documentation and these frameworks is one of mutual dependency: the theoretical frameworks contribute the basic inventory of analytic concepts and procedures that help to ensure the quality and usefulness of a given documentation. Conversely, good documentations provide the empirical basis for the ongoing revision and refinement of the theoretical frameworks.

A further major difference between description and documentation follows from the two assumptions just discussed. This difference concerns the role of primary data within the two frameworks. Within the documentation framework, primary data are of central concern. The goal is to present as many primary data with as much analytical information as possible. Analytic information is given in the form of a commentary (or apparatus) appended to the primary data. Within the descriptive framework, on the other hand, primary data are a means to an end, that is, the analysis of the language system. Analytic statements are of central concern. Primary data are generally not presented in full but only as exemplifications of analytic statements.

2.2. *The basic format of a language documentation*

This section presents a proposal regarding the contents and presentational format of a language documentation. The basic content is determined by the overall purpose of a language documentation, that is, to document the linguistic behavior and knowledge found in a given speech community. The presentational format is determined by the fact that the data assembled in a language documentation should be amenable to a broad variety of further analyses and uses.

Linguistic behavior is manifest in communicative events. The term *communicative event* is intended to cover the whole range of linguistic behavior, from a single cry of pain or surprise to the most elaborate and lengthy ritual. It is also meant to emphasize a holistic and situated view of linguistic behavior. That is, the target of the documentation is not a sound event all by itself, but the sound event as part of a larger communicative setting that includes the location and posture of the communicating parties, gestures, artefacts present, etc. These two features distinguish language documentations from traditional text collections, which primarily contain narrative and procedural texts and generally only document verbal behavior.

The core of a language documentation, then, is constituted by a comprehensive and representative sample of communicative events as natural as possible. Given the holistic view of linguistic behavior, the ideal recording device is video recording. Obviously, various theoretical and practical problems are associated with the task of putting together such an ideal sample. How are comprehensiveness and representativeness defined and attained? What happens when video recording is not possible? These questions are further addressed in the following section.

Metalinguistic knowledge in the sense defined above — that is, native speakers' interpretations and systematizations of linguistic units and events — is manifest in various forms. One form is special communicative events such as language plays (including linguistic jokes and puns) and lists and taxonomies that may be an integral part of poetic and ritual forms of language. Since these manifestations of linguistic knowledge are included in the sample of communicative events forming the core of a language documentation, no special section is needed for them within the overall framework of a language documentation. This also holds for another manifestation of linguistic knowledge, namely the commentary of native speakers on specimens of communicative events (comments such as "this form has such and such implications" or "that utterance may be paraphrased as follows," etc.). These comments are included in the commentary accompanying each communicative event.

However, there may be some areas of linguistic knowledge that are never fully manifest in any one single communicative event or even a very large and comprehensive corpus of communicative events. What I have in mind here are listlike linguistic phenomena such as morphological paradigms, expressions for numbers and measures, folk taxonomies (for plants, animals, musical instruments and styles, and other artefacts), etc. Generally, these will have to be elicited in cooperation with native experts. Given the overall goal of the documentation, the elicitation procedure should be organized to be as natural as possible (at least for expert taxonomies, there will usually be some kind of routine by which the expert transmits his or her knowledge to disciples). Furthermore, the whole elicitation procedure is to be considered a special, somewhat artificial communicative event and is to be documented as such (i.e. including metacomments and digressions of both compiler and native expert, discussions of problems, etc.). These kinds of data are documented in a section called *lists*. The analytical format most closely related to this section is an ethno-thesaurus.

During the work on a language documentation, the compiler will usually also elicit data or discuss language matters independent of a particular communicative event or listlike phenomenon in order to arrive at a better understanding of the language and culture in a given speech community. Since the purpose of the documentation is to make available all the primary data collected by the compiler, a third section, called *analytic matters*, is necessary for accommodating these data.

In order to allow for further analysis and processing of all these kinds of data included in a language documentation, the presentational format has to fulfill certain requirements. There are generally three components to each document (piece of data), viz. the "raw" data in various forms

of representation (transcription, tape, and/or video), a translation (word-by-word/interlinear and free), and a commentary providing additional information as to recording circumstances, linguistic and cultural peculiarities associated with the data segment, comments by native speakers cooperating in the transcription and translation of the segment, problems encountered in transcribing and translating, further data elicited in connection with the segment, etc. In short, everything that happened during recording, transcribing, and translating the data (and eliciting, in the case of elicited data).

The individual commentary for every data segment is complemented by a general introductory commentary that includes general information on the speech community (social organization, geography, history, etc.) and the language (genetic affiliation, typological characteristics, structural sketch, etc.), the fieldwork, the methods used in gathering and processing the data (including notes on the orthographic representation, interlinear glosses, etc.), and the contents and scope of the documentation.

The major difference between the language documentation format and other formats that include primary data — in particular, the standard format of descriptive linguistics, that is, grammar, lexicon, and text collection — pertains to the fact that the language documentation format is uncompromisingly data-driven. Of particular importance in this regard are two features of language documentations: first, no primary data are excluded simply because they do not fit a given analytical format or are not relevant to a particular research goal. Second, the presentation is organized around the documents (communicative events, lists, elicitation topics) rather than following a specific analytical framework.

Note that the difference does not pertain to the fact that a language documentation does not contain analytic information. It does so, but in an unconventional format. For example, a good and comprehensive documentation will include all the information that may be found in a good and comprehensive descriptive grammar. This information, however, will not be accessible in the usual way since language documentations are not organized by grammatical chapters (word order, case marking, adverbials, etc.). Instead, analytical statements regarding, for example, case marking will be found distributed among the structural sketch that forms part of the introductory commentary and the commentaries accompanying individual documents. These statements may be checked and elaborated on by scrutinizing other documents.

This format may be an inconvenience to the person who seeks quick and easily consumable information on case marking in a given language. This inconvenience is, in my view, well compensated for by the fact that a language documentation incorporates information on, and exemplifica-

tion of, case marking phenomena to an extent and degree of detail rarely achieved in conventional grammar chapters. What I have in mind here are comments, for example, on highly idiosyncratic uses of a given case marker, which may be noted and commented on in the commentary accompanying the document in which it occurs but which, in general, will only be accommodated with difficulties in a grammar chapter. Furthermore, and more importantly, in not catering to a particular interest group (e.g. the typologist interested in case marking), chances are higher that a language documentation also contains information of interest for other uses. That is, the data-driven format of a language documentation ensures that it is of potential use to a variety of interested parties. The price to be paid for this feature is that most or all interested parties may find the presentational format somewhat inconvenient.[8]

Note, finally, that the task of compiling a language documentation is not an easy one. Ideally, the person in charge of the compilation speaks the language fluently and knows the cultural and linguistic practices in the speech community very well. This, in general, implies that the compiler has lived in the community for a considerable amount of time. Furthermore, the compiler should be familiar with a broad variety of approaches to language and capable of analyzing linguistic practices from a variety of points of view. These demands will only rarely be met by a single individual. Hence, the compilation of a high-quality language documentation generally requires interdisciplinary cooperation as well as close cooperation with members of the speech community.

3.   Issues in documentary linguistics

Compiling a language documentation according to the model sketched in the preceding section involves at least the following four steps:
 –decisions about which data to collect/include in the documentation;
 –the actual recording of the data;
 –transcription, translation, and commentary;
 –presentation for public consumption/publicly accessible storage (archiving).

The implementation of each of these steps necessitates various practical and theoretical considerations and preparations. Practical considerations concern, for example, feasibility with respect to the speech community as well as the compiler. Obviously, only data for those aspects of linguistic behavior can be collected for the documentation to which the speech community consents and actively contributes (cf. section 3.1 below). In the initial phase of fieldwork, the compiler will be able to handle only

elicited materials and simple texts. A compiler, on the other hand, who has spent a lot of time in the community, is fluent in the language, and knows the culture and language very well will be in a position to put together a much larger and more sophisticated sample.

Discussing and theorizing about all the considerations relevant in language documentation are the concern of documentary linguistics. Practical as well as theoretical work within documentary linguistics does not, however, have to start at ground zero. Many of the issues relevant for language documentation have been discussed in the (sub-)disciplines mentioned above (section 2.1). Thus, much of documentary linguistics' discourse will be concerned with assessing the relevance and feasibility of concepts and procedures developed in other fields and, if necessary, adapting these to the specific problems encountered in language documentation.

In the remainder of this section, some issues in documentary linguistics are discussed in order to illustrate the range and kind of theorizing necessary in this new field.

3.1.   *Limits to documentation: rights of privacy and language rights*

It cannot be presumed that just any data the compiler may happen to get hold of is to be included in the final, publicly accessible version of the documentation. All the contributors will have a say in what can be done and what cannot be done with their contributions. Often the speech community as a whole will also want to have some control over the further processing and distribution of the data. In fact, in some communities, a documentation along the lines sketched in the preceding sections will not be possible at all. This section briefly presents some pertinent examples in order to explore the limits of the present approach to language documentation. I presume without further discussion that the interests and rights of contributors and the speech community should take precedence over scientific interests.

One major constraint on the inclusion of materials into the documentation is the contributors' right of privacy. That is, contributors have to consent to the publication of the materials provided by them. "Publication" is to be understood here in a very broad sense, that is, making a given piece of data accessible to anyone besides the contributor and the compiler. In addition, the compiler of the documentation has to take care that no data are included that may be harmful to an individual or upset the speech community (bad-mouthing, gossip, etc.), even if this possibility is not foreseen by the contributors themselves.

Most issues related to rights of privacy are self-evident. A somewhat tricky issue is, in my experience, the preference of contributors for "clean," edited data, in particular, the preference for having transcripts of spontaneous communicative events look like written texts as found in newspapers or (school)books. This involves not only eliminating false starts, digressions, etc., but often also eliminating the repetitive structures characteristic for spoken language and applying a somewhat arbitrary punctuation. Such editing precludes further analysis in a variety of frameworks, including discourse and conversation analysis and interactional sociolinguistics. I do not have a straightforward solution to offer for this issue. In some instances, however, a compromise will be possible along the following lines: publication in book form of the edited version of the text and storage of the recording and transcript of the original communicative event in a database accessed only for further scientific inquiry.

Apart from restrictions based on the rights and protection of individual contributors, the speech community as a whole — usually represented by its political and/or cultural leadership — may wish to exert its right to have a say in what kinds of data may be collected and to what extent the collection may be made accessible to interested parties outside the speech community. There are basically two motives for a speech community to restrict the extent and public availability of a documentation of its linguistic practices and traditions. One motive is the fact that its linguistic practices involve secret aspects and taboos. A public documentation of such practices would reveal the secrets or lead to the violation of a taboo, generally affecting many or all members of the speech community in a negative way. The other motive is to prevent the exploitation, ridiculing, or improper portrayal of its (linguistic) culture. The interests here are similar to the interests of the creative professions protected by copyright in Western societies.

These two motives are often interwoven and difficult to separate. The consequences for language documentation projects, however, are somewhat different. Therefore, I will briefly discuss them separately and then conclude the section with some remarks pertaining to overlaps and commonalities. I begin with the secrecy motive.

Secrecy is often considered a means of protection of the speech community, directed against nosy outsiders and originating from the pressures of external contact.[9] Although external factors may contribute to the elaboration of secrecy, the major factor in the development and stabilization of secrecy systems seems to be the goal of controlling and inhibiting information flow within a given society (called *internal secrecy* by Brandt). In such societies, access to some forms or areas of knowledge is distributed selectively among its members (often no single member has access to all

information) and usually involves complex and lengthy rites of initiation. Inevitably, at least some aspects of the social organization and practices of this society are linked to the secret knowledge or behavior. Thus, any documentation of secret knowledge or behavior destroys, or at least actively participates in the destruction of, the cultural practices it allegedly documents. Furthermore, depending on the importance of such knowledge and behavior for the overall social organization of the community, the documentation may lead to the disintegration of the speech community itself. Consequently, documentation of such knowledge and behavior is excluded without reservations from the present framework.

The extent to which language documentation is possible under these circumstances depends on the pervasiveness of secrecy in a given speech community. Often secrecy pertains to religious and ceremonial knowledge. If this constitutes a clearly separate domain within the whole network of linguistic practices in a speech community, then a comprehensive documentation of other domains may be possible. If, however, religious and ceremonial knowledge is closely interrelated with political and cultural leadership and thus not clearly separate from much of everyday interaction, then a documentation as envisioned in the present framework will not be possible. This seems to be the case in the Rio Grande Pueblos discussed by Brandt.

In the Rio Grande Pueblos, it is claimed that much of the publicly observable linguistic behavior also has secret ceremonial and religious connotations. To document everyday linguistic behavior by tape recording and/or writing it down carries the danger that aspects of secret knowledge may be "decoded" by further intensive analysis, which is possible only with permanently stored documents. That is, the fact that some linguistic forms relevant to a secret ceremonial may be overheard in daily interactions is considered harmless because no further study of such forms is possible as long as they are not permanently stored in some form. To restrict the transmission of secret knowledge to spoken language is one of the most efficient means for making sure that secret knowledge becomes known only to the proper persons.

That language documentation is not possible under these circumstances does not mean that linguistic research is completely impossible in these and similar societies. It may well be possible to pursue various analytic endeavors, provided that the primary data collected in the process are destroyed afterward. That is, although Pueblo societies may be very unwilling to consent to the recording and publication of communicative events, it may well be possible to write a grammar of a Pueblo language, keeping illustrative examples to the minimum and making sure that they do not contain any objectionable data.

Turning now to the "copyright" motive, note first that this is primarily concerned with external relationships of the speech community, a major distinction to the secrecy motive. A further distinction between the two motives pertains to the fact that a language documentation does not necessarily lead to "copyright" violations. Hence, there is no general and straightforward way of handling this issue. Instead, it is an issue that is basically open to (often lengthy and frustrating) negotiations between the speech community and the compiler(s). The only general recommendation possible is to urge compilers to familiarize themselves with actual "copyright" cases, which nowadays are very common in North America and Australia. This may be of help in increasing awareness of sensitive areas and procedures, which might lead to "copyright" conflicts in other regions of the world as well.

As I understand it, issues of secrecy and "copyright" are not likely to arise in every speech community to the same degree. Instead, they seem to be more prone to occur if a speech community exhibits the following two characteristics: first, the speech community has to be small enough to allow for an effective control of information flow both within the community and between community and outsiders. In particular, pervasive secrecy is, to my knowledge, not found in large speech communities (say, a hundred thousand speakers or more).[10] Second, linguistic practices must play a central role in the totality of cultural practices characteristic for a given speech community. If language is basically considered a convenient means of communication and not seen as closely linked to religious and mythological knowledge and behavior, fewer concerns as to potential dangers and misuse of an extensive documentation may arise.

The preceding remarks are to be taken as very preliminary notes on a topic that requires much more attention and research. At present, the prevalent tendency in linguistic fieldwork is to ignore these issues and to proceed on the assumption that linguistic research of any kind is basically of no concern to the speech community. This prevalent tendency is complemented by the tendency of a minority of researchers with firsthand experience of language-rights conflicts to generalize from their own experience and to presume that secrecy and "copyright" issues necessarily arise in every speech community.[11]

Both tendencies are wrong in my view. What is needed instead is, on the one hand, that every compiler of a language documentation be aware of these issues and take precautions in order to avoid violation of rights of privacy and language rights, irrespective of whether or not conflicts of this kind have arisen before in the geographical and/or cultural area in which the speech community is located. On the other hand, there is a need for further, in-depth empirical and theoretical exploration of these

issues in order to evaluate the practical feasibility and applicability of the documentary approach. If it turns out that, in the large majority of little-known speech communities, issues of secrecy and "copyright" rule out the kind of large-scale documentations envisioned here, the whole approach is obviously doomed to failure.

3.2.   *Parameters for the selection of communicative events*

Given that an extensive sample of communicative events forms the core of a language documentation, the following question arises: what are the relevant parameters in determining the kind and number of communicative events to be included in a language documentation? The obvious practical answer is, as many and as varied communicative events as one can get hold of and manage to transcribe and translate. Although the value of such a practical answer should not be underestimated — in some speech communities, it is difficult to find speakers who consent to the recording of some more or less natural and spontaneous linguistic behavior they are engaged in — there are several reasons for exploring the possibility of a principled, theoretically informed answer to this question within the framework of documentary linguistics.

One major reason is this: it is commonly agreed that conventional text collections, which often include only narratives and procedural texts, are far from sufficient in providing an adequate sample of the linguistic practices found in a given speech community. Hence, if a language documentation is to provide a more adequate sample of such practices, there is a need for some ideas and guidelines as to how the shortcomings of conventional text collections may be improved upon. In particular, there may be a need to develop tools for the collection of communicative events in the event that the direct recording of spontaneous specimens of a given type of communicative event turns out to be unfeasible (this will be discussed further in the next section). Guidelines as well as collection tools presuppose some kind of systematics of communicative events with respect to which the comprehensiveness and representativeness of a given corpus of communicative events may be evaluated.

Discussions of the systematics for communicative events often make reference to the notion of *genre* or *text type*, such as *mythological narrative, description, interview*, etc. These notions, however, have been shown to be very difficult to define empirically (cf., for example, Gülich 1986: 15–19; Biber 1989: 4–6). Furthermore, it is far from clear if it is possible, in principle, to arrive at a cross-linguistically applicable definition of genres. Most of the literature on genres or text types to date has been

concerned with European languages, with a heavy bias toward written language. Therefore, it seems more useful to address the issue of a systematics for communicative events on a somewhat more abstract level, that is, to explore parameters that may be useful in evaluating the comprehensiveness of a given corpus of communicative events. Note, however, that this is not to say that the notion of genre is of no relevance at all to the documentary enterprise. As will be seen in the ensuing discussion, many of the insights gained in the literature on genre are relevant to the problem of determining the kind and number of communicative events to be included in a language documentation. What is not feasible, in my view, is the attempt to establish a universally applicable grid of text types for language documentations.

Given these preliminaries, there are basically two ways to approach the problem of a systematics for communicative events within documentary linguistics. One is to approach the problem from an anthropological point of view and to ask questions such as, what kinds of communicative events occur in a given speech community? How are these conceptualized by the native speakers? What are the features characterizing nativelike communicative conduct? These kinds of questions have been addressed within the framework of the *ethnography of communication* (or *speaking*).[12] This framework rests on the assumption that communicative events are organized in culture-specific ways and it provides concepts and methods useful in probing the characteristic ways of speaking in a particular speech community. To give just one example, the set of native designations for ways of speaking often provides an important key to the native systematization of communicative events.[13] In terms of this approach, a language documentation should include specimens of communicative events for as many native categories as possible.

Note that within the present context some of the more controversial aspects of this framework, such as the notion of communicative competence, are of no particular importance. What is relevant for documentary linguistics is, on the one hand, the idea that communicative events are organized in culture-specific ways and, on the other hand, some of the methods proposed for discovering these ways. For a constructive critique of the framework and its current offspring, see Sherzer (1987), Bauman and Briggs (1990), Hill and Mannheim (1992), Woolard and Schieffelin (1994), inter alia. Closely related, and also of interest for the compilation of a language documentation, are Gumperz's *interactional sociolinguistics* (cf. Gumperz 1982) and the so-called *sociology of knowledge* (Günthner and Knoblauch 1995).

The other method is to approach the problem of a systematics for communicative events within documentary linguistics from a linguistic

point of view and to ask questions such as, is it possible to distinguish types of communicative events with respect to the kinds of linguistic structures that occur? Or, conversely, do certain kinds of linguistic structures only occur in particular kinds of communicative events? Differences in the degree and kind of linguistic structuring of communicative events have been the concern of several linguistic (sub)disciplines, including the following: first-language acquisition, where various acquisition phases are distinguished with respect to the increasing lexical and/or grammatical complexity of childrens' utterances (cf., inter alia, Ochs 1979; Ingram 1989: 32–58); research concerned with the similarities and differences between spoken and written language, where an attempt is made to determine fundamental characteristics of linguistic behavior in these two media (cf., inter alia, Akinnaso 1982, 1985; Biber 1988: 47–58; Chafe 1994: 41–50); genre research (or text typology), where, inter alia, an attempt is made to characterize various text types with respect to features of linguistic structure (say, a novel in distinction to a newspaper advertisement; cf. Gülich and Raible 1972; de Beaugrande and Dressler 1981: 188–215; Kallmeyer 1986; Biber 1988; Günthner and Knoblauch 1995).

In these research traditions, a variety of parameters for the classification of communicative events has been considered; for example, formal vs. informal, literary vs. colloquial, planned vs. unplanned, integration vs. fragmentation, detachment vs. involvement, decontextualized vs. contextualized, elaborated vs. restricted, abstract vs. concrete, written vs. spoken. From among these parameters, the parameter of spontaneity ("plannedness") in the sense of Ochs (1979) seems to be the most general. It is also sufficiently operational to serve as the basic parameter for a comprehensive yet flexible categorizational scheme.[14]

The parameter of spontaneity refers to the amount of time available for planning one's verbal behavior, which varies quite extensively in accordance with the kind of communicative event. Planning here includes the mental preparation of both the content of the message and its linguistic form. From a somewhat different point of view, this may also be interpreted as the degree of control that speakers may exert on their linguistic behavior.

This parameter does not describe a binary taxonomy (planned vs. unplanned). Instead, *unplanned* and *planned* designate the end poles of a continuum of spontaneity along which particular communicative events may be placed. At least the following five major types may be distinguished with respect to this continuum:[15]

Exclamations,   spontaneous and uncontrolled, such as pain cries or signs of surprise. These are very similar to real indexical signs (or symptoms)

since they are directly and often also causally connected with the designated state of affairs (e.g. the intensity of the pain cry in general reflects the intensity of the pain experience).

Directives: short utterances that are completely integrated into a sequence of actions. They generally serve to coordinate actions of several individuals. Typical examples are, in a medical context, "syringe!" or "scalpel!" One-word utterances of children are also related to this type (cf. Ochs 1979: 51 ff., 58 ff., and the literature referred to there).

Conversations,   in which the sequence of speech events is not exclusively controlled by a single individual. Instead, the overall communicative event is constructed interactionally by the participants. In some instances, even a single linguistic construction is coconstructed by two or more participants (cf. Lerner 1991; Ono and Thompson i.p.). Note that this does not mean that participants in a conversation are generally equal in their ability to control the interaction. What is important here is that in any conversational interaction, the possibilities of planning one's own linguistic contributions are, at least to some degree, limited by the necessity of interacting with other participants.

Monologues: communicative events in which a single speaker has considerable control over the speech event and is the sole or primary contributor for an extended period of time. This not only allows for but actually demands a certain amount of planning in order to produce a reasonably coherent speech event.

Ritual speech event,   in which linguistic behavior is reproduced, behavior that has been learned and rehearsed some time before it is performed (the actual performance may include some improvisations). The performance of ritual speech events typically occurs at fixed times and places and is thus often plannable long beforehand. Note, however, the following ambivalence of this type of communicative events with respect to the parameter of spontaneity: the planning and preparation of ritual speech events is dissociated from the actual performance; that is, the actual performance may be highly automatized and thus not require much planning. From this point of view, the performance of ritual speech events shares features with highly automatized speech events closer to the spontaneous pole of the continuum, such as highly conventionalized forms of greetings, etc.

Figure 1 provides an overview of the parameter of spontaneity and the five major types just discussed. Note that no clearcut borders exist

Parameter	Major types	Examples
unplanned		
	exclamative	'ouch!'
		'fire!'
	directive	'scalpel'
		greetings
		small talk
	conversational	chat
		dicussion
		interview
	monological	narrative
		description
		speech
		formal address
	ritual	litany
planned		

Figure 1.   *Types of communicative events according to the parameter of spontaneity*

between the five major types. Instead, various transitional subtypes are to be expected. For each major type, a typical example is given, drawn from the repertoire of communicative events found in European speech communities. Examples located between the five major types exemplify transitional subtypes.

It should be clearly understood that the usefulness of this parameter for documentary linguistics is based on the assumption that spontaneity is correlated with aspects of linguistic structure. In particular, it is assumed that the complexity of linguistic structuring found in a given type of communicative event correlates with the degree of planning and preparation allowed for, and required, on the part of the participating speakers. Linguistic complexity may be manifest in lexical choices and/ or morphosyntactic structure (Ochs [1979] provides some illustrative examples).

Using this parameter as a guideline in compiling a corpus of communicative events, the following points should be kept in mind:

The ability of producing more planned varieties of speech is generally acquired in later stages of the language-acquisition process. The varieties involving the highest degree of planning even require specific, often institutionalized, training. This is obvious, and commonly acknowledged, for ritual speech events. However, the ability for various forms of speaking monologically as well is often not evenly distributed among the members of a given speech community. There are people who are good at storytelling and those who are good at explaining or describing a cultural event or practice. Others may not be very skilled at being the sole or dominating speaker for some time. Often, only a few individuals are able to make public speeches. Accordingly, the compiler of a language documentation has to take care that the person(s) asked to contribute a certain kind of communicative event are actually familiar with this kind of communicative event and have some routine in delivering it.

It is highly probable that the parameter of spontaneity per se and perhaps also the five major types are applicable for systematizing communicative events in all speech communities. It should not, however, be presumed that the examples given in Figure 1 are universally attested. The interview, for example, is a communicative event that is closely linked to Western social science and cannot be expected to be found in many other societies.

The five major types are somewhat heterogeneous. For one, the large majority of communicative events taking place in a given speech community will be conversational. Thus, if overall frequency in everyday behavior were to be taken as a measure for the representativeness of a given corpus, then the majority of the specimens included in a corpus should belong to the conversational type. Furthermore, for the conversational and the monological types far more subtypes exist than for the other types (in Figure 1, this fact is indicated by allocating relatively more space to these two types). Therefore, if the overall representativeness of a given corpus were to be defined in qualitative terms, then a broad variety of monological and conversational subtypes should be included. Finally, a given communicative event may often not be subsumed unequivocally under a single type but may, instead, contain segments belonging to different types. For example, conversations are often interspersed with monological phases or brief directives. Hence, it is not easy in practice to "measure" the representativeness of a given corpus both in quantitative and in qualitative terms.

Given this heterogeneity, it follows that no simple overall scheme for the kinds and number of communicative events to be included in a language documentation may be deduced from the parameter of spontaneity. Instead, this parameter may serve as a guideline in the sense that it allows for the evaluation of a given corpus with respect to the variety of linguistic structures that may be expected to be attested in it. Its major use is to make the compiler aware of potential gaps in the data collected up to a certain stage in the research. Examples of clear gaps include the total lack of directives (either as relatively isolated communicative events or as part of a more complex communicative event) or the fact that all specimens of the monological type basically belong to one subtype.

Ideally, then, a language documentation should contain specimens of communicative events of as many different degrees of spontaneity as possible. This basic "rule," however, may be modified by taking account of more specific hypotheses, which claim that some kinds of communicative events contain particularly rich and important data and are, therefore, to be documented first and foremost. For example, from a grammarian's or typologist's point of view, a case could be made for emphasizing the collection of relatively simple monological communicative events since, in monological speech, the grammatical resources available to the speakers in a given speech community are easiest to detect and are also often pushed to their limit, as it were. Sherzer (1987), on the other hand, advances the hypothesis that verbal art and speech play, that is, communicative events tending even more strongly toward the planned end of the continuum in Figure 1, may be considered as "an embodiment of the essence of culture and as constitutive of what the language–culture–society relationship is all about" (1987: 297). If this hypothesis is correct, a strong case could be made for making documents of verbal art and speech play the center of a language documentation. I am presently not in a position to provide an in-depth discussion of this issue. It is mentioned here as a further example of the kinds of issues that need further study and discussion within the framework of documentary linguistics, in which the practical and theoretical priorities are not identical to those of either typologists or linguistic anthropologists.

So far, the discussion has been limited to specimens of spoken language, based on the (implicit) assumption that the documentation of spoken language will be of central concern in every language documentation. If in a given speech community some linguistic practices make use of other media, say writing or (hand-)signing,[16] these should, of course, also be documented. That is, the parameter of spontaneity is to be complemented by a second parameter, the parameter of modality. Note that the parameter of spontaneity is applicable to all modalities. Thus, with respect to writing, for example, one may distinguish relatively spontaneous forms, such as notes, personal letters, e-mail exchanges, etc., from more planned varieties such as scientific writing and literature.

It is often assumed that the parameter of modality is also of a continuous nature (cf. Akinnaso 1982; Biber 1988).[17] This, however, seems to be a misconception. This misconception probably arose, on the one hand, from the preoccupation of written language research with highly planned and formal varieties of writing, as pointed out by Akinnaso (1985). On the other hand, it may have been fostered by the existence of cross-modal forms of linguistic behavior such as dictation, discussions or lectures based on written notes, etc. The existence of cross-modal

forms, however, should not be conceived of as a transitional phenomenon, providing a continuous link between speaking and writing. Physically, no continuum between speaking and writing (or hearing and reading) exists (cf. also Chafe 1994: 41 ff.). A given specimen of linguistic behavior is clearly either speaking or writing (which does not preclude the possibility that one individual may be engaged in these two forms of linguistic behavior at the same time). These are two totally separate forms of action, controlled by separate neural motor centers. This view is well supported by neurolinguistic evidence, which shows that clear dissociations exist between disorders of each linguistic modality (cf. Shallice 1988: 68–157). Therefore, I assume here that the parameter of modality is categorial rather than continuous.

This concludes the present discussion of a framework for the typology of communicative events from the linguistic structure point of view. To summarize: linguistic practices in a speech community may make use of different media (signing, speaking, writing). For each medium, various degrees of spontaneity may be distinguished. From the point of view of these two parameters, the goal of a comprehensive language documentation, then, is to provide specimens from each modality in as many degrees of spontaneity as possible.

The two approaches to the compilation of a corpus of communicative events sketched in this section — the anthropological approach and the linguistic-structure approach — are based on different conceptual frameworks and aim at two different kinds of comprehensiveness. The results of these approaches, that is, the kind and number of communicative events chosen for inclusion in the corpus, however, substantially overlap and complement one another. Thus, it seems reasonable to assume that the combination of the two approaches should, in practice, result in a sufficiently varied and comprehensive corpus, which will be amenable to further analysis in a broad variety of analytic frameworks.

Note, finally, that we have been concerned in this section with the problem of how to provide for sufficient variety in a corpus of communicative events. This concern originated with a widespread criticism regarding conventional text collections (i.e. that the materials included there do not reflect the variety of linguistic practices in a given speech community to a sufficient degree). The concern with variety, however, should not lead to the neglect of the opposite concern: there is also a need for a considerable amount of repetition in such a corpus. That is, for each type of communicative event, several examples are required in order to be able to determine what is "regular" and what is ad hoc in a given specimen (recall that any distributional analysis crucially depends on the fact that the unit investigated is repeatedly attested).

3.3.   *Quality of data (gathering procedures)*

In language documentation, as in many other sciences, the gathering of data is confronted with the phenomenon commonly known as the *observer's paradox* (cf. Labov 1972a: 113): The object of research is susceptible to change because of the ongoing research process (the presence of the researchers and/or their research tools, etc.). Hence, the question arises as to HOW the data compiled in a language documentation can be, and should be, collected. More generally said, an important area of practical and theoretical inquiry within documentary linguistics is concerned with the evaluation and development of data-gathering procedures.

The question of how to deal with the observer's paradox in language documentation is but one of the topics to be addressed in this area. Another important topic is the exploration of new ways for gathering linguistic data. This is important because it is fairly obvious that the conventional gathering procedures dealt with in linguistic field-method textbooks (e.g. Samarin 1967; Bouquiaux and Thomas 1992) will not suffice for achieving as comprehensive a documentation as envisioned in the preceding sections. In this section, I briefly elaborate on these and related topics in data-gathering methodology. As in the preceding sections, my goal here is not to present "solutions" but, instead, to point out problems in need of further exploration within the framework of documentary linguistics.

The starting point for this exploration is the considerable body of literature on the phenomenon that linguistic behavior changes dramatically if the speakers pay more than the usual attention to how they are speaking. In particular sociolinguists have studied this phenomenon and also experimented with techniques to circumvent or counterbalance the observer effect (cf. Labov 1972a, 1972b; Milroy 1987: chapter 3, for a more recent review). From this research, it can be safely concluded that gathering procedures clearly affect the kind and quality of linguistic data.

For documentary linguistics, a basic — and fairly easily implemented — consequence of this insight is requiring that all data compiled in a language documentation be coded as to their recording/gathering circumstances. This is important since it allows for evaluation of the relevance of a specific piece of data for a given analytical proposal. Within a usage-based approach to grammar, for example, it is important to know whether a particular grammatical phenomon occurred in a fairly natural conversation or only in elicitation.

For evaluating the quality of data, it may again be useful to sketch a typology of communicative events with respect to their "naturalness." The basic parameter of such a typology is the degree to which speakers

are linguistically self-aware, ranging from complete unawareness to paying full attention to linguistic form as in a metalinguistic evaluation of a given form or construction.[18] Linguistic self-awareness on the part of the contributors is influenced in various ways by the compiler(s) of a language documentation. The mere presence of a person known to be investigating linguistic behavior may already have some influence on the contributors' linguistic behavior. This influence increases in accordance with the degree to which the investigator dominates or controls the interaction between the contributors and him- or herself. The control is particularly strong in the case of communicative events that have been "invented" for research purposes. The best-known communicative event of this kind is the (scientific) interview, the linguistic variant of which is called *elicitation*.[19]

Taking the notion of observer-induced linguistic self-awareness as a fundamental parameter, the following basic types of communicative events may be distinguished with respect to "naturalness":

Natural communicative events: communicative events unaffected by any external interference into the conventional communicative routines of the participants. Such events are, in principle, not amenable to documentation since the documentation process itself constitutes an extraordinary factor in the communicative situation.[20]

Observed communicative events: communicative events in which external interference is limited to the fact (known to the communicating parties) that the ongoing event is being observed and/or recorded. Such interference may be caused by the presence of an observer who occasionally takes notes or by the presence of a recording device. That is, all forms of participant observation are to be included here.

Staged communicative events: communicative events that are enacted for the purpose of recording. The important difference between these kinds of communicative events and the preceding ones pertains to the fact that staged communicative events are not "really" communicatively functional, that is, they do not serve any specific communicative purposes other than producing data. Within this category, a distinction may be made between staged events for which only rather general instructions are given (such as "tell me that fairy tale we talked about") and staged events for which more specific props are given (such as pictures, toys, or a film). The former include the kind of communicative events commonly found in text collections (elicited narratives, descriptions, etc.). The use

of props to stimulate particular kinds of communicative events is a fairly recent development. Examples include films (the Pear Story, cf. Chafe 1980), picture books (the Frog Story, cf. Berman and Slobin 1994), and pictures and toys used to generate discourse on a specific topic such as space (the space games developed by the Cognitive Anthropology Research Group in Nijmegen, cf. de León 1991; Levinson 1992).

Elicitation: a type of communicative event invented for conducting linguistic research and documentation. In most communities, it may be safely assumed that this is, in fact, a new type of communicative event for the members of the speech community since giving comprehensive information about everyday linguistic practices is rarely a conventional part of such practices.

With respect to the degree of control the investigator exerts in the overall interaction, the following three "styles" of elicitation may be distinguished (in order of increasing control): (a) *contextualizing elicitation*, where native speakers are asked to comment on or provide contexts for a word or construction specified by the researcher; (b) *translation*, where native speakers are asked to translate a form provided by the researcher into their native language; (c) *judgment*, where native speakers are asked to evaluate the acceptability or grammaticality of a given formation.

A fundamental challenge for documentary linguistics is posed by the fact that the more natural kinds of communicative events are the most difficult to attain. That is, the direct recording of actual (i.e. nonstaged) communicative events is often not consented to by the contributors. This is, for obvious reasons, particularly common for the more informal and personal varieties of communicative interactions (and even if such events can be recorded, it will often be impossible to publish them, since this would incur violation of privacy rights). Furthermore, participant observation without the help of recording devices tends to be of limited value for documenting linguistic practices since it is impossible to note in writing or to reconstruct from memory the linguistic details of a communicative event.

Given this somewhat pessimistic assessment of the possibilities of observing actual communicative events, it follows that one major concern of documentary linguistics in regard to data-gathering procedures will be with the evaluation and further elaboration of elicitation techniques and techniques for staging communicative events. As for staging communicative events, it was already mentioned above that the use of props in this task is a fairly recent development, which certainly warrants further study

and elaboration. Two features of this technique seem to me particularly appealing: first, the use of props offers some direction to the linguistic behavior of contributors without directly focusing their attention on their linguistic behavior (for example, the space games generate discourse in which heavy use is made of space-related concepts and constructions without making explicit reference to these concepts and constructions). Second, data generated by this technique are better and easier to compare on various levels (across speakers of one community as well as across dialects and languages) than miscellaneous data collections.

However, there are also various problems associated with this technique. For example, some of the props that have been used so far have turned out not to be universally applicable (cf. DuBois's [1980] report on his attempts to show the Pear Film in a Maya community in Guatemala). That is, it may very well be that many (most?) props have to be culturally or perhaps even individually adapted in order to successfully stage communicative events (which would obviously affect the comparability potential of this technique). Furthermore, the quality of the data generated by this technique is, to date, far from clear. That is, there is a need to compare such data with data generated by other techniques and to assess specific distortions induced by the props and the overall staging of a communicative event. English and German Pear Film narratives, for example, are conspicuous for the fact that they involve a continuous switch between the narrative event line and comments regarding technical aspects of the film, a feature not found in other kinds of narratives. These and other problems make it clear that there is still a lot of further study needed in order to assess the limits and possibilities of this technique.

This also holds for elicitation, the central technique of descriptive linguistics. This technique has received some methodological attention with respect to its use in Western societies (cf., inter alia, Greenbaum and Quirk 1970; Labov 1975; Milroy 1987; Schütze 1996). Following Briggs (1986), the critical assessment of this method within documentary linguistics has to begin by acknowledging the fact that it involves, in some sense, the invention of a new type of communicative event in those speech communities that are unfamiliar with the research procedures of Western social science. Note that it is not the case that these speech communities generally do not engage in any metalinguistic practices at all. For specific linguistic practices, such as ritual ways of speaking or nomenclatures, there are usually indigenous ways of "teaching." Such indigenous ways of "language teaching" may be a useful starting point for developing — in cooperation with the contributors — an elicitation style that suits the contributors. In particular, it will generally be most

fruitful to conceive of elicitation as a kind of teaching event for which input and control on the part of the native speakers is essential, rather than as some kind of "objective," culturally neutral way of obtaining data to be administered under total control of the researcher.

Seeing elicitation as a kind of teaching event also provides a starting point for assessing various factors that may influence the quality of elicited data. Certainly, one such factor is the degree to which native speaker and investigator have mastered a common medium of communication. That is, it makes a difference whether both parties have to articulate themselves in a foreign language or whether they converse reasonably fluently in the native language. In fact, I would hold that a discussion of more fine-grained topics in syntax and semantics (for example, scope ambiguities) will only be fruitful if the investigator speaks the language well enough to be able conduct such a conversation in the native language. Otherwise, one must always reckon with strong interferences from the language in which the conversation is conducted.

The preceding remarks on elicitation may seem fairly obvious. In my experience, however, discussions concerning the merits of elicitation tend to gloss over the fact that there are many elicitation styles. Hence, I do not think that it is possible to make general claims such as "elicitation is possible and useful" or "elicitation can never produce reliable data." What is needed, instead, is a careful evaluation of various elicitation styles and the factors contributing to their usefulness.

3.4.   *Further issues*

The three issues discussed in the preceding sections obviously do not exhaust the range of topics to be addressed by documentary linguistics. In this section I simply list a few other issues in order to give an idea of the range of topics that need further exploration. This list is not exhaustive.

One important topic pertains to the question of how the communities can be actively involved in the design of a concrete documentation project from the very beginning. How can a documentation project be presented to a community in such a way that the community is likely not only to accept it but also to shape it in essential aspects? In some communities there may exist strong ideas about how documentation should proceed, which do not completely accord with the researchers' plans. How can and should such conflicts be resolved? Closely linked to the issue of the participatory design of a documentation project is the issue of the researchers' involvement in language-maintenance work, which may be

of greater interest to the community than just a documentation. Some ideas and suggestions on these issues may be found in work by Cameron et al. (1992), Craig (1992), England (1992), Jeanne (1992), Watahomigie and Yamamoto (1992), Wilkins (1992), McConvell and Florey (1994), and Wolfram and Schilling-Estes (1995:717f.).

Another complex of issues, which is related to the previous one, is concerned with the technical problems posed by language documentations such as the choice of an appropriate recording and presentation technology (sound recording, video, multi-media applications, etc.), the problem of archiving and maintaining documentations,[21] and the problem of providing and controlling access to documentations.

There is a whole set of issues related to the preparation of the language documents that form the core of a language documentation. For example, it was said above (in section 2) that the documents have to be translated. Hence, questions arise such as, what is a useful translation within the documentary framework? What language or languages should be used as target languages? How can an adequate translation be achieved? Similar issues arise in relation to interlinear glosses. When documenting natural communicative events, there is the problem of how to segment written representations of such events (for example, in clauses and/or intonation units and/or paragraphs?). Another broad area that needs further exploration is the commentary (or apparatus) that is to be appended to each document: what kind of information should be included? How is it to be organized to be maximally accessible? How can redundancies be avoided?

Funding is another important issue. Documentation projects are long-term projects, that is, projects for which it is particularly difficult to get continuous funding (most academic funding agencies have time limits for projects in the range of two to five years). Furthermore, academic funding agencies tend to be unwilling to make substantial funds directly available to communities or community members (rather than researchers). The ideal solution to these and other funding problems would be the establishment of a foundation devoted to documenting and supporting the maintenance of (endangered) languages. Note that there are a number of foundations (including, for example, the very prosperous Getty foundation) devoted to documenting and archiving objects of material culture, in particular art and archeological objects. It is certainly much more difficult to attract substantial donations to the much more abstract and less tangible object of language. But, it seems to me, this is not a totally unrealistic long-term perspective for establishing a funding basis for documentary linguistics.[22]

5.   Conclusion

In this article, a case has been made for the claim that work on previously unrecorded languages involves two essentially separate activities, the activity concerned with collecting, transcribing, translating, and commenting on primary data and the activity concerned with the further analysis of such data in a given analytic framework (in particular, the framework of descriptive linguistics). There are several possibilities of acknowledging, in theory and practice, the separate nature of these two activities. One possibility would be to edit one's fieldnotes, that is, to make the fieldnotes available to other interested parties in a format that allows the uninitiated to work with these data.

The central concern of this article, however, was to explore another, somewhat more radical possibility, that is, to conceive of language documentation as a field of linguistic inquiry and research in its own right. Part of this exploration was an attempt to make explicit some basic assumptions constitutive for the field of documentary linguistics, in particular, the assumption that it is possible and useful to document the linguistic practices characteristic for a given speech community. Based on these assumptions, a framework for language documentation and documentary linguistics was sketched, with special emphasis on the fact that language documentation is NOT some kind of "theory-free" enterprise. Instead, documentary linguistics is informed by a broad variety of theoretical frameworks and requires a theoretical discourse concerned with conceptual and procedural issues in language documentation. For some of these issues, a few preliminary proposals are presented in section 3.

Notes

1. Cf. Householder's (1952: 260f.) famous distinction between "God's truth" and "hocus pocus" attitudes in American structuralism and comments and references on this topic in Hymes and Fought (1981: 148–151).

2. The methodological discussion in fieldwork manuals such as Samarin (1967) or Bouquiaux and Thomas (1992) is primarily concerned with practical aspects in obtaining primary data (suggestions on how to interact with "informants," how to organize an elicitation session, the use of questionnaires, etc.). Note that in several other linguistic subdisciplines, in particular within sociolinguistics and psycholinguistics, the standard of methodological reflection regarding the collection and nature of the data used is much higher than in descriptive linguistics.

3. The handling of primary data in descriptive linguistics is a fairly complex issue that cannot be dealt with adequately in the present contribution. In a more comprehensive treatment, at least the following issues would have to be discussed:

    There are two places for primary data in the "classical" format for language descriptions (grammar, dictionary, and text): the text collection and the examples given in the dictionary and, usually to a much lesser extent, in the grammar in support of an analysis or as an illustration of a possible use for a given form. The increase in the neglect of primary data in the most recent descriptive practice can be gleaned by comparing the ratio of recently produced grammars, dictionaries, and text collections (which is probably something like 10 to 3 to 1). It is also instructive to compare the number of examples appended to each analytic statement in older grammars such as Jespersen's *Modern English Grammar* or Paul's *Deutsche Grammatik* with more recent grammars. A comparison of dictionaries will bear similar results. The fate of text collections in descriptive linguistics is discussed in Himmelmann (1996: 323 f.).

    It should be clearly understood that, in general, descriptive linguists are the last to be blamed for the increasing neglect of primary data in publications. This, rather, is due to the prevailing climate in the field, which is heavily biased toward supporting theoretical work. Ph.D. students are expected to produce theoretical analyses (in some departments, grammars are also acceptable) rather than dictionaries or text collections. Grammars carry much higher prestige in the academic community than dictionaries and text collections (with practical consequences such as getting a job, finding a publisher, etc.). Publishers exert heavy pressure on keeping the manuscript short and to the (theoretical) point.

4. Cf., among others, Coseriu (1974), Harris ([1981] and the contributions in Davis and Taylor [1990]), Hockett (1987), Hopper (1987), Bybee (1988).

5. Reducing spoken language to "language as it may be written down" means neglecting prosodic phenomena and excluding gestures and paralinguistic phenomena from the field of linguistic inquiry proper. Furthermore, in descriptive linguistics there is also a strong tendency to gloss over the many problems involved in segmenting extended stretches of spoken discourse (cf. Serzisko 1992 for references and discussion).

6. Of major importance in this regard are Pawley's (1993 and elsewhere) observations on the difficulty of providing an adequate representation of speech formulas in the descriptive framework.

7. Following Romaine (1994: 22), a speech community is to be understood as a group of people who "share a set of norms and rules for the use of language," not necessarily sharing the same language.

8. This inconvenience may, of course, be overcome by extracting from a language documentation and reorganizing the information relevant to a particular analytical framework. The compiler may be the most efficient person to do this. In principle, however, any person familiar with the framework in question and the handling of primary data is in a position to recast the information given in language-documentation format into the format conventionally used in that framework (a descriptive grammar, for example).

9. The following discussion presents a brief paraphrase of the argument in Brandt (1980, 1981). Brandt's work is based on long-time, first-hand experience with the Pueblo societies in the Rio Grande valley in New Mexico, in particular Taos Pueblo. Her observations, however, seem to be extendable to other societies in which access to and proper use of knowledge is a central concern, and political leadership is intricately and inseparately linked to religious and ceremonial knowledge.

10. This, of course, does not mean that secret societies with secret linguistic practices may not exist in large speech communities (cf., for example, Freemasons' lodges). However, such secret practices are, in general, not intimately linked to everyday linguistic practices, which thus may be documented without endangering the secret practices.

11. See Skutnabb-Kangas and Phillipson (1995) for a much more comprehensive assessment of language rights.

12. A comprehensive statement of the framework by its originator may be found in Hymes (1974). Saville-Troike (1989 [1982]) provides a textbook account.

13. Such a set may be fairly extensive: for German, Dimter (1981: 33) counts 1642 terms for communicative events in the *Duden* (480 of which he classifies as "basic," the remaining ones being "derived"). Stross (1974) lists 416 variants of the word for "speech" in Tzeltal.

14. Cf. Starks (1994) for an empirical demonstration of the relevance of this parameter.

15. Not surprisingly, these five types are more or less directly related to well-known functions of language (cf. Thrane [1980: 2 f.] for a brief synopsis and references).

16. Here, "signing" does not refer to gestures accompanying spoken language (these are considered part of the overall communicative events in which a given specimen of spoken language occurs). Instead, it refers to *alternate sign languages* (a term proposed by Kendon [1988: 4]), i.e. sign languages used under special circumstances by hearing speakers.

    I am not competent to judge the applicability of the present framework to the linguistic practices found in deaf communities.

17. I owe my understanding of this point to Fritz Serzisko, with whom I have had many a helpful discussion on this topic.

18. Cf. Labov's *principle of attention* (1972a: 112). The continuum of contextual styles that Labov defines on the basis of the principle of attention is related to, but not identical with, the following typology of communicative events with respect to "naturalness" (Traugott and Romaine [1985] provide a forceful critique of Labov's continuum). Instead, the phenomena that Labov brings together in one unidimensional continuum are here distributed among three parameters, i.e. spontaneity, modality, and observer-induced linguistic self-awareness.

19. Cf. Briggs's (1986) critical appraisal of the interview as a research tool.

20. Note that "in principle" here does not mean that the documentation process necessarily leads to substantial changes in the recorded linguistic behavior or that those changes that do occur are necessarily relevant to a given research goal. Instead, "in principle" simply refers to the fact that we may not know since we cannot compare the "natural" state of affairs with the documented state of affairs. Furthermore, one exception to this

"principle" may exist: inasmuch as documentation has become an integral part of a given type of communicative event (for example, in parliament or in trials), a case can made for considering such events documented in their "natural" form.

21. Lehmann (1992) explores the idea of a museum for language and languages.

22. One such fund, the Endangered Language Fund, has recently been established. For further information see the Website (http://sapir.ling.yale.edu/~elf/index.html).

References

Akinnaso, F. Niyi (1982). On the differences between spoken and written language. *Language and Speech* 25, 97–125.

—(1985). On the similarities between spoken and written language. *Language and Speech* 28, 323–359.

Bauman, Richard; and Briggs, Charles L. (1990). Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology* 19, 59–88.

Berman, Ruth; and Slobin, Dan I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Erlbaum.

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

—(1989). A typology of English texts. *Linguistics* 27, 3–43.

Bouquiaux, Luc; and Thomas, Jacqueline M.C. (1992). *Studying and Describing Unwritten Languages*. Dallas, TX: Summer Institute of Linguistics.

Brandt, Elizabeth A. (1980). On secrecy and control of knowledge. In *Secrecy. A Cross-Cultural Perspective*, Stanton K. Tefft (ed.), 123–146. New York: Human Sciences Press.

—(1981). Native American attitudes toward literacy and recording in the Southwest. *Journal of the Linguistic Association of the Southwest* 4, 185–195.

Briggs, Charles L. (1986). *Learning How to Ask. A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research*. Cambridge: Cambridge University Press.

Bybee, Joan L. (1988). Semantic substance vs. contrast in the development of grammatical meaning. *BLS* 14, 247–264.

Cameron, Deborah; Frazer, Elizabeth; Harvey, Penelope; Rampton, Ben; and Richardson, Kay (1992). *Researching Language. Issues of Power and Method*. London: Routledge.

Chafe, Wallace L. (ed.) (1980). *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.

—(1994). *Discourse, Consciousness, and Time*. Chicago: University of Chicago Press.

Coseriu, Eugenio (1974). *Synchronie, Diachronie und Geschichte*. Munich: Fink.

Craig, Colette (1992). A constitutional response to language endangerment: the case of Nicaragua. *Language* 68, 17–24.

Davis, Hayley G.; and Taylor, Talbot J. (eds.) (1990). *Redefining Linguistics*. London: Routledge.

de Beaugrande, Robert; and Dressler, Wolfgang U. (1981). *Einführung in die Textlinguistik*. Tübingen: Niemeyer.

de León, Lourdes (1991). *Space Games in Tzotzil: Creating a Context for Spatial Reference*. CARG Working Paper 4. Nijmegen: MPI.

Dimter, Matthias (1981). *Textklassenkonzepte heutiger Alltagsprache*. Tübingen: Niemeyer.

DuBois, John (1980). Introduction — The search for a cultural niche: showing the Pear film in a Mayan community. In *The Pear Stories*, Wallace L. Chafe (ed.), 1–7. Norwood, NJ: Ablex.

England, Nora C. (1992). Doing Mayan linguistics in Guatemala. *Language* 68, 29–35.

Greenbaum, Sidney; and Quirk, Randolph (1970). *Elicitation Experiments in English*. London: Longman.

Gülich, Elisabeth (1986). Textsorten in der Kommunikationspraxis. In *Kommunikationstypologie*, Werner Kallmeyer (ed.), 15–45. Dusseldorf: Schwann.

—; and Raible, Wolfgang (eds.) (1972). *Textsorten*. Frankfurt: Athenäum.

Gumperz, John J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.

Günthner, Susanne; and Knoblauch, Hubert (1995). Culturally patterned speaking practices — the analysis of communicative genres. *Pragmatics* 5, 1–32.

Hale, Kenneth; et al. (1992). Endangered languages. *Language* 68, 1–42.

Harris, Roy (1981). *The Language Myth*. London: Duckworth.

Hill, Jane H.; and Mannheim, Bruce (1992). Language and world view. *Annual Review of Anthropology* 21, 381–406.

Himmelmann, Nikolaus P. (1996). Zum Aufbau von Sprachbeschreibungen. *Linguistische Berichte* 164, 315–333.

Hockett, Charles F. (1987). *Refurbishing our Foundations*. Amsterdam: Benjamins.

Hopper, Paul (1987). Emergent grammar. *BLS* 13, 139–157.

Householder, Fred W., Jr. (1952). Review of *Methods in Structural Linguistics* by Z. S. Harris, *IJAL* 18, 260–268.

Hymes, Dell (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.

—; and Fought, John (1981). *American Structuralism*. The Hague: Mouton.

Ingram, David (1989). *First Language Acquisition*. Cambridge: Cambridge University Press.

Jeanne, LaVerne M. (1992). An institutional response to language endangerment: a proposal for a Native American Language Center. *Language* 68, 24–28.

Jespersen, Otto (1909–1949). *A Modern English Grammar on Historical Principles*, vol. 1–7. Copenhagen: Munksgaard.

Kallmeyer, Werner (ed.) (1986). *Kommunikationstypologie: Handlungsmuster, Textsorten, Situationstypen*. Düsseldorf: Schwann.

Kendon, Adam (1988). *Sign Languages of Aboriginal Australia*. Cambridge: Cambridge University Press.

Labov, William (1972a). Some principles of linguistic methodology. *Language in Society* 1, 97–120.

—(1972b). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

—(1975). *What is a Linguistic Fact?* Lisse: de Ridder.

Lehmann, Christian (1992). Das Sprachmuseum. *Linguistische Berichte* 142, 477–494.

Lerner, Gene H. (1991). On the syntax of sentences-in-progress. *Language in Society* 20, 441–458.

Levinson, Stephen C. (1992). Primer for the field investigation of spatial description and conception. *Pragmatics* 2, 5–47.

McConvell, Patrick; and Florey, Margret (eds.) (1994). *Language Shift and Maintenance in the Asia Pacific Region*. ALI-94 Workshop Reader. Melbourne: LaTrobe.

Milroy, Lesley (1987). *Observing and Analysing Natural Languages: A Critical Account of Sociolinguistic Method*. Oxford: Blackwell.

Ochs, Elinor (1979). Planned and unplanned discourse. In *Discourse and Syntax*, Talmy Givón (ed.), 51–80. New York: Academic Press.

Ono, Tsuyoshi; and Thompson, Sandra A. (i.p.). Interaction and syntax in the structure of conversational discourse. In *Discourse Processing: An Interdisciplinary Perspective*, Eduard Hovy and Donia Scott (eds.). Heidelberg: Springer.

Paul, Hermann (1916–1920). *Deutsche Grammatik*, Bd. I–V. Halle a.S.: Niemeyer.

Pawley, Andrew (1993). A language which defies description by ordinary means. In *The Role of Theory in Language Description*, William A. Foley (ed.), 87–129. Berlin: de Gruyter.

Robins, Robert H.; and Uhlenbeck, Eugenius M. (eds.) (1991). *Endangered Languages*. Oxford and New York: Berg.

Romaine, Suzanne (1994). *Language in Society*. Oxford: Oxford University Press.

Samarin, William (1967). *Field Linguistics. A Guide to Linguistic Field Work*. New York: Holt et al.

Saville-Troike, Muriel (1989 [1982]). *The Ethnography of Communication: An Introduction*, 2nd ed. Oxford: Blackwell.

Schütze, Carson T. (1996). *The Empirical Base of Linguistics*. Chicago: University of Chicago Press.

Serzisko, Fritz (1992). *Sprechhandlungen und Pausen*. Tübingen: Niemeyer.

Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Sherzer, Joel (1987). A discourse-centered approach to language and culture. *American Anthropologist* 89, 295–309.

Skutnabb-Kangas, Tove; and Phillipson, Robert (eds.) (1995). *Linguistic Human Rights*. Berlin: de Gruyter.

Starks, Donna (1994). Planned vs. unplanned discourse: oral narrative vs. conversation in Woods Cree. *Canadian Journal of Linguistics* 38, 297–320.

Stross, Brian (1974). Speaking of speaking: Tenejapa Tzeltal metalinguistics. In *Explorations in the Ethnography of Speaking*, Richard Bauman and Joel Sherzer (eds.), 213–239. Cambridge: Cambridge University Press.

Thrane, Torben (1980). *Referential-Semantic Analysis*. Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs; and Romaine, Suzanne (1985). Some questions for the definition of "style" in socio-historical linguistics. *Folia Linguistica Historica* 6, 7–39.

Watahomigie, Lucille J.; and Yamamoto, Akira Y. (1992). Local reactions to perceived language decline. *Language* 68, 10–17.

Wilkins, David P. (1992). Linguistic research under Aboriginal control: a personal account of fieldwork in Central Australia. *Australian Journal of Linguistics* 12, 171–200.

Wolfram, Walt; and Schilling-Estes, Natalie (1995). Moribund dialects and the endangerment canon: the case of the Ocracoke Brogue. *Language* 71, 696–721.

Woolard, Kathryn A.; and Schieffelin, Bambi B. (1994). Language ideology. *Annual Review of Anthropology* 23, 55–82.