POSITION PAPER

Adequacy in documentation*

Anna Berge

As more and more languages are becoming endangered, our notions of what it means to adequately document a language are changing. Further, while some languages remain undocumented, dictionaries, grammars, and texts are available for more and more languages, enabling researchers to broaden the scope of their documentation efforts. This article examines the question of adequacy in language documentation from a number of perspectives and proposes some general guiding principles for documentation efforts. Points of discussion include the interplay between documentation and description; the potential for diversity in all aspects of documentation, such as diversity of linguistic data, consultants, fieldworkers, and products; and the interaction of the different participants in documentation efforts.

1. Introduction

Not so very long ago, a language was considered well documented if its corpus of materials included a good and comprehensive dictionary, a good and comprehensive grammar, and a sizable number of (usually narrative) texts. The adequacy of this level of documentation has been called into question, especially with the rapid loss or endangerment of many of the lesser-known languages, and most scholars now concede that for a language to be considered well documented, rather more documentation is needed. The question is, what does it mean to document a language *adequately*? This question is in fact fairly complex: it requires a discussion of what "documentation" means, what purpose "adequate" documentation of a language serves, and who participates in language documentation. It is so broad a question that it is difficult to address all of the necessary issues in one paper, although I will touch on some of them in the discussion below. Since

* This material is based upon work supported by the National Science Foundation under grant 0349368. Research on the Aleut language is ongoing, results have been regularly presented at national and international conferences, and publications have appeared or are forthcoming.

language documentation is the theme of this volume, many of these points are addressed elsewhere, and from different points of view.

In this paper, I will address the following points specifically, in order to determine what constitutes adequacy in documentation: what the term *documentation* means; what role description plays in linguistic documentation; what should be documented; who the documenting is for; who gets documented, and who does the documenting; and what adequacy in documentation means for the work of linguists in particular. I also briefly address the purpose of documentation and how this affects attempts to obtain adequate levels thereof, but this point has been copiously addressed elsewhere, and it is also partially contained in the other points I raise. It becomes obvious in the discussions of these points that adequacy in documentation is an ideal and a goal more than a reality. Nevertheless, we can derive some general guiding principles, many of which suggest that, ultimately, adequacy in documentation means letting go of preconceived notions of what it means to document a language as a linguist.

2. What is documentation?

It would not be wrong to say that by language documentation, we mean any record of language usage; in reality, linguists generally view language documentation more narrowly, as work undertaken by an amateur or professional linguist that results in some record of a language for the purpose of furthering the study and understanding of that language. In the past, this would have been understood to involve some fieldwork and to result in products such as dictionaries, grammars, and texts, as mentioned earlier. However, the notion of what constitutes documentation is changing. Himmelmann (1998) has claimed that documentation cannot be considered adequate, or even to be documentation at all, if by that term we mean only the products of research. He advocates the separation of data collection from description, which, as many have pointed out, is a form of analysis, and he identifies the collection, preservation, and dissemination of the primary data as the act of "documentation" (Himmelmann 1998), rather than the linguistic analysis resulting from fieldwork/language collection efforts. (This perspective may be seen as part of a very old debate in the history of linguistics, between descriptive and analytic schools of thought: the former tend to favor massive data collection and description of the data and the latter more theoretical analysis of the given data.) There are good reasons for making this point: historically, there has been a tendency for linguists to depend more heavily on published analyses of data than on the primary data themselves. Especially in the generative and postgenerative grammar period, many descriptions have tended to be opaquely theory dependent and therefore most readily understandable only during the most popular period of the theory being represented. Even traditional grammars assume theoretical positions that make later use of these grammars difficult, such as using improvised terminology, terminology based on another language or linguistic tradition, or terminology employed in nonstandard ways to describe lexical or grammatical categories from the particular point of view of the author. As an example, I cite Kleinschmidt's (1851) grammar of West Greenlandic. This was a stunning, avant-garde work in its time that is still widely referenced today, but, having no larger context within which to discuss ergativity, he proposed terminology and lexical categories for Greenlandic that never became standard within studies of the language. Descriptions invariably filter the data and impose organization on them based on presuppositions and theoretical assumptions of the authors. It therefore makes sense to collect, store, make accessible, and use unedited, original data as the sources of linguistic analyses and descriptions. The suggestion that documentation implies the responsible disposition of the data being collected is noteworthy: it has rarely been a priority among field linguists during their working lifetimes, although it is increasingly seen as vital to documentation efforts.

Himmelmann's work (1998; also in Gippert, Himmelmann, and Mosel 2006), as well as that by Woodbury (2003), Austin (2007), and others, has given rise to serious attempts to codify and unify documentation efforts and create a theory for the practice of documentation in its narrow sense. Elements of the theory include the direction of efforts toward the collection rather than analysis of linguistic data (with the exception of transcription and translation, which are included in collection efforts), the collection of as much data as possible, the standardization of the metadata to describe the data (e.g., information about the time and place of collection, the participants, the topic of the collection session, etc.), and the standardization of data storage/database entry.

Widespread efforts to implement these ideas would, in theory, be good, but it is impractical to assume that professional linguists will stop analysis in favor of merely collecting and storing data, or that academic positions will be made available on the basis of purely documenting activities of this nature. Kaufman (2001: 278) sees this view of documentation as a set of research procedures rather than as a discipline. It is more likely that large numbers of data-management specialists would take on this type of task. It also suggests a certain level of desperation in the documentation of languages and would almost certainly result in other gaps in knowledge, such as the knowledge that comes with experiencing the life, context, and interpersonal dynamic of the collection effort. This approach to documentation therefore must be considered as a starting point for any discussion on the topic, but it is not sufficient, as I suggest later in Section 3.

Implicit in this discussion is the idea that "documentation" refers to purposeful recording of a language, that is, data collection with the intent to document a language, as opposed to a more accidental recording of a language. For the most part, I will assume this viewpoint in the context of a discussion of linguistics, the work of linguists, and the documentation of endangered languages; however, this restriction is problematic, and it could be argued that real adequacy in documentation ultimately requires the inclusion of random language recordings if they exist (see Section 4).

– Adequacy in documentation means at a minimum being responsible for good practices in the collection and preservation of primary data and its associated metadata, and in making the data accessible.

3. What is the role of description vis-à-vis documentation?

Defining documentation solely as the collection and preservation of primary data is problematic and controversial, particularly since it is practically impossible to divorce data from some level of analysis. Even Himmelmann's and Woodbury's understandings of the term "documentation" are modified by the inclusion of basic transcription and translation of the data, which itself assumes some analysis of those data. The very act of collecting data cannot be assumed to be impartial; the questions asked and the answers to them include presuppositions about the data, the purpose of the data-collecting session, the context, the assumptions formed during the session by both the language expert and the documenter, and so forth.

Even given a large corpus of data, we may not have enough information to interpret the data without analysis. Thus, it is difficult to know whether all the linguistic forms and structures have been represented by the available data, whether paradigmatic gaps are intentional or rare, and what types of linguistic elicitations are needed to fill out the corpus of data. For example, coordination in Aleut (or *Unangam tunuu*, as the language is now preferentially called by community members) is rare in texts, but there are many options for expressing coordination, and the choice of options depends on speaker preferences, dialect, and context. Most speakers do not use all available choices, nor are all choices found in all types of discourse. Further, even when a large corpus provides a wide variety of textual and grammatical styles, without analysis, we may not have a basis for understanding the pragmatic and social meanings of linguistic forms. For example, one of the many options for signaling disjunction in Aleut is also a method of signaling polite disagreement, but it is subtle and is infrequent in many textual styles of speech. Likewise, we have a large collection of traditional narrative texts in Aleut, but it is not clear which texts are puns and riddles and which are simply stories, nor are the metaphors clear to modern speakers. In other words, we may understand the words but not the sense of many of these texts.

Finally, raw data are often hard for anyone other than the collector to use. For community members and nonlinguists, raw data can be frustrating and opaque, and if a language should die out, raw data may no longer be interpretable even to future linguists. Himmelmann and Woodbury recognize these points and suggest that descriptive and explanatory material be incorporated into the proposed database of documentation in the form of annotations in appropriate places, rather than in an organized, self-contained work such as a grammar or dictionary; in fact, Woodbury (2003) writes that grammars and dictionaries should not be seen as end products but rather as part of the apparatus of documentation. A further suggestion is that corpus observation is best done in conjunction with metalinguistic discussion, e.g. contextualizing elicitations during the elicitation session, or eliciting speaker interpretations of the data; this component of documentation is often lacking in the published, polished descriptions that are made available. Nevertheless, both authors view the primary responsibility of the field linguist as documenting rather than describing data. Lehmann (1999: 11), however, observes that it does not take a linguist to produce raw data; documentation by linguists includes representations of data, such as phonetic transcription, interlinear morpheme glosses, translations, and so forth, all of which inherently suggest analysis (and all of which are included in "documentation" by Himmelmann and Woodbury). In theory, documentation and description may be viewed as independent, but, in practice, they cannot be separated in the work of a field linguist.

– Adequacy in documentation must include concomitant description, although we can propose some caveats. Although choices in description inevitably include theoretical choices, overly opaque and theory-dependent analyses should be avoided at this stage. Description should enable further documentation rather than focus on theory.

4. What gets documented?

What do we document when we document a language? We can approach this question in a number of ways. In the following discussion, I look first at the form of the linguistic data being collected – such as lexical items, grammatical structures, and types of texts – and then at the variety of the language being treated, e.g., the dialect, the register, the style, and so on.

Regarding the form of the data, we continue to see prototypical documentation/description as consisting of dictionaries, grammars, and texts, supplemented both by recordings of language in use and more and more by Web-based archives. One model of large-scale documentation described by Kaufman (2001) results in precisely these products. In community-based language documentation efforts, the results are often returned to communities in these forms (as well as in versions specially designed for language learning purposes). Traditional documentation of this type is essential, especially for previously undescribed languages. It is enormously time consuming and difficult to generate grammars and dictionaries, which is one reason for continuing to focus on these particular aspects of linguistic documentation. However, more and more languages now have some level of documentation, and it would be irresponsible not to expand the repertoire of linguistic materials.

Adequate documentation/description must include a much greater variety of grammatical data, including a variety of textual genres, speech acts, figurative language, intonation patterns, and so forth, if for no other reason than this: some constructions, terms, etc. are not found in all discourse types; consequently, documenting only one or two types of texts almost certainly results in incomplete documentation of the grammar. Some of the biggest gaps in documentation of understudied languages are found in the newer or perhaps less well-understood fields of linguistics, such as discourse, semantics, cognitive linguistics, and prosody. These fields also contribute to the understanding of a language's structure in sometimes unexpected ways. In my work on Aleut, for example, I have described distributional requirements involving verb moods in clause chains by comparing different textual genres, and I have documented metaphors and metaphor chains by investigating scope of meaning in various lexical items.

Obtaining grammatical variety is more easily said than done. Some types of data, such as narrative texts, are simply more easily collected and analyzed than others; as a result, they are by far the most commonly documented, often to the exclusion of other textual types. Lack of fluency in the language being documented is another reason for lack of diversity in what gets documented: for any field linguist who does not have relatively fluent conversational abilities in his or her language of research, collecting and analyzing colloquial language may not be feasible. Further, a language may simply lack basic tools, such as a dictionary or grammar, which would enable a researcher to conduct more varied documentation. Finally, many documentation efforts are part of much larger initiatives to preserve and revitalize languages, as well as to create products for language communities, all of which tend to focus on traditional stories rather than newer uses of language, and on the rapid production of learning materials such as vocabularies and grammatical sketches. All of these result in the perpetuation of traditional approaches to documentation and description, including in my own work. For almost two decades, for example, I pursued fieldwork with the underlying ideology that language should be studied in context (e.g., syntactic analyses should rely on tokens situated in their discourse context), and that linguistic analyses should reflect actual spoken language. However, until some years ago, most of my data came from narrative texts supplemented by elicitations, with the problems listed earlier playing more or less important roles.

Diversity in what gets documented has often been a luxury that few under-documented languages benefit from. In the past, much of the diversity of linguistic material collected for these languages has not come from the efforts of linguists documenting the language for theoretical purposes, including the production of traditional descriptions. Rather it has come from linguists, ethnographers, anthropologists, and educators working on nontheoretical aspects of language use, such as rules of etiquette in ritualized greetings, introductions, oratory, topics of conversation in everyday life, or language learning materials requiring the collection of a wide variety of basic, often idiomatic, expressions. My shift from collecting and working with narrative texts to the inclusion of conversations, instructions, songs, tokens of motherese, and so on, came when I started working with language revitalization efforts and, in particular, with the production of adult language learning materials. This said, theoretical research on endangered languages has broadened noticeably in recent years to include investigations into discourse, pragmatics, prosody, figurative language use, and more, as evidenced by a rise in relevant presentations and publications, as well as by the rise in interest in the topic of this paper itself, adequacy in documentation.

To put the role of linguists further in perspective, diversity in documentation in general has come from sources other than linguistic documentation efforts: thus, the better-documented languages (e.g., English, Spanish, Russian, French, Chinese, Latin, Ancient Greek, etc.) have enormous corpora of data from written and oral recordings, including newspapers, novels, graffiti, movies, news programs, and so forth, the vast majority of which did not come from linguists. This, of course, brings us back to the definition of documentation: documentation for linguistic purposes is only one form of language documentation. To be sure, linguists play an enormously important role in the documentation and preservation of the lesser-known, less well-studied, and endangered languages of the world, but the task of adequately documenting a language cannot be undertaken by linguists alone (Section 7).

In addition to diversity in linguistic structures, we must also have diversity in sociolinguistic materials, such as records of different dialects, linguistic styles, registers, and even speakers. For example, there is a tendency to document only certain dialects of a language, whether due to easy physical access to those dialect areas, particular relations with speakers, or sociopolitical or socioeconomic considerations. But the choice to document one or another dialect often has important ramifications in the future development and/or study of the language, regardless of whether the dialect comes to be accepted as a standard. The neglect of other dialects will then restrict understanding of the language's historical development and any cultural insights that come with that, not to mention the loss of status that members of an undescribed dialect suffer (see, for example, Adelaar 2001). The effects of social status on the documentation of speakers is discussed more in Section 6.

– Adequacy in documentation must include all the basic phonology, morphology, syntactic constructions with context, lexicon, a full range of textual genres, registers, and dialects, and data from diverse situations and speakers, with enough tokens and a large enough database to be useful in both present and future research. This point has been made by numerous others in the field, although it is only slowly beginning to be implemented in practice.

5. The role of uniqueness in documentation

The role of uniqueness in documentation may be considered as part of what gets documented; however, it is worth considering separately, since it involves a different set of assumptions and problems. Every language has some exciting, typologically unique feature(s) for which it is known to linguists. Once unique features have been identified, there has been a tendency to focus on just those features in documentation and description efforts, presumably to obtain enough data to analyze them and understand how they work within the linguistic system. In Aleut, for example, the typologically unique system of anaphoric marking on nouns and verbs, coupled with an unusual plural agreement system, has been the single most researched feature of the language (see Berge 2009, 2010). However, focusing on the uniqueness of a language may cause us as documenters to miss other particularities that need description. Typologically common features of a language will seem trivially obvious but nonetheless require attention and representation in the corpus for full documentation. Furthermore, some seemingly common constructions may have particularities in usage that go unnoticed without detailed investigation. For example, coordination in Aleut, at first glance, does not show particularly unusual characteristics; however, coordinate structures have interesting and nonnegligible effects on the use of pronouns, on plural concord, and on other features of the syntax which should be investigated. Lehmann (1999) notes that preconceptions about the centrality or peripherality of certain grammatical points have resulted in certain areas of grammar consistently receiving no treatment.

– Adequacy in documentation must include both unique and common constructions, words, and so on, in the language.

6. Who gets documented?

In the documentation of endangered languages, there is often a tendency to focus on a select rather than broad group of speakers. There are many reasons for this: some speakers are better than others; some have more social prestige and are therefore more accepted as representative speakers; and some speakers are easier to work with than others. In addition, some forms of documentation are more easily undertaken, such as collecting vocabulary and narrative texts, as opposed to natural conversations, family interactions, context-specific interactions, etc. The group of speakers documented therefore tends to include older, socially prominent, fluent speakers, and good narrators, and to exclude semispeakers, socially stigmatized individuals, and even child speakers (who are generally less patient with elicitations, less capable of stylistically interesting variations in speech, less likely to produce graceful narratives, and less likely to know specialized vocabulary).

However, focusing on the pure, traditional, older speech or language limits the adequacy of documentation in important ways. For example, ongoing trends in the development of the language may be missed, some of which may be a reflection of very old processes. Thus, syncopation/apocopation in Aleut has been a feature of the language since it was first documented and appears to have had ongoing relevance and ramifications to the grammar. These include increasing differentiation between Western and Eastern dialects of the language in historic times (before its endangerment); this process is still in progress today, especially among speakers from the Pribilof Islands, and the different stages of syncopation are plainly visible by comparing the speech of members of different social classes. Another example is that focusing on traditional subjects tends to result in ignoring slang, jargon, or other forms of specialized language of older as well as younger speakers. For example, terms and phrases related to modern life (money, bank, bars) have been in use in Aleut since before the period of severe endangerment; however, many have not been documented, and there is a distinct preference for collecting Russian-period loans, rather than modern English loans, and even a preference for collecting later Russian religious terminology over recent English borrowings of religious terms. Such selective documentation restricts studies of normal language change and may limit studies of actual, synchronic language use at the time of documentation. Overemphasis on traditional forms, whether by the speakers themselves or by the field-worker, prevents some relatively good speakers from comfortably providing data and may lead to their offering over-analyzed and even incorrect forms. Such a practice also undervalues the speech of people of lower social status, who may be even more fluent, and discounts the speech of less-fluent speakers. Nonetheless, data from these speakers can also provide insights into language complexity. For example, in my studies of Aleut clause chaining, a comparison of clause chain in narratives and conversations of fluent and less-fluent speakers shows that the less fluent speakers have replaced narrative with conversational style. This tells us something about stages of language loss, but it has also brought up differences between styles that until now went unnoticed. Finally, focusing on traditional speech can lead researchers to ignore the speech of young children, if the language is still being spoken by children, unless the focus of research is on child language acquisition. Consequently, many forms of speech (e.g., some types of riddles, puns, and word games, as well as normal interactive routines) that are either unique to children or are more likely to be used or remembered by children, are not documented; the same is true of parent-child linguistic interactions. All of these are the results of value judgments on both the proper forms of the language to be documented and on who can best provide these forms.

– Adequacy in documentation should not be limited to what is considered the purest or most traditional form of the language, nor to a particular speaker population. Speakers of all ages, socioeconomic statuses, etc., should be included.

7. Who does the documenting?

Traditionally, field-workers came from outside the community of speakers, and they tended to work in conditions that assumed a difference in status between the speakers and the field-workers themselves, with the field-workers having the higher status. Field-workers have historically had agendas that were independent of the internal needs or desires of the communities; for example, missionaries documented languages with the aim of eventually gaining converts, scholars did so to advance academic knowledge in their fields and their own careers, and so on. Because the ethics of doing fieldwork have changed dramatically in the past decades, more and more documentation is done with community participation, perspectives, and agendas. This point is addressed at length in other parts of this volume, especially in discussions of cooperative efforts between communities and linguists and the role of the linguist in such collaborations. It is also more and more common to find linguistic field-workers who are native members of a given speech community, with wide-ranging motivations and agendas. The group or person doing the documenting, needless to say, has an enormous influence on what gets documented and on the nature of the documentation itself. A linguist with a particular interest in phonology, for example, may very well produce a comprehensive grammar, but the finer points of the syntax may be missed. Likewise, a person trained by a particular school will most probably perpetuate the school's approach to linguistic investigation. A person working with a particular family, group, or village is more likely to focus on the dialect of his or her consultants than the broader linguistic region. It stands to reason, therefore, that the more people of different backgrounds, training and interests can work on a language, the more completely that language is likely to be documented.

Unfortunately, there has often been a tendency for language workers, whether linguists or community language workers, to take ownership of the language being documented, to the degree that they are reluctant to share the information or even share the ability to document the language with others. In some communities, one person may have achieved a certain social status by being identified as the language worker or language expert, whereas another may become ostracized as a speaker or consultant because of objectionable behavior. In many academies, one linguist is considered the known expert on a language and may protect access to the communities in question, or to the data, or even to the wider scientific community. In some cases, the linguist may be perfectly open, but the scientific community may have acknowledged his or her expertise to such a degree that other voices may not be heard. Although some language documenters are remarkably able and productive, no single person is capable of documenting, let alone describing, an entire language. With endangered languages, where there may not be continuing production of language data without the stimulus of language workers or linguists, it is vital that efforts to document the language be openly encouraged of all people interested in doing so.

– Adequacy in documentation requires as many people as possible to take part in the documentation of a language. These people include any community members willing to record language data, students, and established researchers.

8. Who is the documentation for?

Documentation is not a neutral activity; it is generally conducted with particular needs and agendas in mind for a particular audience. Thus, documentation by academics (e.g., linguists, ethnographers, anthropologists) has, of course, tended to be done for academics; the major results of linguistic documentation have been academic papers, descriptions, and theories. Documentation by missionaries and colonial authorities has tended to be for other missionaries and colonial authorities, either for language learning efforts or for conversion or subjugation. Documentation within communities has tended to be for nonlinguistic purposes (e.g., dissemination of information, entertainment, conservation of ritual or religious custom) but sometimes also for revitalization. In each case, the type of documentation that results is different, and it is good to keep in mind that different needs and agendas result in vastly different types of data.

Today, with increasing involvement of different groups in language documentation efforts, multiple needs and agendas must be negotiated. Each group brings a different idea of adequacy in documentation to the table. Further, there may be multiple ideas and agendas within a single group about language documentation. If we assume a negotiation between linguists and communities (and leave aside nonpurposeful language data collection at this stage), there may be linguists with quite different priorities or theoretical approaches to documentation working together, as well as community members with equally different agendas and desires. In some cases, there may even be restrictions on what is allowed to be documented; this is often true of religious texts or ceremonies, for example. Nevertheless, despite current sociopolitical views, needs, restrictions, and stated intentions, most communities and linguists converge in their desires to see as broad a representation of the language documented as possible.

Linguists and communities especially differ in their requirements of the form, accessibility, and cultural sensitivity of the materials being documented. Thus, as mentioned in Section 2, the vast majority of raw field data – whether in the form of handwritten notes, audiovisual recordings, or digital materials – are not easily interpretable or usable without extensive accompanying annotations. While most researchers would agree that relevant metadata such as date, subject, and participant names should always be attached to the accompanying field data annotations, this is not always implemented in the field. Further, most descriptive and theoretical products of language documentation by linguists have tended to be inaccessible to nonlinguists, including to the very community members participating in documentation efforts. In effect, neither the raw data nor the end products of linguistic fieldwork are easily accessible to the communities, and from a nonlinguist's point of view, this calls into question the adequacy of the documentation.

Finally, there is the important question of who controls the analysis of the data. Although it is necessary to look for linguistic or sociolinguistic patterns in language data, the interpretations of the findings are sometimes objectionable, offensive, and wrong, or they may be perceived as such. In communities with large numbers of people creating their own linguistic and cultural documents (e.g., the English-speaking world, the Yiddish-speaking world, etc.), these interpretations are investigated, discussed, and perhaps eventually rejected by common consensus within the community. In most communities of speakers of endangered languages, or in preliterate or newly literate communities, the opportunity to adjust or correct objectionable or faulty interpretations is limited. Sharing the responsibility and control of the findings and the resulting interpretations of the data is therefore not only vitally important for successful collaborative documentation efforts, it also results in better analyses.

Ultimately, documentation is for community members, future generations, laypeople, and scholars within and outside the community of speakers; in short, it is for anyone interested in the language and culture in question.

– Adequacy in documentation must address the needs of nonlinguists, particularly the needs of the users of the language being documented, as well as the needs of linguists. Further, since the products of documentation differ depending on the agendas of those who create the documentation; diversity in agendas leads to more complete, accurate, and thoughtful documentation.

9. What does "adequacy in documentation" mean specifically for the work of linguists?

If documentation can and should be undertaken by everybody, and should include everything, then what, specifically, should linguists focus on? Should field linguists have a prescribed set of rules for what and how to document a language? How do we best document a language if the goal is adequacy in documentation? These questions are not trivial: we cannot afford not to have a coherent plan of action, especially in the face of massive language endangerment and possible extinction in the coming century. Himmelmann (1998), Gippert, Himmelmann, and Mosel (2006), and others suggest that the task of the linguist is to undertake massive data collection. However, separating documentation from analysis is not ideal, as we have seen (sections 2 and 3), but neither is separating the roles of data collector and data analyst: the trained linguist is unlikely to be content with mere data collection, but the context of data collection provides some grounding for the abstraction of analysis and theory development. Kaufman (2001) proposes large-scale, multilinguist efforts with community language workers. In some respects, the larger a study, the more efficiently and effectively some aspects of documentation can be undertaken; but it is also the case that the more creative and human aspects of language use are more likely to be lost in the scientific methodology of documentation according to a template. Further, while some multi-institution projects are funded every year, the vast majority of linguists are funded for individual fieldwork on small-scale projects. It is therefore unrealistic to expect that either model proposed by Himmelmann or Kaufman can be successfully applied in most cases. At best, one might expect good collaborative and cooperative efforts between individual researchers working on neighboring or related languages so as to maximize the information needed to pursue documentation projects. (Furbee, in this volume, 3–24 offers an alternative solution to the individual linguist bearing the burden.)

The tradition to systematically describe the phonology, morphology, and basic syntax of a language, then to produce a dictionary and record texts arises from just such individual efforts – these are necessary first steps in understanding languages with no or little prior documentation and description. This model of documentation can only really be considered inadequate when we have the luxury of being able to go beyond these basic tools. In languages with these tools already available, the linguist is in a position to branch out and begin documenting and describing other aspects of the language being studied; and as we learn about more complex aspects of language and language use, we need to incorporate these features in our documentation (Section 4). The role of the linguist, therefore, is to systematically document and describe a language at whatever level is appropriate for a given language. Although there are still languages today with little or no documentation, this state of affairs is increasingly rare; and it is time to consider systematizing documentation efforts of other, more complex levels of linguistic structure and usage.

The systematization of documentation beyond the already agreed-upon dictionary, grammar, and texts is complicated by the overwhelming surge in data, types of data, and fields of linguistic inquiry, and the ensuing need to organize, store, and incorporate the data in some coherent way. For example, documentation is greatly helped by new advances in technology, which allow us not only to document more but also to store and make accessible more data than ever before. What we expect out of documentation efforts is far greater than at any previous time, and our expectations change with each major advance in technology. Just a decade ago, fieldwork might have consisted of notebooks, typed computer files, and audio records (tapes as well as digital records). Fieldwork today almost invariably assumes digital (rather than cassette) audio and video recordings, as well as accompanying notes (most probably also in digital form). Whereas many recordings never found their way to permanent storage facilities, and others only did so decades after the data were collected, it is now common to consider (and in many cases require) concurrent archiving, especially digital archiving of the records being produced. There are more and more efforts to unify digital archives (and these are discussed elsewhere in this volume Thieberger and Jacobson, 147–158), so that documentation efforts have more global relevance. However, these efforts are still relatively new, and it will take time to sort out the effects and ramifications of the explosion in data, types of data, etc.

At the same time, our field of enquiry is limited by our understanding: who would have imagined the importance of documenting mundane conversational speech a century ago? Such gaps will undoubtedly be found in future documentation efforts as well. The problem with rigidly prescribing what needs to be documented is that we risk limiting the data or the applicability of the data. Some previous attempts to prescribe what linguistic data should be presented and how they should be cast have resulted in grammars that are hopelessly opaque and do not best reflect the language being documented.

There are no definitive solutions to this dilemma, but the discussion should continue, and the more discussion, the better. Meanwhile, creative approaches to documentation should also be encouraged, and we should approach linguistic studies and fieldwork with fresh eyes, an open mind, and a willingness to pursue a point of interest that is not in the mainstream.

– Adequacy in documentation by linguists should be based on current advances in linguistic understanding and technology, it should include adequate and timely efforts to preserve data and make them available to the greater world, and it should include creative approaches to the study of linguistics.
– Adequacy in documentation means letting go of preconceived notions of what it means to document a language as a linguist.

10. Conclusion

The issues raised in this paper are largely familiar to language documenters, and they should be considered measures of the difficulties involved in the development of a new discipline rather than a critical evaluation of current theory or practice. Real adequacy in documentation comes with flexibility regarding what of a language gets documented and how that language gets documented. It requires large and diverse numbers of participants, large corpora of data, and a cooperative approach on a multitude of fronts: between communities and researchers, between funding sources and documenting teams, and between older and younger generations of speakers or researchers or both. It also comes with letting individual contributors make their own contributions. There are ethical implications to each of the points above and to each of the decisions we make to document or not document. Therefore, achieving adequacy in documentation requires a continuing dialogue about what adequacy actually means.

In addition to stating what should be considered an adequate level of documentation, it would be good to have a measure thereof. One measure of adequacy in documentation and description might be how learnable the language is as a result, since acquiring fluency in a language requires enough data with enough descriptions to reproduce the language outside its normal context. Other measures of adequacy should also be sought.

It is also worth considering the practical reality of trying to achieve adequacy in the documentation of all languages. Of the nearly 6,000 languages estimated to exist today, over half are believed to be seriously endangered, and most are not considered to be adequately documented. There are too few people to do the work, and there are many models of documentation available to follow, not one of which is likely to be sufficient to document all languages in need. Attempts to standardize documentation models are often a response to some of these constraints; hence lists of criteria, proposals to focus on documentation rather than description, the race to document and digitize as much as possible, the assumption (which I do not dispute) that the best chance for adequate documentation is to get as big a corpus of data as possible, and so forth. Unfortunately, there is no magical solution: adequate documentation can only come with time and resources, including the people to do the work. We may know what adequate documentation is, but we do not necessarily have the means to attain such a state for all languages. Nevertheless, we are obligated to document as much as possible.