
Reviewed by Nicholas Evans, Australian National University

This book is a landmark. It is now ten years since Nikolaus Himmelmann’s seminal article Documentary and descriptive linguistics appeared. The intervening decade has seen a growing awareness among field linguists that it is important to distinguish language description from language documentation, so as to provide linguistic data for future researchers, in a way that ensures that:

- the quality of primary data is carefully and constantly monitored and documented, that the interfaces between primary data and various types of analysis are made explicit and critically reviewed, and that provisions are taken to ensure the long-term preservation of primary data so that it can be used in new theoretical ventures as well as in (re-)evaluating and testing well-established theories (p. v, Essentials of language documentation)

Himmelmann’s 1998 article was followed by a rapidly growing movement to develop the techniques and approaches of documentary linguistics and to step up major internationally-funded language documentation programs. Most notable among these are the Volkswagenstiftung’s Documentation of Endangered Languages (DoBeS) program and the Hans Rausing Foundation’s Endangered Languages Project (HRELP), but there have also been important reorientations of funding programs in individual countries, including recent major government-funded research programs in the USA, Brazil, and China, and the BABEL initiative in the European Union. What this movement has lacked, until the appearance of the book under review, has been a single-volume handbook bringing together the ideas and techniques needed to form the practice of would-be documentary linguists. The book is based on an international summer school organised in Frankfurt in 2004 by the three editors, all veteran field linguists. Support from the Volkswagenstiftung enabled them to assemble a veritable A-team of contributors, who were able to put their ideas to the test in a teaching setting before setting them out as chapters in the present book.

Overall the presentations are remarkable for their combination of thoughtfulness and practicality: the volume as a whole strikes the right balance between ideas and applications though the proportions of these elements vary widely between chapters. The open, well-illustrated layout makes the book deceptively readable and approachable given the depth and detail of its content. Occasionally there is a bit too much overlap between chapters (e.g., between chapter 13 on archiving and chapter 4 on data and language documentation), and less cross-talk than one would wish for among the authors. The ordering of chapters is also odd: the chapter by Austin and the chapter by Trilsbeek and Wittenburg contain closely related content but sit at opposite ends of the book, while the set of closely related chapters on ethnographic issues is interrupted by Himmelmann’s unrelated chapter.
on prosody. In these places the book would have benefited from a tighter editorial rein. On the other hand, that individual chapters are self-contained is an advantage if, for example, they are set as individual readings in a field methods course.

Himmelmann’s opening chapter, “Language documentation: What is it and what is it good for?”, sets the scene for the volume, restating and updating the general challenge his 1998 paper outlined for a newly defined subdiscipline of documentary linguistics. One problem that documentary linguists face is to determine where more structured methods are necessary so as to cover all relevant combinations in the more paradigmatically-organized parts of a language. Although documentarist approaches have sometimes been accused of downplaying or rejecting the importance of this sort of data, I believe this is based on a misunderstanding of how language documentation proceeds. On pages 8 and 9 Himmelmann makes clear that he is not arguing for the omission of such domains, but merely for the full documentation of how they are obtained. I am not sure, however, that it is helpful to lump this in as “metalinguistic knowledge,” since a more prototypical reading of the latter term focuses it on those aspects of language that overtly name and consciously theorize about language functions, meanings, and structures—such as the fascinating Kuikuro examples in Franchetto’s chapter (see below). I also find problematic his rather overstated concern with discouraging the investment of effort that goes into publishing reference grammars rather than getting more primary documentary material—but will defer discussion of this point until the latter part of this review.

Arienne Dwyer’s chapter, “Ethics and practicalities of cooperative fieldwork and analysis,” does a good job of making explicit the minefield of diverging expectations through which any fieldwork project must navigate a path. These range from the community’s expectations to those of university ethics permits or institutional review boards and those of national or regional agencies responsible for granting research committees, and on to the expectations of research funding agencies that may see the production of language or other educational materials for community consumption as a diversion of resources away from the purely scientific issues they wish to support, whereas for the communities themselves this is the most valued and motivating outcome. This chapter would have benefited from more references to existing case studies and discussions of the problems outlined; see, for example, Wilkins 1992, Gerdts 1998, and the clear discussion of intellectual property issues in Anderson and Koch 2004.

Dwyer is rightfully critical of the often unsuitable assumptions made by institutional ethics committees because of their historical origins as regulators of biomedical research, where the nature of both research and relationship is very different. Unfortunately this discussion gets buried in a footnote on page 62, although it merits a more central place in the article. A recent article by Widlok (2008) takes these arguments further.

Dwyer enunciates five ethical principles that should guide linguistic field research: (1) Do not harm, (2) Reciprocity and equity, (3) Do some good (for the community as well as for science), (4) Obtain informed consent before initiating research, (5) Archive and disseminate your data and results. Though I agree wholeheartedly with the emphasis Dwyer places on viewing linguistic research as a long-term commitment, both scientifically and

---

1 A further oddity is that just one chapter (Dwyer’s) has its own reference list, while all others are consolidated in the back of the volume.
in terms of human relationships with the community, the chapter does less than it could to show how a more long-term view of research can resolve some of the conflicts she outlines—for example by conceiving of time-linked clusters of projects oriented toward achieving different goals. There are an increasing number of organizations committed to supporting building community capacity to incorporate local knowledge and language into education and other ventures, and these can be an appropriate source of funding for developing community-relevant products of more basic research. Adding a sixth principle that “field languages require long-haul commitments,” with the corollary that researchers should have their own long-term plan overarching the contexts of particular projects, would help work out how to resolve some of the conflicts outlined in the chapter.

In an era when fieldwork is becoming more and more a complex cooperative endeavor involving outside researchers and local team-members in a range of roles, Ulrike Mosel’s chapter, “Fieldwork and community language work,” is a welcome contribution, discussing the pitfalls that can develop from the divergent perspectives of linguists and local language workers. This chapter is likely to become required reading for anyone about to embark on setting up a collaborative project of this type, helpful in terms of foreseeing the project management and workflow kinks that can arise in a complex intercultural team. It also contains some interesting alternative methods for structural elicitation. The following chapter by Peter Austin on “Data and language documentation” is equally practical and clear, focusing on the data-management side of the workflow.

Of the next four chapters, three return to more conceptual themes at the culture-language interface: ethnography of language (Jane Hill, chapter 5), semantics (John Haviland, chapter 6), and ethnography more generally (Bruna Franchetto, chapter 8). Any old-fashioned pencil-and-notebook lone fieldworkers groaning at all the project- and data-management hoops that the Austin and Mosel chapters try to jolly them through should perk up again at this bracket, where the emphasis is less on technical organization and mastery and more on consciousness-raising and heightening the fieldworker’s capacity for interpretation. Haviland begins, “I concentrate on a series of doubt-producing obstacles for the field lexicographer, with some suggestions about how at least to address, if not to overcome them” (129); as he puts it, getting a documentation worthy of the language “relies on both drudgery and ethnographic inspiration, on systematic elicitation and serendipitous discovery” (160). I couldn’t help noticing that, in the structure of the book, it seemed to be the linguists who were put in charge of the drudgery chapters and the anthropologists who got the inspiration chapters.

Jane Hill’s incisive and powerful “The ethnography of language and language documentation” is vintage Hill flying along at her best. She proposes that “documentary linguistics takes up a vision of the integration of the study of language structure, language use, and the culture of language” (113). This prefigures (I think not unrealistically) a rapprochement between linguistics and a reinvented anthropology that has returned from its postmodern holiday ready to engage with falsifiable empirical data again.

In return, “documentary linguists need to be ethnographers, because they venture into communities that may have very different forms of language use from those of the communities in which they were socialized as human beings or trained as as scholars” (113). (As the reader will notice, an exhausting effect of this book is realizing how many things one “needs to be” in order to be a documentary linguist: about the only thing one doesn’t need
to be, it seems, is a grammarian—on which more below. It might have been worthwhile to preface the book with Voltaire’s famous advice that *le mieux est l’ennemi du bien*—the best is the enemy of the good—to avoid scaring too many readers away with the feeling that they could never manage to satisfy all the “need to bes” set out in this book). Hill’s richly-worked anecdotes about working with Nahuatl, Cupeño, and Hopi speakers and the way their cultural backgrounds shape the teaching/documentary encounter are stitched into a deeply serious argument about the need to incorporate a cultural and ethnographic understanding of language into the foundations of research, adding a sense of color and humanity that can readily go missing from the earnest documentarist agenda.

John Haviland’s equally fascinating chapter, “Documenting lexical knowledge,” is full of instructive examples about how to find the particularity in a language’s meaning-system, and is a masterpiece of pedagogical compression. Despite giving an enormous range of investigatory techniques he concludes that “the only remedy [for cracking the more difficult parts of the language system, such as tropes] seems to be wide ranging and systematic ethnographic attention” (157) and that “calling a halt by declaring the database closed is simply an arbitrary rest stop on a very long journey” (161). Unfortunately the only references to “duration” in the book’s index refer to durations of recording and prosodic phenomena. The question of how long documentation projects need to be to accomplish the ambitious goals set out in the book could well have been addressed more systematically with some hard facts about what takes how long to achieve, particularly as the demanding new standards of documentary linguistics substantially drive up the time it takes to process data in the initial phases of research.

Bruna Franchetto’s chapter, though it doesn’t follow on directly from Hill’s and Haviland’s, belongs with them conceptually. The whole vast realm of culture can sometimes seem just too big to reduce to manageable investigatory dimensions, and linguists without anthropological training (the vast majority nowadays) can feel overwhelmed with the problem of knowing what to start asking about. Her chapter grounds the problem by providing a useful and stimulating set of ethnographic topics, including kinship, body parts, material culture, subsistence activities, shamanism, “standard topics in ethnography,” onomastics, and toponyms. Two particularly interesting sections deal with native metalinguistic discourse, throwing up the beautiful Kuikuro term *tisakisü enkgutoho* for deictics and particles, literally ‘made for our words to beach safely’, and with the importance of including sessions showing verbal interactions between native speakers and foreigners in other languages than the one(s) being documented, important for remedying the neglect of field studies of second language acquisition in face-to-face societies and on the shaping of language through contact.

Intonation and prosody are usually the most elusive aspect of a language’s phonology and the hardest area to gain systematic data on. Nikolaus Himmelmann’s chapter 7, “Prosody in language documentation,” resumes the more hands-on, practical style of the early chapters and gives some important pointers on getting a prosodically informative corpus together, including ways of making sure lexical tone is dissociated from intonational contours, and appropriate elicitation methods for changing the information structure of elicited utterances so as to get a wide range of intonation types.

---

2 Though he could usefully have mentioned Hellwig’s (2006) article on field semantics.
Chapters 9 and 10, by Eva Schultze-Berndt and Nikolaus Himmelmann, respectively, continue the hands-on vein. Schultze-Berndt deals with how to annotate texts—what layers to employ, what software to use, what grain to annotate at, and the importance of maintaining standard transcriptions alongside any more fine-grained but deviant representations to facilitate later searching through the corpus.

The discussion of how much to annotate, at the end of the chapter, could usefully have cross-referred to page 122 in Jane Hill’s chapter, recounting the sequence of elation and dismay felt by her student on receiving Ken Hale’s Mountain Pima fieldnotes and finding that the glosses gave out six pages into the investigation, after which the notes are straight Pima. This highlights the tension in being too prescriptive about how to proceed in documentation. Hale’s free-wheeling style allowed him to race after inspired hypotheses through the heat of elicitation, and penetrate as deep into the structure of the language as fast as any fieldworker ever has. Too much insistence on always doing things the way this book advocates can end up stifling this sort of excitement. Yet, as Hill goes on to show, an unannotated legacy like this can leave future investigators scratching their heads.

Himmelmann’s following chapter deals with “The challenges of segmenting spoken language,” both into words and into intonational groups. Reading these two chapters, one realizes what a small proportion of linguistic training is devoted to these issues (unless the student is lucky enough to be taking courses on discourse or conversational analysis). Taken together, these two chapters fill that gap, although (like many parts of the book) their point will be taken up most effectively if they are read in conjunction with hands-on work with real, dirty data.

Most documentarist linguists will face the need to develop a workable practical orthography for the language they are working on, and graduates of typical linguistics courses are not usually prepared to consider all possible relevant factors. Linguistics training produces a strong prejudice in favour of phonemic orthographies, which are not always the ideal solution once all considerations have been weighed up. Frank Seifart’s chapter on orthography development is the clearest and most intelligent discussion of the multiple factors involved that I have read. It takes a sensible practical line on such issues as orthographic depth (e.g., the advantages of a morphophonemic rather than phonemic representation) and the defensibility of underrepresenting some contrasts if they carry a low functional load. He also brings in important considerations from the research on reading and writing, such as the fact that the optimal orthography for a beginning reader is not the same as that for a fluent reader, and that there is an asymmetry between the needs of readers and of writers. Given that initial orthographic discussions for an undescribed language will by definition involve beginning readers, these are important factors to take into account—and raise one issue that could have been given more attention, namely the question of whether orthographies should be expected to undergo some change and adjustment in the first years of use, or whether they should be locked in from the start to avoid confusion.

Ulrike Mosel’s chapter on “Sketch grammar” is another useful corrective to the biases of how linguists are taught to work, namely aiming for the high ideal of a complete refer-

3 Though marred by the slightly inaccurate statement that the graphemes of the Japanese kana system “each refer to a syllable” (277)—“each refer to a mora” would be more accurate, so as to include \+/\ for the codal nasal and \+/\ when used to indicate lengthening of the following consonant.
ence grammar. In practice the process of documentation, with its need to simultaneously develop materials on all fronts (dictionary, glossing conventions for texts, evolving phonological analysis), requires practitioners to work up sketch grammars at various levels of detail.

“Archiving challenges,” by Paul Trilsbeek and Peter Wittenburg, is written by the architects of the world’s premier repository of materials on endangered languages, the DoBeS archive, and sets out the philosophy and design behind their creation. Since archives are the main way future users will access their material, it is essential that field linguists think about how their data will look to users of such archives, and this chapter sets out a minimum set of issues that any documentary linguist must address.

Jost Gippert’s chapter, “Linguistic documentation and the encoding of textual materials,” presents a case study of data loss through the vagaries of font conversion, showing how easy it is for an original text version to end up as gobbledygook after it has passed through a few fonts. He argues that the only way to ensure character integrity is for a combination of Unicode and XML-markup of annotations, though even Unicode is vulnerable to font degradation through its “private use area.” In other words, archived texts will only be safe from the sorts of data loss he outlines in his chapter when Unicode is universally adopted, linguists abstain from the private use facility (which will only happen if the Unicode initiative is receptive to creating standardized new characters as needed), and linguistic analysis tools are XML-based—meaning that alternatives to Shoebox/Toolbox will need to be developed.

The book closes with a chapter by David Nathan, “Thick interfaces: Mobilizing language documentation with multimedia.” By “mobilization” Nathan means “taking linguistic documentation and working with speaker communities and other specialists to deliver products that can be used to counter language endangerment” (364). “Thin interfaces,” in his terms, stick close to the data organization of the archived data, while “thick interfaces” are more complex in their implementation but simplify or dramatize access to the base data for community members such as teachers and other language program developers.

Anyone who succeeds in learning all the lessons of this book will bring back a load of tremendously exciting material from their fieldwork—underlain by the power of multimedia recordings to capture vivid exchanges, faces, and settings, plus the broadened ethnographic lens opened up by the Hill, Haviland, and Franchetto chapters in the middle section of the book. So an average-sized documentation program will generate more than enough data for a rich and stimulating language-teaching curriculum. Nathan’s chapter gives some interesting ideas about the sorts of interfaces that can be derived from such material, and that can help justify the hope often expressed by linguists and communities that the documentation process will assist the community in maintaining its language by channelling lively and condensed language material back for more general exposure. He remains non-committal about whether this is one more task for the field linguist, or something that could in principle be undertaken by someone else with a different skill set, and this will obviously depend on the individuals concerned. But what would have been helpful, in a book on language documentation, would have been to say more about how anticipating the needs of language maintenance and revival programs might point to certain types of data-gathering that would otherwise be overlooked.
I would like to pass now from an overview of the book’s contents, which, as I have indicated, amount to a superbly conceived and executed manual in language documentation, to two general points of criticism—targeted as much at general currents in the documentarist movement as to the book itself.

The first concerns the absence of historical awareness. Documentary linguistics is older than its name, and there have been previous periods in the history of our field—such as Fray Bernardino de Sahagún’s sixteenth-century work on Nahuatl, or the work of Boas and his students in North America—which anticipated many of the tenets and methods of the modern documentarist movement. Sahagún’s work, for example, consciously integrated linguistic and cultural documentation, the gathering of a wide variety of genres, the use of illustrations, and the involvement and training of native speakers in a range of capacities. It would be a shame for documentary linguistics to exhibit the same historical amnesia that has diminished other recent dominant schools in linguistics, since we still have an enormous amount to learn from the methods and approaches used by our predecessors—for example, from philological and hermeneutic methods in textual interpretation. And it is an awareness of what the results of these early documentarists failed to bequeath us despite their best efforts that can best sharpen our own attention to filling out the documentary record.

The second criticism is in some ways more serious. In parts of the book one detects a certain documentarist fundamentalism, which surfaces most clearly in Himmelmann’s argument that linguists should foreswear the writing and publication of descriptive grammars in order to devote their full time and energies to the task of language documentation: “with regard to the economy of research resources, it may be more productive to spend more time on expanding the corpus of primary data rather than to use it for writing a descriptive grammar” (24). But it also crops up elsewhere in the book, as in Mosel’s statement in her chapter on sketch grammars that “a thorough analysis can wait until there is time for a specialist investigation” (307). This “let the grammar wait” position is based on the view that, as Himmelmann puts it (24),

The writing of a descriptive grammar involves to a substantial degree matters of formulation … and organization… These are very time consuming activities which in some instances may enhance the analysis of the language system, but in general do not contribute essential new information on it.

The question is, of course, how far the analysis needs to go.

A number of recent and thoughtful accounts have broached this question. The nub of the problem is identified by Rhodes et al. (2006:3–4):

Himmelmann (1998) has argued persuasively that documentation is distinct from what he calls description, i.e., linguistic analysis. We think this is seriously mistaken. In order to know how far along one has come in documenting a language, one must be able to measure how far there is to go. A crucial part of that measurement is found in the accounting function of analysis. How do we
know when we’ve gotten all the phonology? When we’ve done the phonological analysis and our non-directed elicitation isn’t producing any new phonology.

Let’s consider two brief examples to make this argument more specific, the first from phonology and the second from syntax.

In the realm of phonology, Hyman (2007), shows just how outrageously unnatural are the N+N+N combinations one needs to elicit in order to work through all the possible tone combinations needed to plumb the depths of tone sandhi in Kuki-Chin languages. To check out all the combinations of floating tones that are needed to test particular hypotheses, it is necessary to construct sequences like ‘chief’s beetle’s kidney basket’ or ‘monkey’s enemy’s snake’s ear’. These do not simply, as Himmelmann implies (23), result in the gathering of negative evidence: they generate real physical records.

In the realm of syntax, Keren Rice (2006) has shown the depth of interesting discoveries she made in her investigation of Slave through pursuing a dense nexus of elicited questions targeting the interaction of question words with embedded clauses, of the type ‘where did Mary say John will wait for us?’ By eliciting and permuting structures of this type she uncovered some key principles of how Slave syntax works. Yet once again questions of this level of complexity are so rare in natural language that the generalizations she discovered would be almost impossible to find in a naturalistic corpus. What’s more, a key part of her findings was that some of these sentences are ambiguous. With some word orders, the Slave translation can mean ‘Mary said “where will John wait for us?”’ in addition to the meaning associated with the English sentence. That sort of discovery only comes up with careful probing of meaning. And this will not necessarily be present in a corpus concerned merely with sampling speech.

It is unlikely that even an infinitely large corpus would contain all the combinations needed to work out the answers to questions like this. If a linguist merely records material, without shaping what is to be gathered through their own evolving analysis, future linguists will be deprived of key data. On top of that, to be really useful a corpus must contain discussions of the various ways that each sentence in it can be interpreted in different contexts—a sort of semantically annotated meta-corpus. Again, this can only be produced by embroidering unstructured text with elicited probings—what if you had said X instead? what would it have meant? and so forth. And it emphasizes the fact that there is more to structured elicitation than just asking for grammaticality judgments.

To be fair to his representation of the argument, Himmelmann goes on to say that “it bears emphasizing that documentation does not exclude analysis. Quite the opposite: analysis is essential” (23). This is unproblematic, and tallies with the Hyman and Rice examples above. What I find questionable, though, is Himmelmann’s view that analysis stops well before the completion of a descriptive grammar. My own experience, as both a language documenter and a descriptive grammarian, is that every stage of writing a grammar throws up new questions. The illusion that you understand what is going on, and that your analysis works, retreats like a mirage at every stage of writing the grammar and even while checking final proofs you notice new statements or claims that need checking or

---

4 See Evans and Sasse 2007 for a more detailed statement of this position.
refining. The most important challenge faced by one writing a reference grammar is to construct a description whose thousands of sub-analyses capture the overall Bauplan of the language and at the same time succeed in being mutually consistent. To see this as mere formulation and organization is to grossly underestimate the nature of the analytic challenge. Just having a series of analytic sections that hang off the documentary material runs the triple risk of incompletely pursuing the specific analytic questions, failing to pick up on the interactions between different subanalyses, and representing the analytic claims as a miscellaneous catalogue rather than an organized whole.

I would also add, this time at the level of producing texts—something much closer to the heart of the corpus-building ambitions of the documentarist enterprise—that here, too, the degree of native-speaker involvement and critical engagement increases dramatically at the point where a published product is prepared (say a bilingual edition of a story for school uses, or a first dictionary of a language). Something about the definitive appearance of these products brings out a higher level of scrutiny and a leap to new levels of accuracy in transcription and translation. Both times that I have been involved in producing dictionaries of Australian Aboriginal languages,5 there was a sudden upsurge in interest and in the supplying of new or extended lexical entries at the point where speakers of the language held in their hands a properly-produced book in their language.

For these reasons I think it is a mistake for documentarist linguists to argue that they should consecrate all their time and effort to pure documentary activities at the expense of preparing descriptive grammars or other reference materials. A much more apt strategy is Colette Grinevald’s vision (Craig 2001) of an eternal spiralling upwards through the elements of the classic Boasian trilogy—grammar, texts (now = documentary corpus), and dictionary—with each step forward producing advances and refinements in how the other steps proceed.

To close this review, I would like to draw attention to two facets of the book that reach out in a very positive way to those engaged in language documentation in countries that lack the generous educational or research infrastructure found in Europe, North America, and Australasia. On a world scale, the countries containing the most linguistic diversity, and hence posing the greatest challenges for documentary linguistics in the next decades, are largely those with severe limits on this type of infrastructure. Whether diversity is measured by sheer number of languages, or by deep-time lineages, the result is the same: only three of the twenty-five most linguistically diverse countries in the world are OECD (Organization for Economic Co-operation and Development) countries (Australia, USA, and Canada), and swingeing language death in the latter three countries is likely to move them out of the top 25 in the next few decades anyway.6

A central task facing the community of linguists is thus to recruit, train, and support scholars from developing countries in documenting this vast mosaic. One goal that the book under review achieves outstandingly is to distill the skills needed to undertake language documentation into a form that readers anywhere in the world can take on board,

5 Namely of Kayardild (Evans 1992) and of Dalabon (Evans, Merlan, and Tukumba 2004).

6 See Harmon 1996 for a ranking by numbers of languages, and Evans in press for a ranking by lineages.
with this book in their hands. But a second amplifier is worth mentioning here. Thanks to a welcome initiative by the Volkswagenstiftung, who have made a publication subsidy to Mouton de Gruyter, it is possible for interested researchers and students from developing countries to obtain gratis copies of this book—contact Nikolaus Himmelmann for details.

REFERENCES


EVANS, NICHOLAS. In press. Dying words: Endangered languages and what they have to tell us. Malden, MA: Blackwell/Wileys.


LANGUAGE DOCUMENTATION & CONSERVATION Vol. 2, No. 2 December 2008


Nicholas Evans
nicholas.evans@anu.edu.au