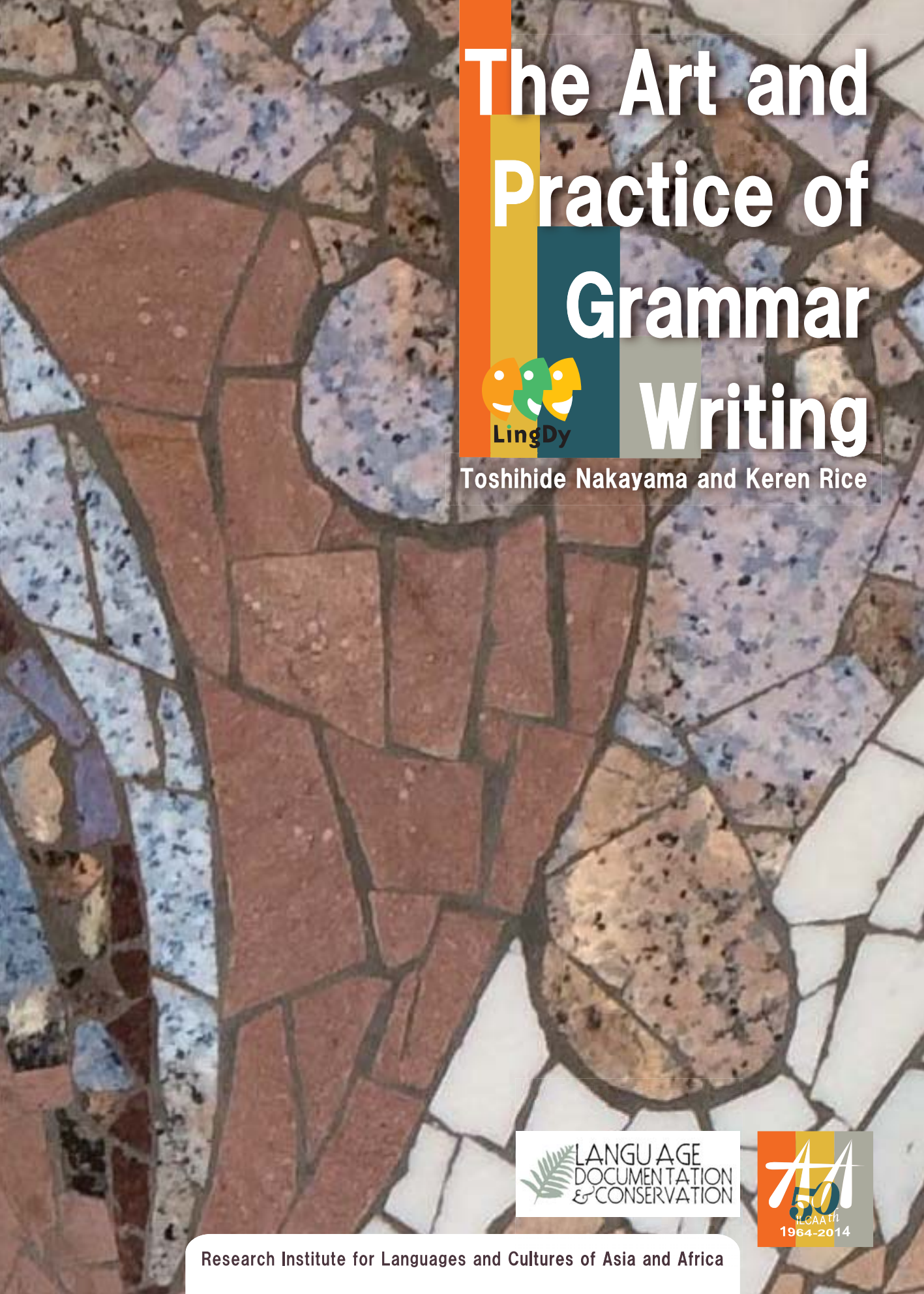


The Art and Practice of Grammar Writing

LingDy

Toshihide Nakayama and Keren Rice

LANGUAGE DOCUMENTATION & CONSERVATION

50 ILCAA th 1964-2014

Research Institute for Languages and Cultures of Asia and Africa

The Art and Practice of Grammar Writing

edited by

Toshihide Nakayama and Keren Rice

Language Documentation & Conservation Special Publication No. 8

Published as a Special Publication of Language Documentation & Conservation

Language Documentation & Conservation
Department of Linguistics, UHM
Moore Hall 569
1890 East-West Road
Honolulu, Hawai'i 96822
USA

http://nflrc.hawaii.edu/ldc

University of Hawai'i Press
2840 Kolowalu Street
Honolulu, Hawai'i
96822-1888
USA

Cover design by Kyoko Nakamura

Library of Congress Cataloging in Publication data
ISBN 978-0-9856211-4-8

http://hdl.handle.net/10125/4579

Contents

1

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 1-5
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4581

Introduction

Toshihide Nakayama[a] and Keren Rice[b]
[a]*ILCAA, Tokyo University of Foreign Studies*
[b]*University of Toronto*

This book grows out of the International Symposium on Grammar Writing: Theoretical, Methodological, and Practical Issues, which was held in Tokyo in December 2009. The symposium and this book project were supported by the Linguistic Dynamics Science Project (LingDy) at the Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, as part of their activity to support and invigorate the descriptive and documentation research of under-studied languages.

The goals of the symposium, echoed in the goals of this book, were two-fold. First we aim to encourage and assist those interested in writing grammars in their endeavor by providing them with information and discussion about methodological and practical aspects of grammar writing. Second we hope to raise awareness about some of the complex issues involved in writing grammars and highlight possible ways of dealing with those issues. In this introduction, we situate the reader by identifying some of the challenges that face the grammar writer in the contemporary context.

1. THE CHALLENGE OF GRAMMAR WRITING. As anyone who has had a goal of writing a grammar can testify, grammar writing is a challenging endeavor. One might think that after undertaking fieldwork, writing up a grammatical description would be a straightforward task, but this is, for most, not the case. There are several factors that contribute to grammar writing being a challenge, and we introduce some of these in this section.

We would like to begin our discussion by addressing why anyone might ever want to write a grammar. To put it simply, many find grammar writing to be extraordinarily rewarding. Valentine (2001), in the introduction to the grammar of Nishnaabemwin, puts this well. He finds writing a grammar to be 'a profoundly humbling experience', adding that it 'provides the enjoyment of thousands of hours of careful scrutiny' (xxxi). (Valentine also says, citing Michael Krauss, that 'even with a hundred linguistics working for a hundred years' one could 'never get to the bottom of a single language'). Valentine says of Nishnaabemwin that 'it is an inexhaustible source of pleasure and challenge for its students' (xxxi). Thus, in addition to challenge, there is enjoyment, pleasure, reward, a sense of accomplishment. These then are some of the reasons that linguists are drawn to writing grammars. What, then, are the challenges?

Here we focus on three major challenges that the grammar writer is likely to face. One arises from the kind of training that students often get and the mismatch between expectations about what a language is and the reality of a language. The second is a consequence of the scope and the magnitude of the project of grammar writing. The third relates to the particular context in which we find ourselves today, with many languages not being passed on to the next generation.

ISBN: 978-0-9856211-4-8

2. MISMATCH BETWEEN FACTS ABOUT LANGUAGE AND THE FRAMEWORK OF GRAMMAR WRITING. Part of the challenge of grammar writing comes from the mismatch that often exists between the reality of language and the goals of a grammar, coupled with our expectations about grammar that have been shaped through what is typically taught in linguistics courses about grammar.

In a grammatical description of a language, we aim to capture the patterns and structures that organize a linguistic system through an analysis of primary linguistic data. Based on reading grammars and on coursework in linguistics, students generally approach linguistic fieldwork with a number of expectations, outlined below:

- Grammar (at least the most important part) consists of general rules: the goal of grammar writing is to identify the general rules and reveal how they form a system;
- Grammatical judgments are invariable: speakers can clearly and consistently tell what is grammatical and what is not, and do not vary about grammaticality judgments;
- Grammar is uniform and invariable: there is a single grammar for a community and therefore everyone in the community share exactly the same structural knowledge;
- The descriptive framework is universal (comparable and applicable across languages).

These common expectations can be contrasted with the reality that students often face in their fieldwork.

- Variability is everywhere: variability in language use sometimes make things seem random rather than rule-governed;
- Fluctuation and indeterminacy in grammaticality judgments: an individual speaker may have difficulty in determining grammaticality; s/he may vary in judgments at different times or for different, very similar, items; judgments quite often have exceptions and fuzzy boundaries;
- A 'grammar' can vary: the 'grammar' of a community is not homogeneous, but varies both within and between individuals;
- Crosslinguistic comparability has limits: even within a universal descriptive framework, there is much room for cross-linguistic variation.

The above set of common expectations is, of course, an idealization: those who have written a grammar of a language are fully aware of this. Students are often at least intellectually aware of what they might find in actual language use; for instance, those who have taken courses in sociolinguistics know about the variation that exists both within a single speaker and within a community, and those who have taken psycholinguistics courses learned that there are not always clear-cut judgments about grammaticality. Nevertheless it can come as a surprise when one begins work on a real language rather than on the constructed problems that we often use in linguistics classes just how much of an idealization is provided in courses. Moving from the real language to what gets written in a grammar requires conscious effort to find the common ground between the patterns and structures that form the grammar and the real language as it is spoken.

The mismatch between what a language is all about and what a grammatical description of a language can realistically capture provides one challenge to the grammar writer.

3. THE SCALE OF GRAMMAR WRITING AS A PROJECT. Grammar writing is challenging for other reasons as well. The project of writing a grammar is substantially larger in scale than many other research projects. The grammar writer is called upon to have comprehensive knowledge of a language, from its phonetics to its discourse structure, coupled with socio-cultural knowledge. While pieces of a grammar are independent, components of the grammar intersect and interact with one another, requiring careful strategy and planning in how to go about writing. The grammar writer must thus put careful thought into how they will complete a project that has no logical end to it, how they will organize the grammar, and how they will relate the pieces of different parts of the grammar to one another.

The scale is large, and there is no particular template for writing a grammar. Much depends on the language itself and on who the author anticipates will use the grammar. There are many complex decisions to make, making grammar writing an art. There is also much to balance: What topics must be covered and which are optional? What language is represented? What kind of depth is needed for each topic? How do the pieces fit together to create a whole?

Even from this brief discussion, it is perhaps possible to see both the pleasure and the challenge that Valentine talks about.

4. CONTEXT IN WHICH DESCRIPTIVE GRAMMAR IS SITUATED. Grammar writing has probably always been a complex undertaking, but writing grammars today strikes us as perhaps even more challenging than it was in the past for several reasons. The most important change that has taken place is the recognition of the endangered state of so many languages of the world. This fact puts grammar-writing projects in a context different from the past and subjects the project to different sets of expectations and requirements. Many questions must be considered in writing a grammar that were not so important in the past. Who is the grammar written for? What kind of language should it represent? Can a single grammar serve all audiences?

Coupled with language endangerment are the rapid changes in technology in the past twenty or so years. One can gather amounts of data that were unprecedented until just recently. The potential scope of a grammar, as discussed above, has always been vast, but the advancement of the technology pushes the horizon even further: it allows more questions to be asked and also allows for new types of analysis and new forms in which grammars can be presented.

Thus, grammar writing in the present day has become more important, and at the same time, more complex and complicated. This is a very good time for us to reexamine the process and method of grammar writing.

5. WHY DO WE NEED YET ANOTHER BOOK ON GRAMMAR WRITING? There have been a few books on grammar writing published in recent years (Ameka, Dench, and Evans 2006; Payne and Weber 2007). One might ask why another book is necessary, given the recent attention to the writing of grammars. This book comes at the question of grammar writing in a somewhat different way than the other books. In particular, the focus of this

book in on practical issues in the actual process of grammar writing, rather than on the abstract values or theoretical frameworks of grammar writing. This was a direct reflection of our concern about the sustainability of the tradition of grammar writing.

While the chapters in this book may well appeal to anyone who decides to write a grammar, especially of an understudied language, we hope that the book will be of particular value to junior grammar writers, and above all to students who are trying to sort out the enormous task that they will take on if they decide to study a language and write a grammar of it for their dissertation.

6. ORGANIZATION OF THIS BOOK. The book contains eight chapters. The chapters by Carol Genetti, Andrew Pawley, and Thomas Payne represent excellent starting points on reading this volume. Pawley writes as someone who has supervised many dissertations that involved the writing of grammars. He gives excellent advice to the person undertaking such a project about how to plan the project, what a realistic thesis is, and so on. While he aims the chapter at Ph.D. students, it contains advice that will be useful to anyone who wants to write a grammar. Payne addresses a number of tensions that arise in writing a grammar: comprehensiveness vs. usefulness, technical accuracy vs. understandability, universality vs. specificity, a 'form-driven' vs. 'function-driven' approach. Most find that it is not possible to be both comprehensive and useful, for example, and Payne gives excellent advice on how to balance these conflicting demands. He ends with an outline of what a balanced grammatical description might look like. Genetti looks at how one might incorporate linguistic typology, argumentation, and theoretical innovation into a grammar, while at the same time maintaining a balanced grammar.

The other chapters address more specific topics. Ulrike Mosel considers how corpus linguistics can serve grammar writing, and how a corpus might be used in preparing a grammar. Marianne Mithun is concerned with data and examples, examining data from phonetic, phonological, morphological, syntactic, semantic, discourse, prosodic, diachronic, and language change spheres and how it can meet criteria of comprehensiveness, accuracy, and sensitivity. Keren Rice discusses the role of phonetics and phonology in a grammar. She provides a survey of how grammars have changed over the years in terms of the content about sounds, and makes recommendations about what aspects of sound structure should be included in a grammar. Terrell points out the importance of documenting particular domains of an endangered language on basis of his work on the Akha shaman chants. Kenneth Rehg discusses the larger role of grammars, and their importance in developing community grammars. Rehg stresses the important role that linguists have to contribute not only to the scientific enterprise, but also to language conservation efforts.

The chapters in this volume are all written by people who have made the kind of contributions that Rehg discusses, and are thus all informed by the knowledge that many languages are not being transmitted to new generations. We hope that the chapters are stimulating to the reader, especially to the person who is new to grammar writing, but we also hope that those who are experienced in grammar writing might come away with something new.

Just like a grammar writing project, the symposium and this book project would not have been made possible without various support from so many people. Although it would not be possible to list them here by their names, we would like to take this opportunity to

thank every one of them for their help. We would also like to thank Shannon Mooney for her assistance with preparation of manuscripts for publication.

References

Ameka, Felix K, Alan Charles Dench, and Nicholas Evans. 2006. *Catching Language: The Standing Challenge of Grammar Writing.* Berlin: Mouton de Gruyter.

Payne, Thomas, and David Weber. 2007. *Perspectives on Grammar Writing.* Amsterdam: John Benjamins.

Valentine, Randolph. 2001. *Nishnaabemwin Reference Grammar.* Toronto: University of Toronto Press.

Toshihide Nakayama
nakayama@aa.tufs.ac.jp

Keren Rice
rice@chass.utoronto.ca

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 7-23
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4582

2

Grammar writing from a dissertation advisor's perspective

Andrew Pawley
Australian National University

Anyone who intends to produce a grammar of a previously little-described language needs to (1) plan the scope, methods and timetable of the data gathering process, (2) think about the conceptual framework that will shape data-gathering and analysis, (3) gather and organize the data, (4) analyse the data, and (5) plan the structure of the written account and (6) write the grammar. The steps are not simply sequential but are to some extent cyclical. This chapter will look at an advisor's role in guiding a PhD student through these steps. It will focus on the following questions: What kinds of data, and how much, are sufficient to base a grammar on? What is a realistic size for a PhD dissertation grammar? What are the main alternative ways of organizing a grammatical description, e.g. in terms of topic divisions and sequencing? What are the *dos* and *don'ts* to be followed in order to make the grammar as descriptively adequate and user friendly as possible? What are the main reasons why some students take forever to complete the analysis and writing process?

1. AIMS. This chapter looks at grammar writing from the perspective of a dissertation advisor. It identifies key decisions that must be made in the course of planning and producing a grammatical description for a doctoral dissertation, and comments on the role of an advisor in dealing with these matters.[1]

It is more than 40 years since I was a PhD student myself, doing fieldwork on Kalam, a Papuan language of New Guinea, and writing a dissertation on the phonology and grammar of that language under the supervision of Bruce Biggs at the University of Auckland. Following those *rites de passage* I got a job as a linguist, and since then have in my turn supervised more than 50 graduate theses, about half of them reference grammars. But I am very mindful of the fact that there are others who have written quite extensively on aspects of the craft of grammar-writing, whereas I have not done so until now.[2] The remarks that follow will draw on the observations of many other scholars as well as on my own experience.

I will consider the following general questions to do with writing a grammar for a dissertation project.

- What is the nature of the task? What is entailed in writing a grammar?
- How should one prepare for such a project?
- What are the main decisions to be made on the way?
- What are the *dos* and *don'ts* in presenting the analysis?
- How not to write a grammar. Some bad strategies in data collecting and writing up.
- What is the role of the advisor in these matters?

1 I am indebted to the two anonymous referees for helpful comments on a draft.

2 For example, Ameka et al. (2006), Bowern (2008), Mosel (2011), Payne (1997) and Payne and Weber (2005), to name just a few.

ISBN: 978-0-9856211-4-8

2. WRITING A GRAMMAR: WHAT IS THE NATURE OF THE TASK AND HOW SHOULD ONE PREPARE FOR IT?

2.1 THE MAIN STEPS. The expression 'to write a grammar' has a number of different readings. A literal reading is roughly 'to give a written account of the grammar of a language' but as a description of what is entailed in producing a grammar this is misleading. The writing bit is just one of several parts of the job. Anyone who would produce a grammar must go through at least the following stages: (1) plan the scope, methods and timetable of the data-gathering process, (2) think about the conceptual framework that will shape data-gathering and analysis, (3) gather the data, (4) organize and analyse the data, (5) plan the structure of the written account and (6) write the grammar.

However, we should bear in mind that the steps are not just sequential but also cyclical to some extent. One begins analysis on day 1 of the data-gathering phase and preliminary analyses will influence or direct the next stage of data gathering. The process of writing up will reveal gaps in the data or flaws of analysis, leading to further data-gathering, reanalysis, and a revised presentation, and so on.

2.2 PLANNING THE PROJECT. Many things have to be planned at the outset, beginning with the choice of a language. One needs to plan the scope and timetable of the research, such as the number and length of field trips and the methods of data collection, and one must give some thought to the model of grammar that will provide the framework for the analysis and write-up.

2.2.1 WHAT KIND OF GRAMMAR? The primary readership of a PhD grammar will be professional linguists. The utilitarian purpose of a PhD dissertation is to demonstrate that the author has achieved a professional level of competence in the field of study and has made a substantial original contribution to it.[3] Two things follow from these circumstances. First, the grammar should be an analytic grammar, a reference grammar, rather than a pedagogical one. Second, it will usually be of a relatively little-described language, at least one for which there exists no good reference grammar.

What about languages for which quite good reference grammar or grammars already exist, e.g. Indonesian, Maori, Mohawk or Quechua? Are there circumstances in which it is acceptable to do a grammar of such a language for a dissertation? Some advisors would say, yes, if there are data that cry out for a fresh analysis, or if the grammarian adopts a particular theoretical framework that one can be confident will reveal new insights. The risks in such cases are that the outcome will end up not being very novel. For reasons of national or academic politics sometimes students choose to describe a prestigious language that is already well documented rather than taking on a minority language that is undescribed but considered less important. I was told that this happened in the 1950s when several

3 While all linguists recognise that writing a descriptive grammar is a very challenging exercise, some would raise practical and/or ideological objections to writing a grammar as a PhD assignment. Instead, they would encourage students to do a dissertation that tackles a particular theoretical problem, or at least to write a theory-focused grammar, one that tests a particular theoretical framework. This, it is argued, will give them a better chance of getting a job in a field where theoretical work is more highly valued than descriptive work.

Indonesian students studying at US universities each chose to do descriptions of Indonesian for their dissertation. In Indonesia at that time there was no career advantage to be gained from choosing to work on a minor language.

A compromise is to do a 'topics in the grammar of X' dissertation, rather than a general reference grammar, where the topics are just those domains where the analyst has an original contribution to make.

2.2.2 CHOOSING A DESCRIPTIVE FRAMEWORK. In principle, a firm decision as to what kind of descriptive framework to use need not be made in the initial planning stages. However, there are good practical reasons why student and advisor should consider this matter early on and come to a joint understanding. For one thing, a student's choice of advisors may depend on decisions made. You don't want your chief advisor to quit in the middle of your writing-up phase because he or she objects to the theoretical framework you are using. Moreover, the kinds of data to be collected will, to some extent, to be dependent on the questions that the framework focuses on.

If a reference grammar is to be readable generations from now it should use a descriptive framework (i.e. a body of analytic concepts and terms) that is familiar to most or all grammarians. This is easier said than done because all descriptions are to some extent theory-specific and specific theories of grammar are notorious for having a short shelf life – and even in their heyday are accessible only to specialists. After about 1960, the research agendas of descriptive and theoretical linguistics diverged sharply. The ultimate aim of theoreticians is to arrive at generalisations about human language and to explain why languages are the way they are and why languages change in the ways they do. Many theoretical linguists have little interest in describing particular languages, regarding them merely as a source of evidence for choosing between competing theories or competing claims about linguistic universals. Descriptive linguists, by contrast, generally maintain an interest in developments in theory as being relevant to the systematic analysis of a language.

Accordingly, most grammar writers of my acquaintance take the view that where possible one should try as far as possible to avoid analytic concepts that are local to particular models of grammar and use instead what R.M.W. Dixon and Matthew Dryer call 'basic linguistic theory' (BLT) (Dixon 2009, 2011, 2012, Dryer 2006). By this is meant, roughly, the analytic concepts that have widest currency among descriptive linguists and typologists. There is a very sizeable body of such concepts. Many are grounded in traditional grammar.

However, BLT is not a fixed thing. The tools of grammatical and phonological analysis continue to evolve and one should be ready to take on board new and useful concepts. Given that such change is inevitable and healthy, it is probably unrealistic to expect that any grammar written today will be an easy read in two centuries time. But by carefully defining key analytic concepts, a grammarian can at least make the path easier for readers.

In any case, the writer of a grammar should be eager to ask questions that have been generated by theory-specific work, if these offer promise of throwing light on the grammar of the target language. Keren Rice (2006:403) remarks that "... the grammar should be informed by theory. This will help make it coherent, and it will allow questions to be asked

that might not come up otherwise. But...theory is not the goal of a grammar". If you want to discuss theoretical issues in depth you write a separate paper or book.

But for some students, who have a keen interest in or allegiance to a particular theoretical model, it is hard to resist the temptation to use that model. And I know at least one linguistics professor who would only agree to chair a student's committee if the student writes the grammar using a particular preferred model of generative grammar with elaborate formalisms. The upside of this was that students gained a rigorous training in that model and, often, achieved some insightful analyses. The downside was that the resulting grammars were overloaded with theory-specific formalisms that are off-putting to all but the most dedicated reader.

2.2.3 ON FIELDWORK. If the data are to come mainly or wholly from fieldwork among speakers of a little-described language in a remote location, how long should a student spend in the field? The field worker should aim to obtain sufficient data to write a first grammar and to gain a pretty good level of competence in the language. Gaining such competence is advantageous for a variety of reasons. The more fluent you are in the language the easier it is to talk to a wide range of people in the community, to argue the point, to follow conversations and monologues, to ask complicated questions and to elicit suitable illustrative examples, and, importantly, the better placed you are to critically evaluate the information that comes in and so to reduce errors.

It is impossible to give a hard and fast recommendation about the time needed to achieve these objectives, because this will vary according to (a) the difficulty of the language, (b) whether or not there are good descriptions available for closely related languages, (c) how difficult fieldwork conditions are, and (d) personal factors. However, as a general rule, I think one needs at least nine or ten months, and preferably a year in the field.

How should the fieldwork be apportioned? How many field trips is an optimal number? How should they be spaced? There are various reasons why it is better to have at least two shorter field trips than a single long one. The first spell of fieldwork should be the longer one, of at least six months, aimed at yielding enough material to draft a substantial part of the dissertation. One can then return home, refresh the body and mind, write up first drafts of a large section of the dissertation, see where all the gaps in the data are, and after six months to a year, return to the field to check and fill in gaps. Nicholas Evans (pers. comm.) argues in favor of three trips: the first and last quite short, say three months each, with a longer middle trip. The reason for preferring a relatively short first trip is that during the early stages of research on an unfamiliar language the researcher needs to spend a lot of time on analysis, simply making sense of the data, and such analysis can be done at home.

Circumstances may sometimes limit the student to a single, long field trip. In these cases it is advantageous to press ahead with analysis and sketch the core chapters while still in the field, so that problems of analysis and major gaps in the data will become evident while there is still time to remedy them.

The situation is different if the field site is easily accessible. And obviously, it is different for those students who have had the advantage of having spent years studying the language before they began their PhD. Students who are members of the Summer Institute of Linguistics (SIL) often have this advantage.

3. SOME NOTES ON DATA-GATHERING

3.1 THE IMPORTANCE OF LEXICAL DATA. Keeping a lexical file is an important part of grammar-writing. The file should include notes on the grammar of each lexical item and examples of use. As Lichtenberk (2008:6) points out, "Grammatical rules, or patterns, are generalizations over various properties of individual lexical items. One cannot write a reasonably detailed grammar...without fairly extensive lexical information". When working on his grammar of Toqabaqita, he found grammar- and dictionary-writing to be mutually beneficial:

> Detailed grammatical analysis enabled a more accurate treatment of the grammatical elements in the dictionary than would otherwise have been the case. And the lexicographical work has been of great importance to the grammatical analysis. In any language, grammatical rules, patterns, regularities are of highly different degrees of generality. Few, if any, hold across the board. Many grammatical patterns are lexically sensitive; they hold for some but not all members of a certain word class.

Of course, it is unreasonable to expect someone who is doing a PhD dissertation on the grammar of a language to try to compile a comprehensive dictionary at the same time. The aim should be to test the grammatical properties of a representative sample of lexical items, so that one can arrive at a reasonably fine-grained treatment of word classes and sub-classes. This brings us to the next question.

3.2 HOW MUCH DATA IS NEEDED TO WRITE A GRAMMAR? How much data is needed to write a reference grammar? Obviously, that depends on how comprehensive the grammar is. One can probably write quite a useful grammar sketch based on a corpus of 5,000 words but it will contain many glaring gaps. 20,000 words will yield a more complete analysis but there will still be many gaps in the kinds and details of constructions represented. 30,000 and 40,000 words will yield still more complete analyses, and so on. But we need to bear in mind that some morphological forms and other grammatical combinations will not occur in a corpus of a million or even 10 million words. Directed eliciting is needed to collect morphological and phrasal paradigms. And that brings us to another question.

3.3 WHAT KINDS OF DATA ARE LEGITIMATE? What kinds of data are legitimate for basing a grammar on? Is elicited data acceptable or should all data come from natural discourse? Is there a desirable balance between the two? And when there is significant variation relating to age, dialect, etc., how one should handle this in gathering samples of the language?

No one would deny the value of a large corpus of natural discourse data. One should collect and transcribe extensive text materials of various genres. But it would be foolish to take an extreme purist position and exclude elicited data. What you can collect in ten months will not be sufficient to answer many questions about grammar and lexicon. You need to elicit paradigms, word meanings and sentence forms and translations, and to train

consultants to give you grammaticality judgments. Of course one should be alert to the dangers of using elicited data and so should check and double check where possible.

The corpus of data obtained during the early stages of fieldwork, when the linguist does not yet know the language well, is likely to contain lots of mistakes and this is especially true of elicited data. I recall the comments of a colleague on the efforts of three linguists who had independently carried out dialect surveys of many Fijian languages and dialects, collecting, among other things, a 300 word basic vocabulary list from each, using Bauan, the lingua franca of Fiji, as the eliciting language. He compared the results from one particular language, collected (as far as I recall) at the village of Namuamua, in the Namosi region of the island of Viti Levu. He said that linguist A visited Namuamua for a single afternoon and his vocabulary list contained 27 errors. B stayed for two days and his list contained 9 errors. (I was linguist B.) C stayed a month and was confident that his list contained no errors. C, who was the commentator, had the additional advantage of having near native fluency in the lingua franca and an impressive familiarity with virtually all of the Fijian languages and dialects.

4. THE SCOPE OF THE GRAMMAR

4.1 HOW LARGE SHOULD A PHD GRAMMAR BE? One of the pitfalls in writing a dissertation is trying to do too much. My supervisor told me firmly: "Remember, the PhD is just a qualification, a ticket to a job. Do a good job but nothing too grand. Leave the grand projects till later". Mark Donohue recalls that when he proposed to add yet another chapter to his already long dissertation draft, he was told gently "You know Mark, you're allowed to do research after the PhD".

OK, but just how large and how comprehensive should a PhD reference grammar be? Let's start with size.

There are some very large reference grammars of previously little-described languages. Exceptional are Keren Rice's grammar of Slave, which runs to over 1400 pages (Rice 1989) and Frank Lichtenberk's grammar of Toqabaqita, 1375 pages (Lichtenberk 2008). These are comparable to the largest grammars of English, e.g. by Quirk et al. (1974, 1100 pages) and Huddleston and Pullum (2002, 1800 pages). My home library contains a few other very other large grammars of Pacific Island languages, e.g. Mosel and Hovdhaugen's (1992) grammar of Samoan is around 800 pages and Alexandra Aikhenvald's (2008) Manambu grammar is around 700. However, none of these works were PhD theses. All were done by established scholars, or teams of scholars, over many years. In the case of the English grammars, the authors could build on centuries of previous work.

What is a reasonable length to aim at for a PhD grammar? I compared 18 grammars published between 1994 and 2008, each grammar being a revised version of a PhD dissertation submitted either to an Australian (13) or a Dutch or German university (5).[4] The languages described are Austronesian (10), Papuan (6) and Australian (2). The figures without parentheses represent the number of pages of the grammar proper. The figures in parentheses indicate the total pages including appendices (usually texts) and references. The year of publication is given; often this is a few years after the dissertation was completed.

4 I do not have on hand details for a comparable sample of grammars submitted as dissertations to North American universities during this time frame.

TABLE 1 Details of 18 PhD dissertation grammars published between 1994 and 2008

AUTHOR	TITLE	NO. PAGES
Australian National University (10)		
Bowden	*Taba: Description of a South Halmahera language* (2001)	408 (451)
Bugenhagen	*A grammar of Mangap-Mbula, an Austronesian language of Papua New Guinea* (1995)	355 (418)
Ezard	*An Austronesian language of the Milne Bay area, Papua New Guinea* (1997)	297 (320)
Farr	*The interface between syntax and discourse in Korafe, a Papuan language of Papua New Guinea*	417 (459)
Hyslop	*The Lolovoli dialect of the North-East Ambae language* (2001)	438 (476)
Jones	*Towards a lexicogrammar of Mekeo (an Austronesian language of western central Papua)* (1998)	553 (601)
Obata	*A grammar of Bilua, a Papuan language of the Solomon Islands* (2003)	281 (329)
Quick	*A grammar of the Pendau language of central Sulawesi, Indonesia* (2007)	601 (716)
Teng	*A reference grammar of Puyuma, an Austronesian language of Taiwan* (2008)	279 (309)
van Klinken	*A grammar of the Fehan dialect of Tetun* (1999)	322 (355)
Other Australian (3)		
Eades	*A grammar of Gayo, a language of Aceh, Sumatra* (2005)	316 (350)
Pensalifini	*A grammar of Jingulu, an aboriginal language of the Northern Territory* (2003)	240 (262)
Sharp	*Nyangumarta, a language of the Pilbara region of Western Australia* (2004)	392 (429)
Netherlands (4) and Germany (1)		
Dol	*A grammar of Maybrat, a language of the Bird's Head Peninsula, Papua Province, Indonesia* (2007)	290 (328)
Klamer	*Kambera, a language of Eastern Indonesia* (1994)	336 (368)
van den Heuvel	*Biak: description of an Austronesian language of Papua* (2006)	423 (473)
van Staden	*Tidore, a linguistic description of a language of the North Moluccas* (2000)	355 (566)
Wegener	*A grammar of Savosavo, a Papuan language of the Solomon Islands* (2008)	330 (372)

It can be seen that 16 of the grammars (excluding appendices and texts) fall between 240 and 440 pages, with just a couple of outliers in the 600 range. The median length is 336 pages. Appendices with texts usually amount to 30-50 pages but in one case (the Tidore grammar of van Staden (2000) they run to 200 pages. Over the years I have generally advised students that 300-350 pages, excluding appendices, should be ample. However, my impression is that in recent decades PhD grammars have grown larger. If so, no doubt one

factor driving it is competition: survival of the fittest in the job market, Darwinian natural selection.[5] Another factor (see discussion below) is progress in the field: we know more than we used to about some aspects of grammar and pragmatics.

4.2 WHAT SHOULD BE IN A REFERENCE GRAMMAR? HOW SHOULD IT BE ORGANIZED? It is generally agreed that a grammar should include descriptions of the phonology, morphology, syntax and semantics of a language. However, within these broad limits grammars vary considerably in scope and degree of detail. My impression is that reference grammars today, including those done as PhD dissertations, cover a wider range of phenomena than those done, say, in the 1950s, 60s and 70s, particularly in the domains of syntax and pragmatics. It is also my impression that the modern grammars tend to be more discursive and readable than those produced at the height of the structural and transformational-generative grammar eras. In those times grammarians were more concerned with form than function. Descriptions usually said less than nowadays about the semantic characteristics of word classes and little about such things as pragmatics and information structure. Treatments of syntax were more limited than today. Grammatical relations like subject and direct object were not so carefully defined.

However, it is also my impression that modern descriptions tend to be less rigorous than grammars written by structuralists of 40 and 50 years ago in one respect: their treatment of the combinatorial possibilities of constituents. Even before Chomsky, structuralist grammars sought to specify all and only the possible combinations of elements within whatever units were being described. Modern grammars tend to contain fewer generative formulae but offer a more expansive and more readable account of grammatical constructions. I think there is room in a grammar for both approaches. One can always insert formulae predicting the possible combinations at the end of a more expansive discussion.

When it comes to the finer details of a description—which grammatical phenomena to treat and in how much detail and in what order, most us would agree with those, like Rice (2006:400-1), who advises against following a predetermined outline because each language demands its own strategy. She illustrates by referring to several grammars that differ markedly in the way they are organized.

However, there are some favoured patterns. I compared the contents of 17 different grammars of Austronesian and Papuan languages which were originally submitted as PhD theses. They fell into two classes. Most consist of 7 to 10 chapters, some run to 14-16 chapters. I also looked at some larger grammars that are not PhD theses, which have between 20 and 40 chapters.

In the case of grammars with 7-10 chapters, most of the chapter titles in the different grammars are essentially the same and the order of chapters is very similar. A typical sequence is:

5 I am aware of one PhD grammar that is right off the scale: Alexandre François's description of Mwotlap, a language of north Vanuatu. This totals 1033 pages or about 470,000 words, excluding appendices. It was completed in four years including a year's fieldwork. During the same period the author managed to do the research for, write up and publish a 350 page grammar of another language of Vanuatu. But mere mortals should not attempt such feats.

1. Introduction. (Also called The language and its speakers, or similar.)
2. Phonetics and phonology.
3. Word classes. (Also called Parts of speech.) Sometimes includes phrase types.
6. Derivational morphology.
5. The noun phrase. (Also called Referring expressions.).
6. The verbal complex.
7. Basic clause structure.
8. Complex sentences.

The content and organization of the PhD grammars with 14 to 16 chapters differs from these in two ways: (i) They give separate chapter status to items such as pronouns, numerals and adpositional phrases, which in other grammars are treated within one of the standard chapters. (ii) They devote whole chapters to construction types that are particularly prominent or elaborate in the language, e.g. serial verb constructions, possessive constructions. Three examples follow:

Lolovoli, NE Ambae (Hyslop 2001)
1. Introduction. 2. Phonology. 3. Basic clause structure. 4. Word classes. 5. Noun phrase. 6. Adjuncts. 7. Possessive and associative constructions. 8. Spatial reference. 9. Verb phrase. 10. Serial verb constructions. 11.Valency change and rearrangement. 12. Reduplication. 13. Existential, equational and locational clauses. 14. Subordinate clauses. 15. Coordination.

Taba, S. Halmahera (Bowden 2001)
1. Introduction. 2. Phonetics and phonology. 3. Morphological and syntactic units. 4. Parts of speech. 5. Basic clause types. 6. An overview of clausal syntax. 7. Nouns and noun phrases. 8. Verb morphology and valence. 9. Possession and related constructions. 10. Quantifiers. 11. Demonstratives and directionals. 12. Serial verb constructions. 13. Adpositional phrases. 14. Clausal modifiers. 15. Questions and requests. 16. Interclausal relations.

Fehun Tetun, East Timor (van Klinken 1999)
1. Introduction. 2. Phonology and morphophonemics. 3. Word classes. 4. Derivational morphology. 5. Numerals, classifiers and the numeral phrase. 6. Pronouns and determiners. 7. The noun phrase. 8. Prepositions and the prepositional phrase. 9. The clause. 10. Auxiliaries. 11. Adverbs and verbal modifiers. 12. Serial verb constructions and prepositional verbs. 13. Complementation. 14. Beyond the clause.

Now let us look at the chapter headings in a very large grammar, with 40 chapters, the average length of the chapters being 33 pages.

Toqabaqita, Malaita, Solomon Is. (Lichtenberk 2008)
1. Introduction. 2. Phonology. 3 Grammatical profile. 4. The verb phrase. 5. The particles in the verb complex. 6. The noun phrase. 7. The noun-phrase internal particle group. 8. Possessive and associative noun phrases. 9. Nominalizations and deverbal nouns. 10. The prepositional phrase. 11. Coordination of noun phrases and prepositional phrases. 12. Compounding. 13. The demonstratives and the demonstrative adverbs. 14. Constructions with inclusory pronominals. 15. Tense and aspect. 16. The sequential subject markers. 17. Negation. 18. Mood. 19. Interrogatives. 20. Imperatives. 21. Low-individuation of participants derivations. 22. Reciprocal and related situation types. 23. Self-contained situations. 24. Unrestricted choice. 25. Comparisons of inequality. 26. Locational, existential, and possessive sentences. 27. Verbless sentences. 28. Coordination of clauses. 29. Complement clauses. 30. Relative clauses. 31. Conditional sentences. 32. Concessive clauses. 33. Reason clauses, purpose clauses, and purpose nominalizations. 34. Consequence clauses. 35. Temporal relations. 36. Deranked subordinate clauses. 37. Direct speech. 38. Topicalization. 39. Focusing. 40. Toqabaqita, Solomon Islands Pijin and English. Appendix: Texts. Notes. References. Index.

At first glance it may appear that the Toqabaqita grammar treats five times as many different topics as the PhD grammars containing just eight chapters. Certainly, the range of topics is considerably greater. However, the main difference is not so much the range of topics as the depth of coverage, the amount of detail to be found in the large grammar. Topics that occupy one section of shorter grammars, or that perhaps receive only passing mention, get a whole chapter to themselves in the Toqabaqita grammar.

With respect to the ordering of chapters Rice (2006) quotes a general principle from the Cambridge grammar series:

> Basically, if an analytic decision concerning category X needs to refer some facts concerning category Y, then the chapter dealing with Y should be positioned before that dealing with X.
>
> However, things are often not quite that simple. Whichever order is chosen, there is bound to be a need for extensive cross-referencing.

4.3 SHOULD A GRAMMAR BE CONCERNED WITH HOW TO SAY THINGS IDIOMATICALLY? I have always been concerned by the fact that knowing the core grammatical rules of a language will not allow you to speak that language idiomatically, saying things the way a native speaker says them. Writers of reference grammars traditionally are not much interested in what things people commonly say in a language and how they say these things, beyond the question of what is grammatical. Only a very few reference grammars contain sections on 'How to say things', ways of talking about everyday topics, even when there are very clear rules or conventions for doing so. Try looking in a grammar for an account of the conventions for telling the time of day. An exception in my sample of grammars of Pacific Island languages is Robert Bugenhagen's grammar of Mangap-Mbula, which has an appendix that begins:

> It would be a tragedy to read through all the listings of word classes, rules and trees of a grammar and still not have any idea how to express some of the everyday things which make up a large part of people's normal conversation.
> (Bugenhagen 1995:374)

Under this heading the author goes on to treat 1. Existential and presentative constructions. 2. Location. 3. The weather, time of day, etc. 4. Movement. 5. Ownership and other types of relationships. 6. Emotion. 7. Conveyance. 8. Affectedness, different types of causation. 9. Speech/quotations.10. Want, try, believe, promise, persuade, know, see and think (verbs that take sentential complements).

Bugenhagen wrote his grammar in the early 1990s. In today's grammars you are likely to find some of these topics treated in the main body of the grammar, because they concern constructions where morphosyntax craves correlation with function. Bugenhagen's topics 1, 2, 5, 8, 9 and 10 fall into this category. Others are less likely to be found in a grammar. It would be an interesting exercise to compare a sample of grammars and see how many of them give some prominence to ways of talking about particular subject matters, such as weather, or emotions, or conveyance.

5. THE FINER DETAILS OF PRESENTATION

5.1 SOME QUESTIONS. The previous section took a broad view of the contents of a grammar. Let us now consider some of the finer, nitty-gritty details of presentation. For example: Should the morphosyntactic part of the grammar begin with an overview of the main types of clause constructions before getting into the details of word formation and word classes? Should there be a separate chapter on word classes or should each word-class be introduced when dealing with the type of construction headed by that word-class (e.g. verbal clauses, noun phrases, etc)? Should each chapter begin with a summary of what it is about? What notational devices should be used, e.g. tree diagrams or square brackets to show constituent structure? How many examples should one give to illustrate a particular point? And for each construction type should one try to include a compact generative grammar in the form of explicit, concise formulae that will (aim to) predict all and only the possible strings, or should one be content to discuss constructions in a more informal way, with limited use of generative formulae?

5.2 SOME DOS AND DON'TS. A recent issue of Studies in Language (Payne and Webber 2006) contains papers by a number of linguists reflecting on issues in grammar writing. Michael Noonan's contribution (Noonan 2006) includes a list of dos and don'ts, based on suggestions provided by various experienced grammarians. The dos and don'ts concern, not so much the things that everyone agrees should be done in a grammar, like a thorough description of the morphology and syntax, but some of the things that tend to get left out or done not so well. Noonan divides them into three broad categories: User friendliness, descriptive adequacy and comprehensiveness. A summary of some of these prescriptions follows.[6]

6 I have omitted a few of Noonan's items, including those that we have already discussed, and have renumbered them accordingly.

- **User friendliness**

1. Provide a detailed index and table of contents. This helps readers find information.
2. The text should be divided into numbered and titled sections and subsections, with ample cross-referencing. Important terms should be highlighted by boldface.
3. Provide plenty of examples. Made up examples are appropriate for presenting information about basic constructions [when the writer is certain of their correctness: AP] but naturally occurring examples should otherwise be used.
4. Provide interlinear glosses (morpheme by morpheme) translations, as well as free translations for all examples.
5. A typological sketch of three to five pages should be included at the beginning of the grammar. This gives the reader a quick overview of the most important elements in the grammar.

- **Descriptive adequacy**

6. Use standard IPA characters to present information about the phonetics of the language.
7. Give instrumental documentation of the acoustic properties of vowels, the duration of segments, and tone and pitch accent systems.
8. Provide a full description of segmental and suprasegmental contrasts and the evidence for these.
9. Provide a full description of distributional patterns of elements of the phonology, in terms of syllables, words and whatever other units are relevant.
10. For morphologically complex languages, provide not just lists of affixes but tables with full paradigms showing combinations of all relevant morphemes.
11. Define grammatical categories used in the grammar. Don't assume that word classes and grammatical relations (subject, direct object etc) are givens.
12. The choice of labels for grammatical features is not as important as a thorough presentation of the facts.
13. It's better to admit ignorance of a poorly understood grammatical feature than to say nothing about it. Saying nothing can be misread as indicating that the feature is lacking.
14. The absence of a feature should be noted, if that feature might be expected to occur on areal, genetic or typological grounds.
15. Indications of frequency of grammatical elements and constructions should be provided where appropriate.
16. It is best to describe morphology mainly with a form-to-function orientation and syntax mainly with a function-to-form orientation.
17. A vocabulary of all the lexemes which occur in the grammar should be provided.
18. A collection of texts, at least 20-30 pages, with morpheme glosses and translations, should be included.

- **Comprehensiveness**

19. The writer should consult survey questionnaires and well-regarded grammars to make sure that important topics are not missed.
20. The grammar should contain information about genetic and areal affiliations of the language.

21. The grammar should contain information about how the data was obtained and about the sociolinguistic context. The latter should include the number and location of speakers, the age demographics of language use, the degree and nature of multilingualism, degree of literacy, etc.
22. There should be ample reference to previous scholarship on the language and the culture of the community.

Noonan ends with two items that are desiderata rather than requirements:

23. A good dictionary is a powerful adjunct to a good grammar. It will contain much grammatical information supplementing that in the grammar.
24. Where practical, audio and video recordings should be made of various genres.

Most grammarians would surely agree with most of these points although the question arises whether all of prescriptions 1-22 should apply to PhD grammars, which should not aim to cover everything in depth. The most contentious claim among these might be 16, saying that it is best to describe morphology mainly with a form-to-function orientation and syntax mainly with a function-to-form orientation. This is a useful general rule of thumb but the issues here are complex and one should be wary of hard and fast recommendations.

To Noonan's list I would add:

25. Begin each chapter with a summary of what it is about.
26. When describing the internal structure of complex constituents (e.g. noun phrase, verb, verbal complex, transitive clause) give explicit statements of combinatorial possibilities.
27. For the published version of the dissertation, include an index giving page references for key topics and terms.

6. HOW NOT TO COMPLETE A DISSERTATION: SOME BAD STRATEGIES IN DATA-GATHERING AND WRITING UP. A proportion of PhD students never finish their grammars. In my experience, the reason is seldom loss of interest, except when this is allied to long-term depression. Grammarians generally love their work. Sometimes the failure to finish is because of external factors — illness, lack of money, etc. — but leaving these factors aside, the main causes of incompletions in my experience are the following:

1. *Database addiction*. In many cases the symptoms of database addiction can also be spotted quite early. The student says he or she cannot begin serious writing until an extensive corpus of data has transcribed, annotated and analysed, or otherwise processed. At first the advisor accepts this line but time goes by and the database keeps growing and growing, and very little in the way of chapter drafts appears, the advisor realizes that something is wrong. The database has become an excuse for not writing up analyses, and has led to a form of writer's block.

2. *Theory addiction*. A commentator on a draft of this chapter refers to dissertation writers who, chronically insecure about their command of theory, fail to complete

because "there's always one more paper to read, or the theory shifts and the writer feels compelled to revise".

3. *Perfectionism*. Usually, it doesn't take long to spot a perfectionist. He or she is someone who is reluctant to show you any chapter drafts until they are close to perfect and who consequently makes very slow progress. The perfectionist spends an inordinate amount of time worrying about analyses and revising drafts, or talking about the problems of analysis, without actually getting much onto paper.

4. *The all chapters at once strategy*. In this strategy, you write fragments of most or all chapters at once. (I use 'chapter' loosely for 'section on a major topic'.) First you decide on a format for the grammar including chapter topics and subtopics. Then you search your database for examples that are relevant to each of the topics and sub-topics and insert these examples in the relevant subsections. Then you start to write bits of commentary on these examples. After a while you have written bits about sections 2 and 7 of chapter 3, sections 3 and 5 of chapter 4 and sections 2, 4 and 6 of chapter 6, and so on. No chapter is ever more than half finished. After a while the advisor discovers this is happening, tells you it is a bad method, and asks you to complete a draft of one chapter at a time so the advisor can give feedback. You say this is impossible because chapters 3, 4 and 6 are interdependent. Or you promise to change your ways but cannot actually bring yourself to do so.

A milder variant of type 3 is where the student submits drafts of a sequence of chapters to the advisor(s), say 2-3, and gets back comments on each but presses ahead with writing drafts of chapters 4, 5 and 6, without looking carefully at these comments. The upshot is that the faults of writing style, analysis, etc. exhibited in chapters 2-3 are repeated in 4-6.

Of course bad practices are not long-term problems if they can be changed. But when they reflect deeply ingrained character traits, it is a different story.

7. THE ROLE OF THE DISSERTATION ADVISOR(S). Up till now I have assumed that the voice of the advisor is constantly to be heard addressing the various issues discussed, but I have seldom made explicit reference to the duties of the dissertation advisor(s) (whom we call in the Antipodes, the supervisor(s)). Perhaps I should be more explicit. In the department at the Australian National University (ANU) where I have taught for the last 20 years, there is normally a panel of three supervisors. Two are primary supervisors: there is a chief supervisor or chair of the panel and a co-supervisor who has almost equal responsibilities with the chair. Both are expected to advise the student at all stages of the project. The third supervisor's main duty is to read the completed draft, once the two main supervisors are reasonably satisfied with this. At the ANU the thesis is sent to a set of external examiners, usually three. The student's thesis panel nominates the examiners but plays no part in the actual examination of the dissertation. In the USA the system is different. The committee consists of a chair, who is the principal advisor, and several other members, who play lesser roles but typically read the dissertation when it is complete. The completed dissertation is examined in-house by the student's committee.

There are times where the advisor(s) and student need to be in regular contact to discuss issues and other times where the student can proceed independently. In the beginning, when the details of the project is being planned, there should be regular exchanges. The same applies in the later stages, when the student is writing the grammar.

The task of describing a little-known language normally requires extended fieldwork, often in a remote place. In the case of students doing fieldwork in a context of which they have no previous experience it is desirable, but not always possible, for an advisor to accompany the student into the field on the first trip and stay for a time.

Faced with a student who expresses a wish to do a grammar for his/her PhD, I generally begin by discussing the pros and cons of such a choice as a PhD topic and whether the student has the training for it. No one should try to write a grammar without a good grounding in the fundamentals of descriptive linguistics. Ideally, one should have taken courses in and read widely in all the basic fields of descriptive and typological linguistics, and looked carefully at number of reference grammars. One should read what literature there is on the target language, and on the language family to which it belongs, and should also read such literature as there is on the way of life of the communities speaking these languages.

The advisor must assess whether a student is suited by training, abilities, work habits and temperament to complete all the steps in a grammar-writing project. Obviously, one should not encourage or accept a student to do such a project if s/he seems unsuited to the task. However, in the absence of a track record of having previously completed a similar task, e.g. a master's thesis, this is always a bit of a gamble. There are various stages where students can lose their way.

Unsurprisingly, the most successful PhD grammar writers are those who have all the desirable qualities: they have sharp and enquiring minds, are well-trained in theory and skilful in data-collecting and analysis. They are enthusiastic and hard-working, enjoy fieldwork and are well-organized. They have the good judgment and flexibility to recognize and accept good advice and to question that which is dubious.

One remembers with a warm glow those students who had all these qualities. One young woman, in particular, was good at everything and dedicated to her language but where she really stood out was in the way she organized her two main advisors (myself and my colleague Malcolm Ross) during the writing up stages. One would have to say she took charge and, in the nicest way, controlled her advisors like puppets on a string. She lived 400 km west of Canberra but each month would come to the city and stay for a week near the university and make appointments to see each of us separately. She would email each advisor in advance with a set of very clear questions to be discussed at the meetings. If there was a conflict of advice between advisors this would lead to further set of well formulated questions. On each visit she would leave chapter drafts with each advisor and expect us to have our comments on these drafts ready for her next visit. The two advisors never dealt with the same draft at the same time. The student arranged things so that advisor A read the first draft and, after she had revised according to A's comments, advisor B got the revised draft. All her requests were made with such politeness, and every piece of advice was used so well that we did not mind being exploited with maximum efficiency. It came as no surprise that an excellent dissertation was completed in quick time and published soon

after, or that in the many years since her PhD was completed she has maintained a close and productive association with the community whose language she described.

References

Aikhenvald, Alexandra A. 2008. *The Munambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.

Ameka, Felix, Alan Dench and Nicholas Evans, eds. 2006. *Catching language: the standing challenge of grammar-writing*. Berlin: Mouton de Gruyter.

Bowden, John. 2001. Taba: *Description of a South Halmahera language*. Canberra: Pacific Linguistics.

Bowern, Claire. 2008. *Linguistic fieldwork: a practical guide*. New York: Palgrave MacMillan.

Bugenhagen, Robert. 1995. *A grammar of Mangap-Mbula, an Austronesian language of Papua New Guinea*. Canberra: Pacific Linguistics.

Dixon, Robert M.W. 2009-10. *Basic linguistic theory.* vol. 1. *Methodology*. (2009); vol. 2. *Grammatical topics* (2011); vol. 3. *Further grammatical topics*. (2012). Oxford/New York: Oxford University Press.

Dol, Philomina. 2007. *A grammar of Maybrat, a language of the Bird's Head Peninsula, Papua Province, Indonesia*. Canberra: Pacific Linguistics.

Dryer, Matthew S. 2006. *Descriptive theories, explanatory theories, and basic linguistic theory*. In Ameka et al., pp. 207-234.

Eades, Domenyk. 2005. *A grammar of Gayo, a language of Aceh, Sumatra*. Canberra: Pacific Linguistics.

Ezard, Bryan. 1997. *An Austronesian language of the Milne Bay area, Papua New Guinea*. Canberra: Pacific Linguistics.

Farr, Cynthia. 1999. *The interface between syntax and discourse in Korafe, a Papuan language of Papua New Guinea*. Canberra: Pacific Linguistics.

François, Alexandre. 2001. *Contraintes de structures et liberté dans l'organisation du discours. Une description du mwotlap, langue océanienne du vanuatu*. Thèse doctoral. Université Paris IV–Sorbonne.

Huddleston, Rodney and Geoffrey Pullum. 2002. *The Cambridge grammar of the English Language*. Cambridge: Cambridge University Press.

Hyslop, Catriona, 2001. *The Lolovoli dialect of the North-East Ambae language*. Canberra: Pacific Linguistics.

Jones, Alan. 1998. *Towards a lexicogrammar of Mekeo (an Austronesian language of western central Papua)*. Canberra: Pacific Linguistics.

Klamer, Marian. 1994. *Kambera, a language of Eastern Indonesia*. Canberra: Pacific Linguistics.

Lichtenberk, Frantisek. 2008. *A grammar of Toqabaqita*. (2 vols) Berlin: Mouton de Gruyter.

Mosel, Ulrike. 2011. Morphosyntactic analysis in the field: a guide to the guides. In Nicholas Thieberger, ed. *The Oxford handbook of linguistic fieldwork*, 72-89. Oxford: Oxford University Press.

Mosel, Ulrike and Even Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Scandinavian University Press.
Noonan, Michael. 2006. *Grammar writing for a grammar reading audience*. In Payne and Weber, 351-365.
Obata, Kazuko. 2003. *A grammar of Bilua, a Papuan language of the Solomon Islands*. Canberra: Pacific Linguistics.
Payne, Thomas E. 1997. *Describing morphosyntax. A guide for field linguists*. Cambridge: Cambridge University Press.
Payne, Thomas E and David J Weber, eds. 2006. Perspectives in grammar writing. Special issue of *Studies in Language* 30(2). Benjamins: Amsterdam/Philadelphia.
Pensalifini, Rob. 2003. *A grammar of Jingulu, an aboriginal language of the Northern Territory*. Canberra: Pacific Linguistics.
Quick, Phil. 2007. *A grammar of the Pendau language of central Sulawesi, Indonesia*. Canberra: Pacific Linguistics.
Quirk, Randolph, Sidney Greenbaum, Jeffrey Leech and Jan Svartvik. 1974. *A grammar of contemporary English*. London: Longmans.
Rice, Keren. 1989. *A grammar of Slave*. Berlin: Mouton de Gruyter.
Rice, Keren. 2006. *A typology of good grammars*. In Payne and Weber, 385-415.
Sharp, Janet. 2004. *Nyangumarta, a language of the Pilbara region of Western Australia*. Canberra: Pacific Linguistics.
Teng, Stacey Fang-ching. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Canberra: Pacific Linguistics.
van den Heuvel, Wilco, 2006. *Biak. Description of an Austronesian language of Papua*. Utrecht: LOT.
van Klinken, Catharina. 1999. *A grammar of the Fehan dialect of Tetun*. Canberra: Pacific Linguistics.
van Staden, Miriam. 2000. *Tidore, a linguistic description of a language of the North Moluccas*. PhD dissertation, Universiteit Leiden.
Wegener, Claudia, 2008. *A grammar of Savosavo, a Papuan language of the Solomon Is*. MPI Series in Psycholinguistics 51. Niemegen: Max Planck Institute for Psycholinguistics.

Andrew.Pawley@anu.edu.au

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 25-52
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4583

3

The data and the examples: Comprehensiveness, accuracy, and sensitivity

Marianne Mithun
University of California, Santa Barbara

Good grammars are read by diverse audiences with a wide variety of interests. One might not write a reference grammar in exactly the same way for all potential users, but particularly in the case of under-documented and endangered languages, it is likely that whatever is produced now will be consulted for answers to questions beyond those originally anticipated. A good grammar can provide more than descriptions of patterns the grammarian has noted at the time of writing; the examples it contains can provide a basis for future discoveries and new uses. It thus makes sense to consider the types of data that might best meet the needs of current and future readers, some of which we cannot even imagine at present. For some purposes, sensitive, typologically-informed elicitation is necessary, while for others, material drawn from unscripted connected speech is crucial. Here the potential contributions of examples of each type are considered for descriptions of phonetics, phonology, morphology, syntax, discourse, prosody, language change, and language contact.

Writing a grammar can be one of the most demanding projects a linguist undertakes, but also one of the most exhilarating. On the one hand, it requires a broad mastery of general linguistics, a deep understanding of the language to be described, and dedication to a potentially monumental task. On the other, it can be immensely satisfying: an opportunity to see beautiful systems in all their richness and complexity, to watch their interactions, and to appreciate the language as a whole. Especially for those writing a grammar for the first time, there is much to think about before plunging in: the intended audience, the topics to be covered, the organization, the style, and more. The focus here will be on just one aspect of the enterprise: the data that form the heart of the work and the selection and presentation of the material in examples. If the data come from direct collaboration with speakers, it is useful to reflect on the kinds of examples that will form the basis of the grammar during both the documentation process and the grammar-writing process.[1]

Readers have traditionally approached reference grammars with certain expectations, anticipating basic descriptions of the sound system, word formation patterns, and sentence structures. Such information is still fundamental to a good grammar. But as the field of linguistics progresses and the community of grammar readers widens, ideas are evolving about the kinds of information that can move the field ahead and serve the needs of more users. Theoretical and technological advances are stimulating us to ask new questions and

1 I would l like to express thanks to Toshihide Nakayama and Hideo Sawada, who organized the International Symposium on Grammar Writing as part of the Linguistic Dynamics Science Project at the ILCAA, Tokyo University of Foreign Studies. The symposium stimulated us all to think more deeply about the process of grammar writing and prompted useful discussion. I am especially grateful to the Mohawk speakers who have contributed their insights into the issues discussed here, in particular Charlotte Bush, Josephine Horne, Kaia'titáhkhe' Jacobs, and Watshenní:ne' Sawyer.

ISBN: 978-0-9856211-4-8

providing new tools for answering them. At the same time, the looming threat of language loss is inspiring more and more communities to mount language programs aimed at the documentation, preservation, and especially revitalization of their heritage languages. Distinctions between academic and community scholars are blurring: dedicated scholars coming from both directions are bringing increasing expertise and sophistication to their work. All are interested in clear, rich descriptions of the language, though sometimes their priorities differ. All of these developments mean that more kinds of readers are approaching grammars with a wider range of hopes.

Good grammars are read by a variety of linguists, including those specifically interested in phonology, morphology, syntax, discourse, typology, language change, and language contact, as well as relations among language, culture, and thought, and more. Reference grammars are increasingly consulted by those undertaking language revitalization projects, as a basis for planning language classes, preparing classroom materials, developing larger scale curricula, and compiling pedagogical grammars. They may be picked up by other community members simply interested in their heritage, by cultural anthropologists, by ethnohistorians, and others. One might not write a reference grammar in exactly the same way for all of these audiences: a grammar aimed at syntacticians might differ from one designed for language teachers. One written for typologists might differ from one aimed at interested community members. But particularly in the case of underdocumented and endangered languages, it is likely that whatever is produced will be consulted by a wider audience than originally anticipated. It is thus useful to try to imagine, when assembling the data that will form the basis of the grammar and choosing the examples that will ultimately appear in it, the kinds of information that might best meet the varied and evolving needs of current and future users.

There was a period in linguistics when sparseness was highly valued: students were sometimes told that no more examples should be included in a discussion than the bare minimum necessary to justify particular statements. Now, with more widespread recognition that the issues of concern to linguists are constantly evolving, and the acceleration of language loss in the world, the importance of rich data is becoming increasingly recognized. The examples in a grammar should provide evidence for the generalizations made, but they can also serve as a basis for further discoveries. The number and complexity of examples that can and should be included in a grammar necessarily vary from one situation to the next. For some non-linguists, short descriptions with small numbers of simple examples may be more accessible. Sometimes there are practical limitations on the length of a printed book. But if the grammar is to be the only description of the language, particularly if the language is threatened, other issues may be in play. Quantity is crucial for revitalization projects: language teachers and learners need more than a few forms to understand and absorb a pattern. Because the audience for a good grammar will be diverse, examples will be consulted for a variety of purposes, often beyond those envisioned by the author. Every example must thus be accurate on all levels: phonological, morphological, syntactic, lexical, and pragmatic.

Of course different languages show complexity in different areas of structure and use. Accordingly, good grammars will vary not just in the complexity of their examples, but also in the distribution of this complexity. Some points to consider when assembling data and examples for a grammar are discussed in the following sections. Not all are equally

relevant for all languages, all situations, or all times: languages are disappearing, the circumstances in which they are spoken are evolving, and technological possibilities are expanding. Many of the points discussed here are illustrated with examples from Mohawk, an Iroquoian language from northeastern North America, but for the most part, the details of the examples are less important than the principles they were chosen to exemplify.

1. BASIC DOCUMENTATION. If the grammar is to be based on material assembled by the grammarian in collaboration with speakers, it is useful to begin thinking about the kind of data that will provide a good foundation at the outset of the documentation work. With accelerating progress in technologies for audio/video recording and data management has come increased attention to procedures for language documentation. Himmelmann characterizes the field of language documentation as 'concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties' (2006:v). Discussions of issues involved in documentation can be found in Chelliah (2001), Mithun (2001, 2007), and Chelliah and de Reuse (2011), as well as the collections in Gippert, Himmelmann & Mosel (2006), Grenoble & Furbee (2010), the Language Documentation & Description series edited by Peter Austin and published by the Hans Rausing Endangered Languages program, and especially in the online journal *Language Documentation & Conservation* (http://nflrc.hawaii.edu/ldc/). These works cover such topics as kinds of data that should go into the record, recommendations for recording equipment and practices, formulas for metadata, formats for portability, preservation, and dissemination, ethics, and models of collaboration between speech communities and outside scholars. These collections also provide useful lists of web resources, links to software for transcription and annotation, metadata and corpus management, fonts, encodings and keyboard assignments, and speech analysis.

Not all of the data collected during a documentation project will figure equally in a grammar. Neither elicited nor spontaneous data are sufficient on their own, and the optimal balance between the two is an ongoing matter of discussion. Elicitation can provide orderly paradigms and parallel forms for comparisons. But also important is as large a corpus as possible of unscripted connected speech in a range of genres, both monologue of various types and especially conversation. Different genres can provide evidence of different aspects of the language. Furthermore, spontaneous speech is more likely to be idiomatic, providing examples not just of basic grammatical structures, but of grammatical constructions along with the kinds of lexical substance that is conventionally attached to them.

2. BASIC PRESENTATION OF EXAMPLES. A very useful discussion of the selection and presentation of examples for grammars is in Weber (2007). Weber points out that examples must illustrate the claim being made, they must be accurate, they should illustrate a range of uses, they should be structurally diverse, and, where possible, they should be culturally interesting. One should be certain that they are free of gender bias, project a good image of the speakers and their culture, and do not embarrass particular individuals or groups. Weber provides some ordering principles, suggesting that examples that best illustrate the claim being made be ordered first, that simple examples appear before more complex ones, that more typical ones precede more unusual (marked) ones, and that ambiguous cases (those

that could be interpreted either in such a way as to support the claim or in some other way) be ordered last.

Weber further points out that examples should be framed in the grammar: their relevance to the point being made should be explained in the prose immediately before or after them. Claims and examples should be integrated, in order to avoid lengthy descriptions followed by lengthy series of examples.

Choices about the layout of examples and the amount of information to include with them will vary with the nature of the language, the situation, and the points being made. Particularly for languages with morphological complexity, examples are typically presented in a multi-line format. Some common conventions for formats and abbreviations for grammatical terminology are laid out in the Leipzig Glossing Rules, assembled by linguists at the Max Planck Institute for Evolutionary Anthropology and available at http://www.eva.mpg.de/lingua/resources/glossing-rules.php.

An example from Mohawk, which can show relatively elaborate morphology, is in (1).

(1) Mohawk example: Watshenní:ne' Sawyer, speaker p.c., Bridge 2.43[2]

Tsi	*niió:re'*	*tsi*	*kahnekí:io*
tsi	ni-io-r-e'	tsi	ka-hnek-iio
so	PRT-N.PAT-be.far-STATIVE	so	N.AGT-liquid-be.good.STATIVE
so	it is that far	so	it is liquid good

'The water was so good that

enwá:ton'	*nenhshnekì:ra'*.
en-w-aton-'	ne=en-hs-hnek-ihra-'
FUT-N.AGT-be.possible-PFV	the=FUT-2SG.AGT-liquid-consume-PFV
it will be possible	the you will liquid consume it

you could drink it.'

Here the top line presents the utterance as spoken, in the standard community orthography. Some publishers require that this line be in italics, while others prefer a basic roman font. Sometimes boldface is preferred, especially by communities who want to highlight the importance of the language being described.

2 Abbreviations for glosses follow the Leipzig Glossing Rules (http://www.eva.mpg.de/lingua/resources/). Additional abbreviations are AGT GRAMMATICAL AGENT, CONTR CONTRASTIVE, I INDEFINITE OR GENERIC GENDER, N NEUTER GENDER, PAT GRAMMATICAL PATIENT, PRT PARTITIVE, Z ZOIC GENDER. Unless otherwise noted, Mohawk examples are given in the standard practical orthography. The phonetic values of the symbols are essentially like those of the IPA, with the following exceptions. Orthographic <i> is a palatal glide [j] before vowels, but the usual high front unrounded vowel [i] elsewhere. The digraph <en> represents a nasalized caret [ʌ̃], and the digraph <on> a nasalized high back rounded vowel [ʊ̃]. The colon <:> indicates vowel length. Acute accents <ó> indicate stress and rising tone, and grave accents <ò> stress and falling tone. The apostrophe <'> represents glottal stop [ʔ]. Stops <t> and <k> are automatically voiced before other voiced segments. The sequence <ti> before a vowel is pronounced as an affricate [dʒ]. Further abbreviations are EP epenthetic vowel; LK linker (the vowel *-a-* inserted between noun and verb stems in compounds under certain phonological conditions, and between verb stems and derivational suffixes.

The next line is a parsed line, showing the internal morphological structure of each word. Here boundaries between morphemes are indicated with hyphens: en-w-aton-'. Boundaries between clitics and their hosts are shown by an equals sign: ne=en . . . If the language shows extensive phonological processes, the grammarian can decide how abstract the representations of the morphemes should be. The particular Mohawk words in (1) do not show many phonological processes. When the article *ne* occurs before a word beginning in a vowel, it often cliticizes and is reduced to *n=*. In (1) it is represented as spoken on the top line (*n*), but shown in its full form on the parsed line (*ne=*). Stress placement and vowel length are not properties of individual Mohawk morphemes, but rather of full words. Neither is represented in the basic forms of the morphemes on the second line.

The third line provides glosses, the meaning or function of each morpheme. Lexical glosses are given in roman type (be.good), and grammatical terms are given in small caps (STATIVE). When a single morpheme in the second line corresponds to a multi-word gloss in the third line, the words in the gloss are linked by periods: the gloss for the Mohawk *-iio*, for example, is given here as be.good.STATIVE. No period separates person and number, as in 2SG, the gloss for the second person singular pronominal prefix *-hs-* 'you'.

Because Mohawk morphology is sometimes complex, it can be useful to readers to have a word-by-word translation as well, as on the fourth line in (1).

Finally, the last line provides a free, idiomatic translation, usually surrounded by single quotation marks. It is important to keep all associated lines together on a page. Particularly when the order of information in the target language contrasts strongly with that in the free translation, it may be easier to present the full free translation as a separate unit at the end of the example.

Where publishers permit, it can be useful to use a smaller font for the interlinear analysis lines, like the second, third, and fourth lines in (1). Some readers are less interested in the analysis and find these lines distracting. In some electronic formats it is possible for readers to see only as much as they wish on a computer screen at one time. They may even be able to hover over certain elements to bring up further information, such as the morphological analysis of words, additional forms and/or meanings of a morpheme, the discourse context surrounding the example, or even sound. And possibilities are increasing all the time.

The Mohawk sentence in (1) is presented such that each group of lines represents a separate prosodic phrase or intonation unit, that is, it is arranged by prosodic rather than syntactic structure. Such an arrangement can be useful, displaying structure that would be lost if examples were broken into lines on a purely syntactic basis.

The layout in (1) is of course not appropriate for every purpose or every language. In a section of a grammar that lists distinctive sounds, single words and glosses are usually sufficient. In a paradigm, simple lists of related words followed by their meanings are usually most effective. In a discussion of lexical categories, on the other hand, an additional line might be useful that identifies parts of speech. If a language is usually written with a non-roman orthography, such as the Cherokee syllabary or Chinese characters, an additional line of transcription might be important. A language might have such simple phonological structure that a separate parsed line is not necessary: the top line could be segmented. Morphological or syntactic structure might be so straightforward that a separate line for literal translations, like the fourth line in (1), is unnecessary.

Finally, various additional kinds of information about the example may be useful. In (1), the speaker has been identified. Some speakers are willing to be credited for their contribution to the work, and their identity can provide important information about differences among dialects, ages, genders, etc. Other speakers prefer to remain anonymous, and of course their preferences should be respected. In this example, there is also an indication of where the sentence can be found in the corpus: it occurred 2 minutes and 43 seconds into a narrative now identified as the Bridge text. Such annotation can allow readers to check things for themselves and access further information such as discourse context and prosody.

3. PHONETICS AND PHONOLOGY. Examples of words containing each of the distinctive sounds in the language and their variants have long been a basic component of most grammars. Now that audio files can be included with grammars, and grammars can be published in electronic formats with embedded sound, more phonetic information can be included with the description. The accessibility of audio data is a wonderful advance for all readers, both those hoping just to learn about the language and those hoping to learn to speak it. Advances in tools for acoustic analysis are making new kinds of visual displays possible, such as vowel spaces and pitch traces.

The potential value of such displays can again be illustrated with Mohawk. The language shows a two-way tone contrast on stressed syllables. The tones are not simply level high and level low. Each has a distinctive pitch contour or melody. In long, stressed syllables, what is referred to as high or rising tone consists essentially of a rise in pitch. What is referred to as low or falling tone first rises more quickly to a point higher than a basic rising tone, then plunges steeply to a point below the baseline. The effect can be described in words, as here, but, a pitch trace can make things clearer. A comparison of the two pitch contours can be seen in Figure 1 created with Praat software (http://www.fon.hum.uva.nl/praat/). The word *onón:ta*' 'hill' [onú:daʔ] with rising tone was pronounced twice, followed by the word *onòn:ta*' 'milk' [onù:daʔ] with 'falling' tone, also pronounced twice.

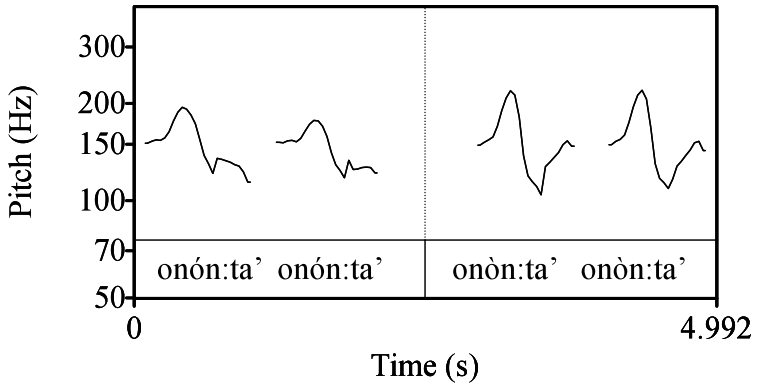


FIGURE 1: Rising versus falling tone contours in Mohawk.

Many languages show special phrase-medial or phrase-final phonological effects. After recording words in isolation, eliciting them in a carrier phrase so that the target word is phrase-medial can be revealing. For interested grammar readers, pitch traces can again provide an effective visual display. In Mohawk, stress is basically penultimate. Open syllables are lengthened and carry one of the two pitch contours seen in Figure 1 above. When a word with penultimate stress and rising tone on an open syllable is followed by another word, however, the pitch continues to rise into the next syllable. This effect can be seen in Figure 2. The first phrase consists of two words: *Sonkwehón:we kenh?* 'Are you Native?' The second is a single word, nearly the same as the first: *Konkwehón:we* 'I am Native'. Both show rising tone and length on the penultimate syllable *hón*. (The orthographic digraph *on* represents a high back nasalized vowel [ụ].) Phrase-medially however, as in the question here, the pitch continues to rise higher into the posttonic syllable *we*: *Sonkwehón:wé ken?*

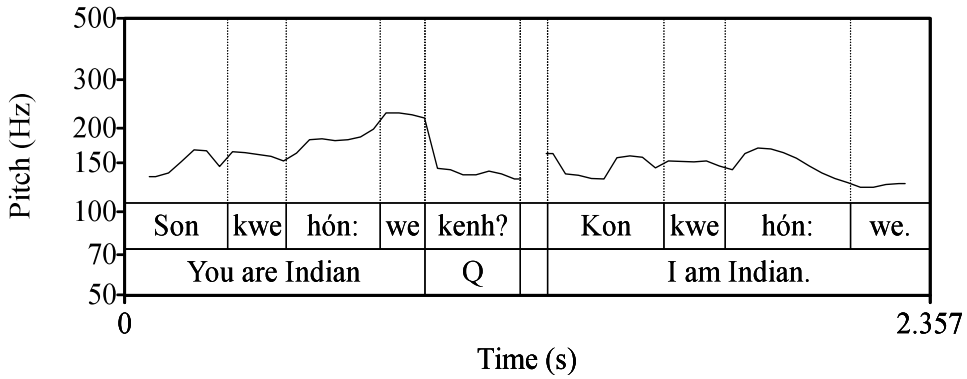


FIGURE 2: Mohawk pitch in context.

4. MORPHOLOGY AND LEXICALIZATION. Readers generally expect a good reference grammar to list all affixes or morphological processes, with their ranges of form and function. They expect a complete description of the variant shapes of each. Particularly where a grammar is going to serve multiple purposes and audiences, examples of all of these are important.

The point can again be illustrated with Mohawk. All Mohawk verbs contain a pronominal prefix identifying their core arguments: one argument for intransitives, two for transitives. I may discover that the second person dual prefix 'you two' of imperatives has the form *seni-* before *k*, but *tsi-* (IPA [dʒj]) before *a*.

(3) Mohawk dual pronouns

seni-ká:we	'Paddle, you two!'
tsi-atkáhtho	'Look, you two!'
tsi-átien	'Sit down, you two!'

I should check to see what shape this prefix takes before other consonants and vowels. I may look through my data and discover that the form *seni-* also occurs with all other consonants in the language. Rather than simply stating this fact, it is good to provide examples of each.

(4) Mohawk dual pronouns before consonants

seni-tákhe	'Run, you two!'
seni-hrárho	'Pull ashore, you two!'
seni-ráthen	'Climb up, you two!'
seni-nóhare	'Wash it, you two!'
seni-'niá:ken'n	'Escape, you two!'

But verb stems beginning with other vowels are rarer, and there may be no examples even in an extensive corpus of them with second person dual pronominal prefixes. If I have identified specific stems that begin with the crucial vowels, Mohawk speakers easily provide dual commands with them, so long as the combinations make sense.

(5) Mohawk dual pronouns before vowels

sen-itskó:tak	'Stay seated, you two!'
sen-è:iahre	'Remember, you two!'
sen-ó'kwat	'Dig, you two!'

If we generalized from the form in (3), we would be wrong—the form *tsi-* only occurs before the vowel *a*.

Restricting the number of examples to a small set can make a grammar quicker to read. In some cases it can also make it more difficult for readers to grasp the principle in question, particularly when the description is couched in a very formal framework. And it can result in missed opportunities: there may be complexities that are not evident from one form alone, complexities that were not anticipated at the time of writing. In addition, language teachers need more than one example to teach a pattern. Especially for purposes of language revitalization, rich exemplification is crucial.

Full inflectional paradigms, in which the stem is kept constant, are typically appreciated and heavily used. Presenting information about forms in two places, once grouped by category (second person dual), and once summarized in paradigms, may not be maximally parsimonious, but particularly where the patterns are complex, it can be space well spent.

There is a traditional expectation that inflection is fully productive, that forms exist for all possible combinations of stems and pertinent inflectional categories in a language. But gaps in inflectional paradigms do exist, and they can be difficult to discover from spontaneous speech alone: the forms in question simply never occur. They can sometimes be uncovered through elicitation, but it is crucial that the elicitation be sensitive. Speakers need to be aware that their sense of what actually exists in the language is precious. Specification of the distinction between possible and actual forms is a valuable part of the grammar. Gaps can be significant.

As noted, all Mohawk verbs contain a pronominal prefix identifying their core arguments, as in *rák-hsere's* 'he is chasing me'. Similar prefixes can be seen in kinship terms. Relatives are described in Mohawk not as possessions (like the English *my grandfather*), but as relationships. The senior member of a relationship is indicated much like the grammatical agent of a verb, and the junior member much like the grammatical patient of a verb.

	Verb	Kinship term
(6)	*rák-hsere's*	*rak-hsót=ha*
	M.SG/1SG-chase-STATIVE.DISTR	M.SG/1SG-be.grandparent.to=DIMINUTIVE
	'he is chasing me'	'he is grandparent to me' = 'my grandfather'
(7)	*rí-hsere's*	*ri-ièn'=a*
	1SG/M.SG-chase-STATIVE.DISTR	1SG/M.SG-have.as.offspring=DIMINUTIVE
	'I am chasing him'	'I have him as offspring' = 'my son'

The prefixes distinguish three persons, three numbers, and four genders. As a result, paradigms can be large: *shonkenihsótha* 'we two, our grandfather' = 'he is grandparent to us two' (M.SG/1DU), *ietshihsótha* 'you all, your grandmother' = 'she is grandparent to you all' (ISG/2PL), *shakotihsótha* 'their grandparents' = 'they are grandparent to them' (M.DP/3DP), and many more.

Gender is distinguished only in third person in Mohawk.

(8) Mohawk genders

Masculine	Male persons, a few obviously male animals such as 'bull', 'rooster'
Neuter	Inanimate objects
Zoic	Most animals, some female persons
Indefinite	Generic persons ('one, they') and other female persons.

Two different genders are used for referring to female persons: Zoic and Indefinite (also termed Feminine-Zoic and Feminine-Indefinite). The factors underlying the choice between the two are subtle and intriguing. All speakers agree, however, that when discussing one's grandmother or mother, only Indefinite verbs are used, as a sign of respect. For 'she is sewing', one would never say *ka-'níkonhs* (*ka-* 'she ZOIC)), only *ie-'níkhonhs* (*ie-* 'she INDEFINITE) if referring to one's grandmother. For 'I like her', one would never say *ke-nòn:we's* (*ke-* 1SG/ZOIC), but only *khe-nòn:we's* (*khe-* 1SG/INDEFINITE).

Paradigms of kinship terms are very large, as noted, but they show some surprising inflectional gaps. There are no Indefinite kinship terms for grandmothers or mothers, even though the forms they would have if they did occur are obvious.

(9) Kinship term gaps

	ak-hsótha	'she (ZOIC) is grandparent to me' = 'my grandmother'
no	**ionk-hsótha*	'she (INDEFINITE) is grandparent to me'
	ake-'nisténha	'she (ZOIC) is mother to me' = 'my mother'
no	**ionke-'nisténha*	'she (INDEFINITE) is mother to me'

These gaps are particularly surprising in light of the fact that in verbs, only Indefinite forms are used for these very persons. In fact the prefixes on kinship terms and associated

verbs do not match. In (10) below, the prefix on 'my grandmother' is Zoic, but the prefix on 'she is sewing', referring to the same person, is Indefinite.

(10) Kinship mismatches

Ak-hsótha	*ie-'nikhonhs.*
ak-hsot=ha	ie-'nikhon-hs
Z/1SG-be.grandparent.to=DIMINUTIVE	I-sew-HAB
she (Z) is grandparent to me	she (I) is sewing

'My grandmother is sewing.'

It turns out that the verb forms are an innovation. The Indefinite category was originally used only as a generic: 'one, they'. It retains this function in all of the modern languages in the family. At a certain point, however, people started using it as a sign of deference to certain women. This innovation has now worked its way through the entire Mohawk verbal paradigm: all verbs have both Zoic and Indefinite forms. But it has not fully penetrated the kinship paradigms. The Mohawk terms for 'my grandmother' and 'my mother' are holdouts: only the original Zoic forms are used. (These originally referred to all third persons, before the introduction of special Masculine pronouns.) The lag in development of the grandmother and mother forms is not altogether surprising. These are among the earliest terms learned and used by children, and among the most frequent. They are learned as chunks and deeply entrenched, rather than assembled online, thus making them especially resistant to change.

A sensitivity to the difference between actual and possible forms is crucial. Actual forms are those that are known and used spontaneously by speakers. Possible forms are those that a speaker may be able to create on demand. As morphological patterns are uncovered, it can be easy for everyone to be carried away with the magical regularities. Speakers can unwittingly create new forms by analogy, without considering whether the forms are actually used. A grammar should provide a record not simply of what could exist in the language, but of what speakers recognize as established entities in their lexicon, words they have heard. Careful elicitation practices are key here: speakers need to know that their intuitions about what is actually said are valued.

In the domain of word formation (derivation and compounding), the distinction between actual words and possible words is all the more important. Derived forms that actually exist provide a record of concepts that speakers have codified. It should go without saying that non-native grammarians should not invent forms, particularly derived forms, even when they feel they have understood the general principles in play. The verb in (11) was provided by one linguist to demonstrate that 'unaccusative' verbs allow incorporation of their 'subjects'.

(11) Mohawk word?

Wa'-ka-wí:r-en'-ne'

FACTUAL-N.SG-baby-fall-PFV

'The baby fell.'

Surprised to see this example in print, I consulted a group of speakers. Their reactions were strong. "That's absolutely terrible!", "Even the worst speaker wouldn't say this!", "Oh my gosh!" (and worse). They commented that this word had obviously never been presented to any speaker for approval. Even though Mohawk is polysynthetic, with many long, complex words and highly productive morphology, speakers have a keen sense of which words are part of the language and which are not. The verb stem used when an animate being falls is actually *-ia't-en-'n-* 'body-lie-INCHOATIVE', with incorporated noun stem *ia't-* 'body'.

(12) Mohawk word
Wa'-ka-ià:t-en-'n-e'.
FACTUAL-N.SG-body-lie-INCHOATIVE-PFV
'She fell.'

This stem *-ia'ten'n-* is lexicalized, a recognized vocabulary item. It reflects a larger pattern whereby verbs are classified as having physical, mental, or abstract effects, by means of the incorporated nouns *ia't-* 'body', *'nikonhr-* 'mind', or *-rihw-* 'matter'. It is especially unfortunate when invented examples like that in (11) become part of the published record of an endangered language. If examples are chosen from a corpus of actually occurring forms, they can serve as a valid resource for those seeking to learn the language, as well as for those seeking to discover further generalizations about the language, like the body/mind/matter classification.

5. SYNTAX. It has been observed that descriptive grammars written during the first half of the 20th century tended to concentrate on phonology and morphology, while those written later include more extensive discussions of syntax (Cristofaro 2006:138, Rice 2006:239, and others). Much of this difference can be attributed to progress in our understanding of the kinds of syntactic constructions that exist in different languages and the ways they differ. But when describing and exemplifying syntactic structures, even greater care must be taken in selecting data than when describing phonology and morphology. Phonological and morphological structures tend to be more automated, with components usually below the consciousness of speakers. For the most part, speakers produce complex words on demand without danger of phonological or morphological 'mistakes': few stumble over choices among allophones or allomorphs. Strings of words, however, are another matter.

One way in which elicited and constructed examples can be problematic is the distribution of information over lexical categories. A linguist offered the example in (13) as evidence that Mohawk *-'ke* is an adposition. (Spelling, glossing, and analyses have been adjusted here to conform with community conventions and standard analyses.)

(13) Mohawk adposition?

Thí:ken	*o-nont-á-'ke*	*ió-hskats*	*ne*	*o-kwir-e'=shòn:'a.*
that	N.SG-hill-LK-place	N.SG-be.pretty	the	N-tree-NOUN.SUFFIX=DISTR

'On that hill, the trees are pretty.'

The argument being made was that NPs cannot be adjoined to a clause unless they bind some gap or pronoun inside that clause, but that NPs governed by an adpositional element are not subject to this restriction. The sequence 'on that hill' must thus be an adpositional phrase. Even if the sentence were acceptable (which speakers agree it is not), it would not prove the principle. The morpheme *'ke* is actually a derivational suffix that derives nouns referring to places. It is not relational: it does not specify a grammatical relation like case endings or adpositions in other languages. Mohawk nouns with such derivational endings can serve as core arguments of clauses, not just adverbs: 'I know that town (town-place)', etc. A reasonably-sized corpus of unscripted speech would show this. Some placenames contain this ending, and some do not. Furthermore, derived terms for places can even serve as the input to further derivation. The name of one Mohawk community, for example, is *Kahnawà:ke*. The term for residents of the community was formed by the addition of another derivational ending. The term for 'heaven' is literally 'sky place'. The term for 'angel' is literally 'sky place resident'.

(14) *Kahnawà:ke*	*Kahnawa'kehró:non'*
ka-hnaw-a'ke	ka-hnaw-a-'ke=hronon'
N-rapids-LK-place	N-rapids-LK-place=resident
'Rapids place' = *Kahnawà:ke*	'*Kahnawà:ke* resident(s)'

(15) *karonhià:ke*	*raronhia'kehró:non'*
ka-ronhi-a-'ke	ra-ronhi-a-'ke=hronon'
N-sky-LK-place	M.SG-sky-LK-place=resident
'heaven'	'angel'

The creation of the sentence in (13) above seriously distorts the syntactic structure of Mohawk.

Because most good grammars are going to be consulted by an ever-widening range of readers with varying purposes, it is crucial that all data in them be accurate on every level. The sentence in (13) was published with errors in its transcription, segmentation, analysis, and glossing. But perhaps more serious is the fact that even if the individual words were accurate, the sentence is not Mohawk. One speaker explained politely, 'That's not something that would ever be said. Perhaps it was written by someone trying to learn. You might forgive them for that.' When asked what she might say in such a situation, she suggested (16).

(16) Mohawk counterpart: Kaia'titáhkhe' Jacobs, speaker p.c.

VERB	DEM
Ióhskats	*thí:ken,*
io-hskats	thiken
N.PATT-be.pretty	that
it is pretty	that

VERB
tiokwiró:ton'.
t-io-kwir-ot-on-'
CISLOCATIVE-N.PAT-tree-stand-DISTR-STATIVE
there it tree stands here and there

This sentence is instantly recognizable as robustly Mohawk. It exemplifies a common pattern of expression. Speakers manage the flow of information such that each significant new idea is introduced in a separate intonation unit or prosodic phrase. In this construction, a basic idea is first introduced by a verb, a complete clause in itself: 'it is pretty', followed by the demonstrative *thí:ken* 'that'. The demonstrative serves as a sort of place-holder, signalling that further elaboration is to come. The distribution of information over words is entirely different from that in (13), which contains a verb and two noun phrases, each with a determiner; (16) consists of two verbs plus a demonstrative. The speaker who suggested (16) noted that she could not imagine a situation where the hill would be mentioned in the same sentence. She hypothesized that if for some reason it were needed, it would be introduced in another sentence.

The linguist who produced the tree sentence in (13) also produced the sentence in (17) below as an example of a predicate nominal construction. (Again spelling and analyses have been regularized.)

(17) Mohawk lexical categories?

Kanónhsa'	*thí:ken*	*o'nerohkwa'kénha'*.
ka-nonhs-a'	thiken	o-'nerohkw-a'=kenha'
N-house-NOUN.SUFFIX	that	N-box-NOUN.SUFFIX=DECESSIVE

'That old box is a house.'

Asked about this one, speakers all agreed that someone might come up with this if they were just learning the language and trying to translate from English. The word *kanónhsa'* has the morphological structure of a Mohawk noun, but this word is not normally used for real houses. Nouns for immovable entities like buildings are incorporated into verbs. The normal way to refer to a house, if it is not incorporated into another verb, is in (18). As one speaker commented, 'A house can't just be there in a vacuum'.

(18) Standard Mohawk term
kanónhsote'
ka-nonhs-ot-e'
N-house-be.standing-STATIVE
'(standing) house'

The Mohawk speaker who provided (16) above suggested that to try to convey the meaning apparently intended for (17), one might say something like (19).

(19) Mohawk idiomatic alternative: Kaia'titáhkhe' Jacobs, speaker p.c.

Wahatenonhsónnia'te'	*thí:ken,*
wa-ha-ate-nonhs-onni-a't-e'	thiken
FACTUAL-M.SG.AGT-MIDDLE-house-make-INSTR.APPL-PFV	that
he house.made himself with it	that

'He made himself a house with it, that thing,

o'nerohkwa'kénha'.
o-'nerohkw-a'=kenha'
N-box-NOUN.SUFFIX=DECESSIVE
former box
box.'

This is the same Mohawk construction seen earlier, actually more interesting grammatically than that in the invented example in (17). The first prosodic phrase ends in the demonstrative *thí:ken* 'that', promising further elaboration to follow. It shows that ideas expressed in nouns in English are often expressed in verbs in Mohawk. It also shows a different conception of the argument structure, in which humans play the grammatically prominent roles, typical of Mohawk.

It might be countered that (19) would not illustrate the structure that (17) was constructed to show, which is true. But there is a deeper issue here that grammarians should consider. Languages can differ syntactically in ways beyond word order or conditions for omitting pronouns. They can vary in how ideas are distributed over lexical categories, over predicates and arguments, over clauses, and over sentences. If the data in a grammar consist only of constructed examples and sentences translated from another language, we stand to miss much of what that language has to teach us beyond what we already know. In her article 'Writing culture in grammar in the Americanist tradition', Jane Hill (2006) eloquently discusses various ways grammars might reflect the culture of which a language constitutes a part. We can do more than provide translation equivalents of what we consider theoretically relevant sentences in another language. We can strive to capture glimpses of how speakers package their thoughts, what they choose to say and how they choose to say it.

The choice of examples of complex sentences raises similar issues. A substantial corpus, with a variety of kinds of speech represented, is likely to provide ample exemplification of complex constructions, probably of more types and greater complexity than one might think to elicit. At the same time, not all constructions one might think of will necessarily surface. Another linguist recently noted the theoretical importance of Chomsky's 1977 article 'On *wh*-movement', suggesting that investigating parallels in other languages might lead to further understanding of their syntax. Chomsky contrasted the two English sentences in (20) to show that movement of the *wh*-word to the front of a sentence is unbounded, as in (20a), but that the presence of a noun phrase like *the story* in (20b) blocks the movement because of a subjacency violation. The brackets [ ] indicate the position from which the question word is said to have moved.

(20) a. What did Susan say Mary thought John should persuade Bill to buy [ ]?
b. **What* do you believe the story that Mary bought [ ]?

For some languages, translating these sentences then asking for grammaticality judgments might produce clear answers and interesting results. For others, it risks clouding the picture. Just as progress has been made in our understanding of syntactic patterns, so too, are advances being made in our understanding of discourse patterns. It is not surprising that sentences comparable to those in (20) do not occur in even a substantial corpus of unscripted Mohawk speech. Speakers do not combine long strings of ideas like those in a single sentence, for systematic reasons involving information structure. Should we fill in the gap with elicitation? It might be possible to find a Mohawk speaker who could be persuaded to translate them into Mohawk, or to give grammaticality judgments about an interviewer's translation of them. But we might ask what such judgments would represent. Speakers of most endangered languages are bilingual: that is usually part of the story of endangerment. There are still excellent, articulate Mohawk speakers, though all of them are also good speakers of English. Faced with hypothetical sentences like these, even the speakers themselves cannot know how much they are tapping into their intuitions about English. If Mohawk translations of such sentences became part of the record of the language, they would certainly misrepresent its discourse and information structure, that is, the way speakers actually distribute information over clauses and sentences.

6. DISCOURSE AND INTERACTION. Spontaneous connected speech offers insight into fundamental features of a language, in many cases the kinds of features that make the language special. All connected speech is not the same, however. Particular features often show different degrees of elaboration in different kinds of discourse. It is important to draw examples for a grammar not just from monologue, but also from interactive conversation, normally the kind of speech that is both the most frequent and of most relevance to revitalization projects.

The importance of genre can again be illustrated with examples from Mohawk. The passage in (21) opened a story written by a group of Mohawk language teachers, all excellent first-language speakers.

(21) VERB
Tewakhwishenhé:ion
te-wak-hwish-enhei-on
DUPLICATIVE-1SG.PAT-strength-die-STATIVE
my strength has died
'I was tired

PARTICLE	VERB	VERB
sok	*iohsnó:re'*	*onkità:wha*'.
sok	io-hsnor-e'	wa'-w-ita'w-ha-'
so.then	N-be.fast-STATIVE	FACTUAL.1SG.PAT-sleep-ANDATIVE-PFV
so then	it is fast	I went to sleep

so I quickly went to bed.

PARTICLE	VERB
Sok	*wa'-k-atà:sw-aht-e'*
sok	wa'-k-ata'sw-a-ht-e'
so.then	FACTUAL-1SG.AGT-go.out-LK-CAUS-PFV
so then	I extinguished

Then I turned off my light

PARTICLE	VERB
tanon'	*ia'ká:rate'*.
tanon'	ia'-k-arat-e'
and	TRANSLOCATIVE-1SG.AGT-lie-PFV
and	I lay down there

and lay down.'

Morphologically and syntactically, the passage is fine. The words are all well formed and idiomatic. The distribution of information over nouns and verbs is typical of Mohawk speech; there are few nouns in the entire story. There are just two particles that relate ideas to each other: *sok* 'so then' and *tanon'* 'and'.

Compare the style of (21) above to that of (22) below, from a conversation. A group of friends were discussing an old man they had known as children.

(22) Charlotte Bush, speaker p.c., Onkw A 41.00

PARTICLE	VERB
Tanon'	*raonkwe'táksen*.
tanon'	raw-onkwe-'t-aks-en
and	M.SG.PAT-be.a.person-NMLZ-be.bad-STATIVE
and	he is person bad

'And he was cross.'

Watshenní:ne' Sawyer, speaker p.c.

VERB	PARTICLE	PARTICLE
Rorihwakwénienhs	*nen'*	*nè:'e;*
ro-rihw-a-kweni-enhs	nen'	nè:'e
M.SG.PAT-matter-LK-be.able-HAB	that	that
he is matter competent	CONTRASTIVE	

'He was respectful though;

VERB	PARTICLE	
rorihwakwenienhstòn:ne	*nek*	*tsi*
ro-rihw-a-kweni-enhst-onhne	ne-k	tsi
M.SG.PAT-matter-LK-be.able-HAB-PAST	the-only	as
he had been matter competent	but	

he used to be respectful but

PARTICLE	PARTICLE	PARTICLE	VERB	PARTICLE
khere'	*kati'*	*kenh*	*thitewana'kón:nihskwe'*	*wáhi'*
khere'	kati'	kenh	thi-te-wa-na'kw-onni-hskwe'	wahi'
perhaps	in.fact	Q	CONTR-1INCL.AGT-PL-anger-make-PAST.HAB	TAG
I guess	in fact	Q	we used to make him mad	TAG

I guess in fact we used to make him mad, didn't we.'

This passage contains noticeably more particles than that in (21). In the first line uttered by the second speaker, the sequence *nen' nè:'e* marks a focus of contrast with the preceding comment. In the second line, the particle *nektsi* 'but' announces a contrast to come in the third line. In the third line, the particle *khere'* is a sort of inferential indicating that the speaker is imagining the reason behind the man's crossness. The particle *kati'* signals that this sentence is relevant to the preceding discussion. The particle *kenh*, normally the yes/no question marker, here serves to suggest slight doubt. The final particle *wáhi'* is a tag, comparable to English 'isn't it' or here 'didn't we'. This is an example of one of the common interactive uses of the Mohawk tag *wahi'*. This was a co-constructed narrative. The tag signaled that the speaker was not setting herself up as the only one knowledgeable about the topic, and solicited participation from the others. She was successful: her comment was met with affirmation from the others, one of whom then added to the story.

Data from elicited translations, and even from texts constructed by speakers sentence by sentence as in (21) above, are typically poor in the devices that speakers use spontaneously to shape their messages, highlight significant information, background peripheral or familiar ideas, mark contrasts with previous statements or general knowledge, show links to previous threads of discussion or fresh starts, interact with their audience, and much more.

7. PROSODY BEYOND THE WORD. A key element of linguistic structure is prosody. Technological advances have made it possible to observe patterns of pitch, intensity, rhythm, and phonation closely and even quantitatively if desired, and to include descriptions of these patterns in grammars. Examples were seen in Section 3 of visual displays of the pitch countours associated with distinctive tones on Mohawk words. Similar displays can add helpful information to descriptions of larger stretches of speech. The relation between grammar and prosody is not isomorphic: in some cases the two run in parallel and reinforce each other, but in others they convey different information. Neither can be fully predicted from the other.

Prosody is typically a significant component of question constructions, for example, but the prosody of questions varies across languages. With acoustic analysis, we can see the prosodic patterns that accompany different kinds of questions and include visual representations of them in the grammar. English yes/no questions tend to show a rise in pitch. Their Mohawk counterparts generally do not. The difference can be seen by comparing the two frames in Figure 3, produced by the same bilingual speaker. The first shows the pitch contour of the Mohawk question *Tentéhse' kenh*?. The peak occurs on the stressed syllable of the first word. The contour contrasts sharply with its English counterpart 'Will you be back?', which shows a steady rise.

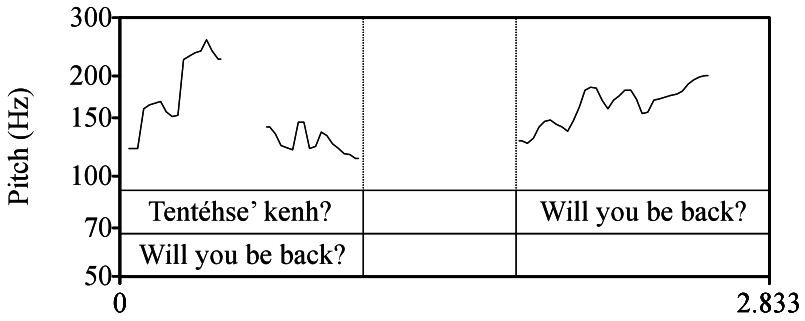


FIGURE 3: Mohawk and English prosody

Mohawk question-word questions show a similar fall in pitch, as can be seen in Figure 4.

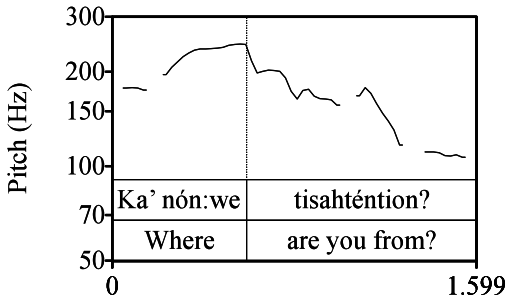


FIGURE 4: Mohawk lexical gap question

Tag constructions with *wáhi'* also show a final fall in pitch, as can be seen in Figure 5, a pitch trace of the final line of example (22) above.

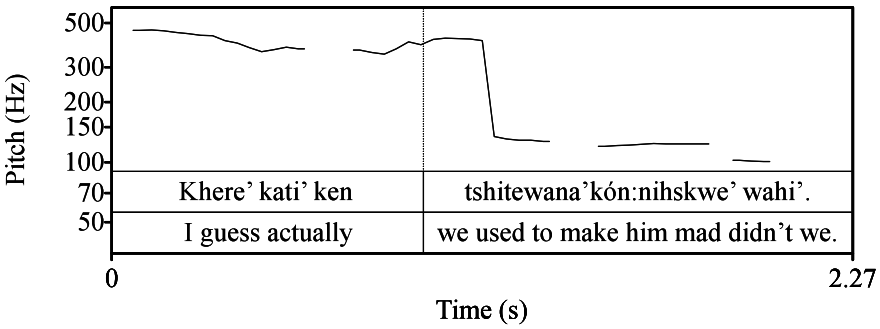


FIGURE 5: Falling pitch in tag construction

Prosodic patterns can also be revealing for studies of syntactic complexity. Looking at words in print, one might come to the conclusion that Mohawk speakers simply speak in sequences of simple sentences. Consider example (23) below.

(23) Mohawk: Billy Two Rivers, speaker p.c., B2R 50.17

Wahskwé:ni'	*á:re'*	*thé:nen'*	*wahsì:ron'*	*kenh?*
wa-hs-kweni-'	are'	othenen'	wa-hs-ihron-'	kenh
FACTUAL-2SG.AGT-be.able-PFV	again	something	FACTUAL-2SG.AGT-say-PFV	Q
you managed	again	anything	you said	Q

'You couldn't manage to say anything different, eh?'

Both of the verbs, *wahskwé:ni'* 'you managed' and *wahsì:ron'* 'you said it', are finite and could be used on their own as independent sentences.

Wahskwé:ni'.
'You managed it.'

Wahsì:ron'.
'You said it.'

The free translation, later provided by another speaker who participated in the conversation, indicates that the utterance was understood as one complex sentence. The prosody shows the same structure: the two clauses 'you managed' and 'you said something' were integrated under a single intonation contour, with a pitch reset only at the beginning, on the stressed syllable of the first word of the first clause (*wahskwé:ni'* 'you managed'), and a continuous fall until the end of the second clause. (The precipitous drop here is due to the falling tone on *wahsì:ron'* 'you said it'.)

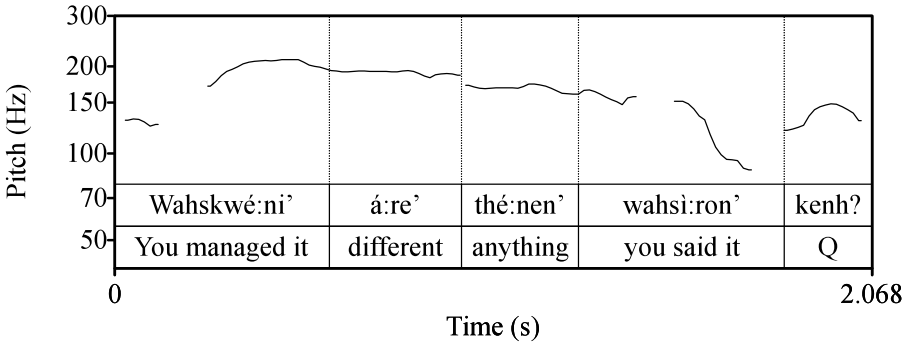


FIGURE 6: Single integrated prosodic envelope for complex sentence

Examples of prosodic structure are all the more important in descriptions of discourse and information structure. The basic unmarked prosody of a Mohawk sentence shows a

progressive descent in pitch from one stressed syllable to the next. The pitch trace for example (24) is in Figure 7.

(24) Basic simple sentence: Watshenní:ne' Sawyer, speaker p.c., Onkw A 3.35

Ó:nenhste'	*ken'k*	*nikontihnenié:son's*	*tanon'*
o-nenhst-e'	ken'=k	ni-konti-hneni-es-on's	tanon'
N-corn-NOUN.SUFFIX	small=only	PARTITIVE-Z.PL-height-be.long-STATIVE.DISTR	and
corn	just small	so they are variously tall	and

'The corn is very short and [it all seems to be doing poorly].'

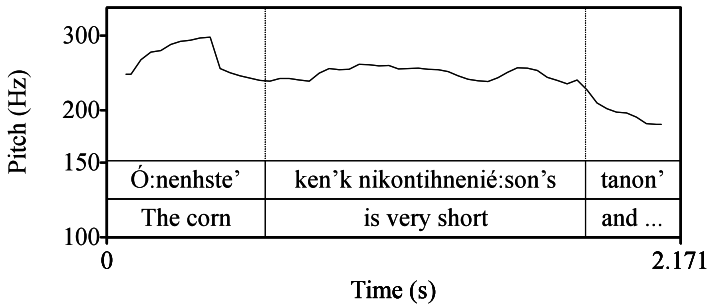


FIGURE 7: Basic declination over a prosodic phrase

When examined in print alone, the example in (25) appears to show the same structure.

(25) Topicalization: Watshenní:ne' Sawyer, speaker p.c., Onkw A 41.15

Akhsotkénha'	*wahonwatinónhsani'*	*iatathróna'*.
ak-hsot=kenha'	wa-honwati-nonhs-a-ni-'	i-atat-hrona-'
Z/1SG-be.gp.to=DECESSIVE	FACTUAL-I/3PL-house-LK-lend-STATIVE	M.DU-REFL-be.with-STATIVE
my late grandmother	she house lent them	they two are with each other

'My late grandmother rented a house to a couple.'

But this is a topicalization construction, used when the discourse topic is shifted to a different referent. The construction is usually used when the new topic has already been mentioned or is associated with another referent under discussion. This sentence was part of a lively conversation among a group of half a dozen people. It was the opening to an anecdote. The speaker had just said "I have another story". She then continued to recount her grandmother's experience.

This topicalization construction shows a distinctive pitch contour. It begins on a high pitch, but after the topicalized element, here *Akhsotkénha'* 'my late grandmother', there is a brief break, then a pitch reset on the stressed syllable of the following clause, here *nón*.

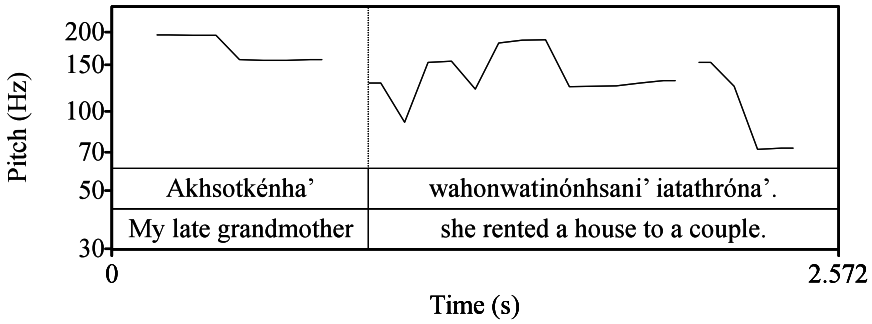


FIGURE 8: Topicalizaton

The previous example in (24), 'The corn is very short', was not a topicalization construction. The speakers had been discussing a trip they had taken the day before and the things they had noticed along the way. The corn was not mentioned again.

A number of other constructions are distinguished essentially by intonation. Now that the tools are available for us to see the patterns and display them, it makes sense to consider including such displays in grammars.

8. TYPOLOGY AND THE SELECTION OF EXAMPLES. An awareness of the kinds of grammatical categories and patterns that recur crosslinguistically is an important tool for grammar writers, making it possible for them to recognize patterns more quickly, identify points of general linguistic interest, and know which issues to pursue further. It is also useful when it comes to selecting examples for the grammar. In her 2006 article 'The organization of reference grammars: A typologist user's point of view', Sonia Cristofaro provides a good illustration of such benefits. As she notes, Givón (1980, 2001) and others have observed that the forms of complement constructions can vary across and within languages.

(26) Complementation scale for English: Givón (2001:43)

i.	Co-lexicalized complement	*She let go of the knife.*
ii.	Bare-stem complement	*She let him go home.*
iii.	Infinitive complement	*She wanted him to leave.*
iv.	For-to complement	*She'd like for him to leave.*
v.	Subjunctive complement	*She wished that he would leave.*
vi.	Indirect quote complement	*She said that he might leave later.*
vi.	Direct quote complement	*She said, "He might leave later."*

When multiple complement constructions coexist within a language, various factors affect their distribution. One is the semantics of the matrix predicate.

Noonan (1985, 2007) distinguishes the following classes of complement-taking predicates.

(27) Complement-taking predicate types: Noonan (1985, 2007)
- i. Utterance: *say, tell, report, promise, ask*
- ii. Propositional attitude: *believe, think, suppose, assume, doubt, deny*
- iii. Pretence: *imagine, fool, pretend, make believe*
- iv. Commentative (factives): *regret, be sad, be odd, be significant, be important*
- v. Knowledge: *know, discover, realize, find out, forget, see, hear, dream*
- vi. Fear: *be afraid, fear, worry, be anxious*
- vii. Desiderative: *want, wish, desire, hope*
- viii. Manipulative: *cause, force, make, persuade, tell, threaten, let, order, request, ask*
- ix. Modal: *can, be able, ought, should, may, be obliged*
- x. Achievement: *manage, dare, remember to, happen to, get to, try, forget to, fail, avoid*
- xi. Phasal (aspectual): *begin, start, continue, keep on, finish, stop, cease*
- xii. Immediate perception: *see, hear, watch, feel, imagine*

Armed with a framework like this, the grammar writer can select examples of complementation such that each matrix type is represented. Examples can first be sought in unscripted speech, then gaps can be checked with judicious elicitation.

9. LANGUAGE CHANGE AND THE SELECTION OF EXAMPLES. Languages are constantly evolving, as speakers work to make sense out of the patterns they perceive, repair apparent irregularities, and extend those patterns to express new ideas. Many such developments are gradual. A particular construction might appear first only in certain lexical contexts, then spread to others lexical item by lexical item. A frequently-used collocation may become ever more frequent, used in more and more situations, and as a result become more general in meaning. Erstwhile syntactic constructions may be used so often that they begin to blend and erode phonologically, ultimately evolving into grammatical and discourse markers. This dynamism is a fundamental feature of any living language, one that should be included as part of the description where possible. A rich body of examples from spontaneous speech in a grammar can often provide a snapshot of such processes in progress.

An example of such a process can be seen in developments of the Mohawk verb root *-ehr-* 'think, believe, want'. Like other verb roots, it appears in verbs in various aspects and tenses, and with all persons and numbers. It often occurs as the matrix verb of complement constructions.

(28) Mohawk verb *-ehr-* 'think, believe, want': Joe Deer, speaker, Sose 1.20

Wà:kehre'	*akwé:kon*	*tenkhenonhwará:ton'*.
wa'-k-ehr-e'	akwek-on	t-en-khe-nonhwaraton-'
FACTUAL-1SG.AGT-think-PFV	be.all-STATIVE	DUPLICATIVE-FUT-1SG/3PL-greet-PFV
I thought	all	I will greet them

'I thought I would greet everybody.'

(29) Mohawk verb *-ehr-* 'think, believe, want': Billy Two Rivers, speakerp.c., B2R 59.55

Áhsehrek

a:-hs-ehr-ek

OPTATIVE-2SG.AGT-think-CONTINUATIVE

'You would think

tóka	*ráonha*	*iahawennà:reke'*.
toka	raw-onha	i-a:-ha-wenn-a-hrek-e'
maybe	M.SG.PAT-alone	TRANSLOCATIVE-OPTATIVE-M.SG.AGT-word-LK-push-PFV
maybe	himself	he would word push

maybe he himself would be pushing for the language.'

(30) Mohawk verb *-ehr-* 'think, believe, want', Kaia'titáhkhe' Jacobs, speaker p.c. B2R 11.54

Kwah	*<u>í:kehre'</u>*	*thi,*	*á:re's*	*kawennókeri*.
kwah	i-k-<u>ehr</u>-e'	thiken	are'-s	ka-wenn-okeri
just	PROTHETIC-1SG.AGT-<u>think</u>-STATIVE	that	again-DISTR	N-word-be.gathered
really	I think	that	again	it is word shrunk

'I really <u>think</u> that the words are compacted.'

The examples above are from conversation, but similar examples are easy to elicit. In spontaneous speech, however, the verb *í:kehre'* 'I think' appears in constructions that differ to varying degrees from prototypical complement constructions. Sometimes what would seem like a part of the complement clause precedes the matrix.

(31) Kaia'titáhkhe' Jacobs, speaker p.c., B2R 51.10

Ì:se'	*<u>í:kehre'</u>*	*sahtentiòn:ne'*
ise'	i-k-ehr-e'	sa-ahtenti-onhne'
2	PROTHETIC-1SG.AGT-think-STATIVE	2SG.PAT-leave-STATIVE.PAST
you	I think	you had left

'You, <u>I think</u> you were away.'

There is evidence that this verb is taking on modal meaning, indicating less than total certainty on the part of the speaker. The exchange below is interesting for two reasons. One is that the speaker utters the word *í:kehre'* twice in one sentence, once inside of a simple clause. The other is that another participant in the conversation then expressed agreement, but it was not with the apparent matrix 'I think', but rather with the apparent complement: 'they've planted pole beans'.

(32) Interaction: Sose 3.30

A	*Tanon'*	*<u>i:kéhre'</u>*	
	tanon'	i-k-ehr-e'	
	and	PROTHETIC-1SG.AGT-think-STATIVE	
	and	I think	

'And <u>I think</u>

ienakarótha'	*<u>i:kéhre'</u>*	*rotiiénthon*.
ie-nakar-ot-ha'	i-k-<u>ehr</u>-e'	roti-ient-hw-on
I.AGT-pole-stand-HAB	PROTHETIC-1SG.AGT-think-STATIVE	M.PL.PAT-lie-CAUS-STATIVE
one pole stands	I think	they have planted

they've planted pole beans.'

B	*Én:,*	*rotiiénthon*	*kwah*	*í:ken.*
	en:	roti-ient-hw-on	kwah	i-ka-i
	yes	M.PL.PAT-lie-CAUS-STATIVE	just	PROTHETIC-N.AGT-be
	yes	they have planted	really	it is

Yes, they really have.' (Not: 'Yes, you really do.')

Further examination of spontaneous speech shows additional developments. The verb now also appears in a much reduced from *khere'* with the meaning 'maybe, perhaps'. Speakers no longer feel that it contains the first person pronominal prefix *k-* 'I'. It often occurs as part of the sequence *khere'kati'ken*, apparently a combination of *khere'* 'perhaps' + *kati'* 'in fact, actually', and the interrogative *kenh* which adds uncertainty. Speakers are unsure about whether this sequence consists of three elements or just one.

(33) Modality: Joe Dove, speaker: Sose 9.20

Khere'kati'kenh			*tóka'*	*sahontenhní:non'.*
khere'	kati'	kenh	toka'	sa-hon-aten-hninon-'
perhaps	actually	Q	maybe	REPETITIVE.FACTUAL-M.PL.AGT-MIDDLE-buy-PFV
maybe			perhaps	they sold it again

'I guess maybe they sold it.'

Such progressions of grammatical development are generally not as evident in elicited or translated material. Elicited examples might be simpler and illustrate a basic grammatical point more clearly, but if all examples are elicited or constructed, the dynamism inherent in the language will be missed.

10. LANGUAGE CONTACT. The potentially powerful role of contact in shaping grammar is becoming ever clearer, as more detailed documentation of more languages is becoming available. Grammar writers often make a conscious effort to exclude all non-native features from their descriptions and examples. Particularly in the case of endangered languages, it is important to many communities to have a record of the traditional form of the language as it was spoken before the encroachment of a competing majority language. At the same time, bilingualism has long been the norm in many communities, even before contact with larger world languages. It can enrich languages, as bilingual speakers exploit the distinctions offered by two systems in order to express themselves more eloquently in each. But it can also erode the distinctivenesss of a minority language, as patterns are remodeled to mirror those of the majority language. In the end it is communities who have the most to say about what they hope to see documented in a grammar: the most traditional patterns of the heritage language, or the modern usage of skilled bilingual speakers.

The difference is not always obvious. Lexical borrowing can be evident, particularly when the source language is well known. Structural borrowing can be more difficult to spot and evaluate, but it can have more profound effects. Bilingual speakers may create patterns in one of their languages modeled on those in the other with no transfer of lexical substance. They may simply increase the frequency of an existing minor pattern in one language to match the frequency of a comparable major pattern in the other, or extend it to more contexts. But if most or all examples in a grammar are obtained through elicitation

and/or translation, the description may not even reflect the current state of the language. It can be difficult to determine whether the similarities are actually representative of the language or an artefact of the methodology.

Possible contact effects can be obscured even when speakers themselves are constructing examples. Even good speakers can produce structures during a translation process they would never utter spontaneously. When one skilled Mohawk speaker assembled a pedagogical grammar, he produced the question and answer pair in (34).

(34) a. *Í:seks kenh ne kanà:taro?*
you eat Q the bread
'Do you eat bread?'

b. *Í:keks tiótkon ne kanà:taro.*
I eat always the bread
'I always eat bread.'

The words are phonologically and morphologically well-formed. The question 'Do you eat bread?' correctly illustrates the position of the interrogative particle *kenh*, immediately after the first constituent. But the speaker who created this example would not talk like this.

The question in (34a) shows a word order similar to that in English, where direct objects routinely follow the verb. But constituent order in Mohawk is not governed by syntactic relations. It is pragmatic: essentially, the most newsworthy information appears early in the clause (often after various orienting and other discourse particles). In yes/no questions, the focus of the question appears initially, followed by the interrogative particle *kenh*, as here. But otherwise the word order in the question above is unusual, with its focus on the eating rather than the bread: 'Do you <u>eat</u> the bread?'. The use of the particle *ne* introduces a second complication. Mohawk *ne* often appears in the same kinds of contexts as English *the*, but it has a subtly different function: it indicates that the referent has already been mentioned or evoked in the current discussion: 'the aforementioned'. The question in (34a), presented in the grammar with no previous context, is thus pragmatically odd in several ways. It might be appropriate if you had been telling me that you bake a lot of bread, and I wanted to ask you whether you actually eat that bread. To ask a more general question about whether someone eats bread, a usual form would be one like that in (35), with the bread first and no *ne*.

(35) *Kanà:taro kenh í:seks?*
bread Q you eat
'Do you eat bread?'

The answer in (34b) above, *Í:keks tiótkon ne kanà:taro* is also unidiomatic, perhaps again reflecting English patterns. The word *tiótkon* 'always' normally supplies important information and tends to occur at or near the beginning of the clause in Mohawk, as in (36).

(36) *Hén:,* *tiótkon* *ne* *kanà:taro* *í:keks.*
yes always the aforementioned bread I eat
'Yes, I <u>always</u> eat bread.'

In this case, it is the process of assembling examples that has produced contact effects.

11. CONCLUSION. As our technology and understanding of language progress, so too can our ideas about the kinds of data that might be useful to a wider range of users and that can lead to new discoveries. We are learning more about the intimate relations between structure and substance: speakers do not simply know abstract grammatical patterns on the one hand, and lists of morphemes and words on the other. The strengths of bonds between constructions and particular lexical items probably fall along a continuum. We are learning more about relations between structure and context: speakers select morphological and syntactic constructions for a variety of reasons, often involving the discourse context and the interpersonal situation and goals. We are becoming increasingly aware of the role of prosody in syntax and discourse. We are also becoming more conscious of the constantly evolving nature of linguistic structures and the forces that shape them, both language-internal mechanisms and language contact. If the examples in the grammar are accurate on all levels of structure, they should be useful for learning more about all of these areas of inquiry.

There are now many good grammars that provide models of how to choose effective examples. Among the basic guidelines that have been discussed here are the following.

1) Nature of the data
   a. Drawn as much as possible from spontaneous connected speech, in a variety of genres, criticually including ample conversation
   b. Augmented by elicited examples for clear pronunciations of individual words, completeness of descriptions of allomorphy and paradigms, and illustration of contrasting structures
   c. Representative of the range of known typologically significant variables
   d. Accompanied by surrounding context where pertinent
   e. Generous in quantity
   f. Culturally appropriate, all else being equal

2) Presentation of examples
   a. Interlinear analysis and glossing for languages where this is not immediately obvious. Different amounts of interlinear information are appropriate for different languages. Interlinear lines may show such things as morphological segmentation, underlying forms, morpheme glossing, and/or literal word-by-word translations.
   b. Where appropriate, references to locations of the examples in texts and/or audio recordings that would allow the reader to see them in their discourse context or hear them.
   c. Insofar as possible, use of standard glossing conventions such as the Leipzig Glossing Rules.

3) Additional aids
   a. Figures showing such acoustic information as vowel spaces, pitch contours, vowel spaces, etc.
   b. Traditional paradigm tables

The special value of data from unscripted connected speech was recently brought home to me when I consulted two grammars of the same language. The first provides clear lists of phonemes and allophones, case endings, demonstratives, pronouns, and quantifiers. It contains examples of complex noun phrases combining all of those elements. There are lists of tense, aspect, and mood endings, reflexives and reciprocals, causatives, negatives, and passives. There are examples of simple sentences, conjoined sentences, relative clauses, adverbial clauses, and complement clauses. All of the kinds of structures a typologically-informed grammarian of this period would seek out are exemplified. The second grammar lists the same elements, though transcriptions are not always as clean. But the two grammars differ in a striking way. The first could be a description of a language almost anywhere in the world. The second is immediately obviously a grammar of a language indigenous to California. This second grammar contains numerous affixes and clitics not mentioned in the first, markers even a typologically sophisticated grammarian might not think to look for. They represent categories that are highly developed in languages indigenous to California, languages that are genetically unrelated but that have been spoken by peoples in close contact over centuries. The areal affiliation of the language comes across robustly in the examples, and not just because of mentions of acorns rather than rice. It is obvious from the ideas speakers chose to express, the distinctions they chose to specify, and the distribution of information over words, clauses, sentences, and larger stretches of speech. As might be suspected, examples in the first grammar were drawn almost entirely from elicitation, direct translations of English models, which they generally parallel word for word. Those in the second came from a vast collection of texts of varied kinds. In those it was the speakers who chose what topics to discuss and what to say about them. The writer of this second grammar was acutely tuned into the genetic and areal context of the language, and he was able to note significant similarities and differences between it and its closest genetic relatives and neighbors. What is perhaps surprising is the fact that this second grammar was written nearly a century before the first. Because of its grounding in extensive unplanned speech, it continues to provide valuable information about issues even this alert grammarian could never have thought of at the time.

References

Austin, Peter K. 2003 on. *Language description and documentation*. Endangered Languages Project. London, UK: School of Oriental and African Studies.

Chelliah, Shobhana L. 2001. The role of text collection and elicitation in linguistic fieldwork. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*. 152-165. Cambridge, UK: Cambridge University Press.

Chelliah, Shobhana L. and Willem J. de Reuse 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer.

Chomsky, Noam. 1977. On WH-movement. In Peter W. Cullicover, Thomas Wasow, and Adrian Akmajian (eds.), *Formal syntax*, 71-132. New York: Academic Press.

Cristofaro, Sonia. 2006. The organization of reference grammars: A typologist user's point of view. In Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 137-170. Berlin: Mouton de Gruyter.

Gippert, Jost, Nikolaus P. Himmelmann & Ulrike Mosel (eds.) 2006. *Essentials of language documentation*. Berlin: Mouton de Gruyter.

Givón, Talmy. 1980. The binding hierarchy and the typology of complements. *Studies in Language* 4:333-77.

Givón, Talmy. 2001. Verbal Complements and Clause Union. Chapter 12 of *Syntax*. II:39-90. Amsterdam: John Benjamins.

Grenoble, Lenore & Louanna Furbee (eds.) 2010. *Language documentation: Practice and values*. Amsterdam: John Benjamins.

Hill, Jane H. 2006. Writing culture in grammar in the Americanist tradition. In Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 609-628. Berlin: Mouton de Gruyter.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 1-30. Berlin: Mouton de Gruyter.

Mithun, Marianne 2001. Who shapes the record: The speaker and the linguist. . In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*. 34-54. Cambridge, UK: Cambridge University Press.

Mithun, Marianne 2007. Grammars and the community. In Thomas E. Payne & David J. Weber (eds.). *Perspectives on grammar writing*. 45-70. Amsterdam: John Benjamins.

Noonan, Michael. 1985. Complementation. In Timothy Shopen (ed.), *Language typology and syntactic description II: Complex constructions*. 42-140. Cambridge: Cambridge University Press.

Noonan, Michael. 2007. Complementation. In Timothy Shopen (ed.), *Language typology and syntactic description II: Complex constructions*, 2nd edn. 52-150. Cambridge: Cambridge University Press.

Rice, Keren. 2006. Let the language tell its story? Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 235-269. Berlin: Mouton de Gruyter.

Weber, David J. 2007. The linguistic example. In Thomas E. Payne & David J. Weber (eds.). *Perspectives on grammar writing*. 199-213. Amsterdam: John Benjamins.

mithun@linguistics.ucsb.edu

4

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 53-67
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4584

On the role and utility of grammars in language documentation and conservation

Kenneth L. Rehg
University of Hawai'i at Mānoa

The National Science Foundation warns that at least half of the world's approximately seven thousand languages are soon to be lost. In response to this impending crisis, a new subfield of linguistics has emerged, called *language documentation* or, alternatively, *documentary linguistics*. The goal of this discipline is to create lasting, multipurpose records of endangered languages before they are lost forever. However, while there is widespread agreement among linguists concerning the methods of language documentation, there are considerable differences of opinion concerning what its products should be. Some documentary linguists argue that the outcome of language documentation should be a large corpus of extensively annotated data. Reference grammars and dictionaries, they contend, are the products of language description and are not essential products of language documentation. I argue, however, that grammars (and dictionaries) should normally be included in the documentary record, if our goal is to produce products that are maximally useful to both linguists and speakers, now and in the future. I also show that an appropriately planned reference grammar can serve as a foundation for a variety of community grammars, the purposes of which are to support and conserve threatened languages.

1. INTRODUCTION.[1] During the past decade, an increasing number of linguists have taken up the task of creating a lasting, multipurpose record of the world's many endangered and minority languages. These efforts, commonly referred to as 'language documentation' or alternatively 'documentary linguistics', have drawn new attention to the tools, methods, and products of basic linguistic research.

The emerging field of language documentation arose in response to an increasing awareness that many of the world's approximately 7,000 languages are likely to be dead or moribund by the end of this century. How many languages will be lost is, of course, unknown; estimates range between 30 and 90 percent. What is certain, however, is that the empirical foundation of our discipline is rapidly eroding.

There is much to be done. Michael Noonan (2006:352) estimates that we have adequate documentation for approximately 500 languages, preliminary documentation in the form of short grammars and dictionaries for perhaps 2,000 languages, and little or no documentation for the remaining 4,500. If Noonan is right, this means that we have satisfactory documentation for only approximately 7% of the world's languages, typically those that are *least* endangered.

Clearly, then, the central challenge to the discipline of linguistics is to document as many endangered languages as possible, and, where appropriate, to assist in their maintenance. The issue I wish to address in this chapter is the role of grammars in language documentation. More specifically, I want to argue that a grammar that is produced for,

1 I wish to thank Carol Genetti, an anonymous reviewer, and the participants at the LingDy International Symposium on Grammar Writing for their many helpful comments on this chapter.

ISBN: 978-0-9856211-4-8

with, and by speakers of the target language can play an essential role in both language documentation and conservation.

2. THE ROLE OF REFERENCE GRAMMARS IN LANGUAGE DOCUMENTATION. While there is widespread agreement about the goals and methods of language documentation, not everyone agrees on what its products should be. The traditional goals of fieldwork were to produce a grammar, a dictionary, and a collection of texts, commonly prioritized in that order. Documentary linguistics, as conceived by Himmelmann, stands those goals on their head. He argues that the primary goal of language documentation is to build an extensive corpus of texts, while the position of grammars and dictionaries is less certain.

In part, this reordering of priorities is a consequence of how one defines language documentation. Himmelmann broadly characterizes language documentation and its goals as follows:

> Language documentation is concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties. (Himmelmann 2006:v)

> The goal [of language documentation] is to create a record of a language in the sense of a comprehensive corpus of primary data which leaves nothing to be desired by later generations wanting to explore whatever aspect of the language they are interested in… (Himmelmann 2006:3).

In Himmelmann's view (2006:17), the "well-established format for language documentation consisting primarily of a reference grammar and a dictionary…is, strictly speaking, a format for language *description* and not for language *documentation* proper".

While one can, of course, establish such a division of labor in theorizing about the tasks involved in compiling a comprehensive record of a language, in practice, one cannot easily separate language documentation from language description, nor is it necessarily desirable to do so. At the University of Melbourne, for example, language documentation and language description are seen as yin and yang components of an undertaking that, tongue in cheek, they suggest might be called 'descriptamentation'.[2] Similarly, at the University of Hawai'i, the goal of our program in language documentation and conservation is to bring data, documentation, analysis, description, and theory together in one seamless whole, which we simply call 'documentation'.

Drawing a distinction between language documentation and language description is, in a more fundamental sense, beside the point. The primary concern ought to be about what kinds of materials will best serve the potential users of the records we compile for languages. Himmelmann's characterization of the goal of a language documentation—that it provide a record of a language for 'later generations'—is well-taken. However, it runs the risk of being far-sighted in both senses of this term. That is, it is forward-looking, but, like visual far-sightedness, it may result in an inability to focus clearly on the near present. I believe that language documentation should be concerned not only with future utility, but with present value as well.

2 See <http://www.linguistics.unimelb.edu.au/research/projects/greatthings.html>.

What then might be the role of reference grammars in language documentation? I would argue that the answer to this question is contextually contingent. If one is working with a language spoken by only a very small number of elderly speakers, and if there is no interest in revitalizing the language, then the best strategy is probably to build as large a corpus as possible, based on an on-going analysis of the language to insure that critical data is not being overlooked. A reference grammar can come later.

Regardless of the vitality of a language, however, the compilation of a large corpus of data is an essential part of the record.[3] I do not contest this claim. However, if one is working with an endangered or threatened language that might be revitalized, then the development of a grammar that is comprehensive and theoretically-informed should be assigned high priority, *if* our goal is to produce a record of a language that is not only representative and lasting, but one that is also maximally *useful*. The question, of course, is useful for whom?

3. THE UTILITY OF REFERENCE GRAMMARS FOR LINGUISTS. Reference grammars are, or at least ought to be, useful to linguists, both in their roles as consumers and producers, as I discuss below.

3.1 LINGUISTS AS USERS OF GRAMMARS. One issue that has been given insufficient attention in the literature on language documentation is the usefulness of a corpus of data for which there is no grammar, or for which there is only a sketch grammar. It might be interesting to query those who manage the archives in which such corpora are stored to see to what extent linguists make use of them. I have never undertaken this task formally, but the anecdotal information I have is that they are, in fact, underutilized by linguists, presumably because linguists judge the effort required to make effective use these corpora to be incommensurate with the potential reward. This is hardly surprising. Except for those specialists who are working on a particular language, or perhaps a specific language family, linguists will always prefer to work with those languages for which we have reliable reference grammars (and dictionaries). A reference grammar provides a one-stop source of basic information on a language and, if available, is typically the first resource that a linguist will go to if his/her interest is in the grammar of the language.

This volume and other like it (e.g. Ameka, Dench, & Evans 2006; Payne & Weber 2006) are, in fact, typically focused on how reference grammars can be made more useful to linguists. Why? Because reference grammars are as basic to linguistic research as ingredients are to chefs. The impact on our discipline of compiling corpora *instead* of writing reference grammars is unlikely to be positive.

3.2 LINGUISTS AS AUTHORS OF GRAMMARS. While it is obvious that reference grammars are useful to linguists as users, it is perhaps less commonly observed that

3 Himmelmann (2006:24) speaks of 'economy of effort', suggesting that "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". I disagree. Much of the work involved in compiling a corpus for a language can be better carried out by trained native speakers, thus leaving the linguist free to undertake the analysis and description of the language.

reference grammars can also play an important role for linguists as the producers of such products.

First, it should be noted that the kind of careful analysis of data required to write a reference grammar provides at least a partial check on the adequacy of the corpus. Simply collecting a massive amount of data without a detailed analysis of its content is certain to result in an inadequate sample of the language. I witnessed an extreme example of this as a graduate student. A number of faculty and graduate students in anthropology were preparing to work in Melanesia and, during the course of a preliminary site visit, had recorded many hours of narratives in the language of the community in which they planned to work. No analysis of the data was done until they returned to the university, whereupon they discovered that that did not have a *single* example of how to ask a question. A basic rule-of-thumb among experienced field workers is that analysis must be an on-going task, so that one has a clearer idea of where the holes are in the data. The idea that one could collect a sufficiently large corpus that would provide answers to any question one might have about the data is simply unrealistic. Analysis provides a check on the adequacy of the data, and writing a description of the data provides a check on the adequacy of one's analysis.[4]

Second, the writing of a grammar provides the linguist with an important 'discovery procedure'. The idea of discovery procedures emerged during the heyday of structuralist linguistics. The goal of such procedures was to provide the linguist with a set of tools which, if properly employed, would lead one to a unique and accurate analysis of a language. In phonology, this resulted in 'tests' such as those for 'minimal pairs', 'complementary distribution', and 'free variation'. While we now know that such discovery procedures are flawed and unreliable, working linguists also know that such procedures are nevertheless useful.[5] And working linguists who have written a reference grammar also know that writing such a grammar is a valuable discovery procedure in its own right. The challenge of writing down one's analysis of a language, in such a way that it is comprehensible to (some) others, invariably leads one to ask questions about the data that might otherwise not have arisen. (Ask anyone who has written a grammar.) The consequence is that the record of the language provided by the linguist is much enriched by this experience.

Another benefit to the linguist of writing a grammar is that it promotes 'whole-system thinking'. Much of the linguistics literature during the last four decades of the 20th century focused on developing theoretical claims about specific, and often narrowly defined, aspects of language, commonly based on limited data extracted from the work of others. While it is an incontrovertible fact that research of this nature has enormously advanced our understanding of language, it is also true, I think, that linguists who solely engage in research of this nature are likely to have a very different view of how languages work than linguists who have attempted to provide a comprehensive description of all aspects of the grammar of a single language. The latter attempt requires one to see how isolated facts fit together to

4 Himmelmann also advocates analysis of the data one is including in a corpus, but it is unclear what level of analysis he has in mind. His comment (2006:28) that Hockett (1958) and Gleason (1961) might serve as excellent introductions to linguistics suggests that his views of the field substantially diverge from those held by most American linguists.

5 For example, these discovery procedures fail to provide one with a means by which underlying diphthongs can be distinguished from surface diphthongs. See Rehg 2007.

form the whole, and that experience typically has a transformative experience on how one understands language and works as a linguist. It should also be noted that lasting advances in linguistic theory are typically the consequence of an encounter with data that resist analysis within the frameworks of existing theories. Fieldworkers and writers of grammars commonly experience what might be called 'theory lag'. The challenge, then, becomes one of revising the theory so that it accommodates the data—hopefully never the reverse.

Finally, the professional value of writing a grammar must be considered. In most academic institutions, contributions to the discipline, rightly or wrongly, are measured in terms of numbers of publications and the venues in which they are published. The building of corpora and work with endangered language communities typically, and most unfortunately, count for little. In some departments, perhaps especially in the United States, even the writing of reference grammars and dictionaries is not recognized as an important scholarly contribution. In response to this narrow and harmful conception of the field, the Linguistic Society of America recently endorsed a resolution put forward by its Committee on Endangered Languages and their Preservation that calls for recognition of the scholarly merits of language documentation.[6] That resolution notes that "the products of linguistic language documentation [including reference grammars and dictionaries] ...are…fundamental and permanent contributions to the foundation of linguistics, and are intellectual achievements which require sophisticated analytical skills, deep theoretical knowledge, and broad linguistic expertise".[7] Consequently, one can hope that those who do fieldwork will continue to build corpora, but, where appropriate, then take the next logical steps—to write grammars and produce dictionaries. And they should do so without guilt.

4. ARE REFERENCE GRAMMARS USEFUL ONLY TO LINGUISTS? Reference grammars are useful to linguists—or at least to some linguists.[8] Few will contest that claim. Indeed, the relatively small body of literature that exists on writing grammars typically focuses on how linguists can make grammars more useful to other linguists. And, if one's intended audience is other linguists, that is an appropriate concern. Other linguists may necessarily be one's only audience. A graduate student writing a grammar as a PhD dissertation will have other linguists as his or her primary audience—namely the members of the dissertation committee. Linguists writing grammars of languages, all of whose speakers are illiterate, or none of whom speak the language the grammar is written in, will also appropriately write for other linguists. But, are reference grammars necessarily useful only to linguists? Himmelmann (2006:19) observes:

> Grammars…provide little that is of direct use to non-linguists, including the speech community, educators, and researchers in other disciplines...

6 See <http://www.linguisticsociety.org/resource/resolution-recognizing-scholarly-merit-language-documentation>.

7 It is remarkable that the field of linguistics pays lip-service to the importance of language documentation, but, in some cases, discourages it in practice. One is reminded of Cicero's complaints about the philosophers of his day: "…impeded by their zeal for learning, they desert those whom they ought to protect" (Strange & Zupko 2004).

8 Reference grammars are especially useful for formalists and typologists. It should therefore come as no surprise that linguists working in these areas have been some of the strongest supporters of efforts to deal with the endangered language crisis.

Lise Dobrin (2009:619), too, has observed that there is a "great gap between academically produced knowledge about language on the one hand, and real-world problems on the other".

Clearly, these concerns are well-motivated. Even when circumstances permit writing to a broader audience—that is, in contexts where non-linguists might be able to make use of a grammar—linguists nevertheless all too often write only to other linguists. This shortcoming, however, is not, and need not be, true of all grammars. The question, then, is how can we develop grammars that are useful to a broader audience, including non-linguists? More specifically, the issue I wish to consider is how do we approach the task of writing grammars so that there is some hope they might be useful for the communities with which we work?

5. WORKING WITH COMMUNITIES. What I wish to propose here is that, where appropriate and possible, we write reference grammars that are accessible to the speakers of the language and then subsequently make use of these grammars to develop community grammars.[9] To illustrate this approach, I will provide a case study of a reference grammar that was written *on* Pohnpeian, both *for* and *with* speakers of this language, and then discuss a community grammar that was written *by* two Pohnpeians, utilizing the content of this reference grammar.[10]

5.1 THE UTILITY OF THE POHNPEIAN REFERENCE GRAMMAR. Pohnpeian is a Nuclear Micronesian (Austronesian) language spoken on the island of Pohnpei in the Federated States of Micronesia. At present, the language has approximately 28,000 speakers. While most speakers of the language under the age of 50 are, to varying degrees, bilingual in English, the use of Pohnpeian remains robust.

The *Ponapean Reference Grammar* (reflecting an older name for the language that was in use at the time this grammar was written) was developed at the University of Hawai'i as part of a project called the *Pacific Languages Development Project* (PLDP: 1970-1974). The PLDP targeted all the major and several of the minor languages of what at that time was known as The Trust Territory of the Pacific Islands.[11] Its goals were (1) to develop standard orthographies, (2) to produce reference grammars, (3) to compile bilingual dictionaries, and (4) to train Micronesians to serve as indigenous linguists. The grammar for Pohnpeian was written by the author of this chapter with the assistance of Damian Sohl, a native speaker who was a participant in this project.[12]

---

9 Of course, even reference grammars written exclusively for linguists can serve as the foundation for community grammars, assuming a linguist is available to explain the grammar to the community.

10 See Czaykowska-Higgins (2009:22-25) for an insightful discussion of various approaches to fieldwork. The model I am advocating here is the one she calls *community-based research*.

11 The Trust Territory of the Pacific Islands was later partitioned into the Federated States of Micronesia, the Republic of the Marshall Islands, the Republic of Palau, and the Commonwealth of the Northern Mariana Islands.

12 A sister publication in the form of a bilingual Pohnpeian-English Dictionary was published as Rehg & Sohl: 1979.

A second project, called the *Bilingual Education Program for Micronesia* (BEPM: 1974-1983), brought Micronesian educators to the University of Hawai'i to train them in the principles and practices of bilingual education and to teach them to use the literacy documents that were developed in association with the PLDP.[13] Consequently, prior to publication, I was able to use various drafts of this grammar in a seminar on the structure of Pohnpeian that I regularly taught in conjunction with this project. The grammar was thus reviewed, commented on, and revised as a result of input from approximately thirty native speakers over the course of approximately six years.[14]

The Ponapean Reference Grammar (PRG) was specifically targeted for speakers of Pohnpeian. In the first sentence of the first paragraph of the Preface, I noted:

> My purpose in writing this book has been to provide a description of the major grammatical features of Ponapean for the reader who has little or no training in the analysis of language. Although this work is intended primarily for native speakers of Ponapean who are bilingual in English, I hope it will also be useful to others whose interests have brought them to the study of this language.

I further commented:

> I have endeavored throughout this volume to keep its content as clear and as simple as possible….I have tried to minimize the usage of [technical linguistic terms], but where they permitted a better or more efficient explanation of the data, I have not shied away from them. I have taken care, though, to define each technical term as it is introduced and to illustrate it with numerous examples.

That is, this grammar was written using the tools of linguistic theory that were available at that time, but the analyses resulting from the use of those tools were presented so as to be comprehensible to a broad audience, most especially educated Pohnpeians. It is thus useful to bear in mind that one's theoretical framework need not dictate one's descriptive framework.

I hasten to point out that I take no credit for deciding on the audience for this grammar, nor for this approach to writing it. These were the guidelines for all the grammars published as part of this project. I also do not intend to hold the content of this grammar up as a model for others. It was largely written while I was a graduate student, and though the data it contains are accurate (having been examined by multiple speakers), the scope of the grammar, as well as the analyses it contains, could certainly be improved upon.[15] Further,

13 A third undertaking, called the Pacific Area Language Materials Development Project (PALM: 1975–1983) developed vernacular language materials in a variety of content areas for many of the languages of Micronesia, including Pohnpeian.

14 See Rehg 2004 for a brief description of the Micronesian projects in which this grammar was written and taught.

15 Some linguists have told me that it is not possible to write such a grammar, that explaining linguistic concepts to non-linguists would result in grammars of excessive length. I would recommend to those who take this position that they examine the grammars produced for the Micronesian languages to see how this task was managed and to judge for themselves the extent to which these grammars succeeded.

because this grammar was written for a broad audience, and, because the publication of these grammars and dictionaries was subsidized in part by the Micronesian government, they were made available at a very low cost and continue to be relatively inexpensive.[16] I fully recognize, of course, that the PRG was produced under highly advantageous circumstances that are far from typical. Nevertheless, a consideration of the ways in which this grammar has been useful has potential relevance for linguists and language communities elsewhere.

What then has this grammar been good for? Some linguists have been able to make use of it, perhaps especially phonologists. But, one would hope this to be a minimal outcome. The more central question for this chapter is of what use has it been to the Pohnpeian community? Because I have continued to work on Pohnpeian and have made multiple visits to the island since the grammar was published in 1981, I now feel qualified to provide at least a brief response to this question.[17]

First, the grammar has clearly impacted the way in which speakers, especially younger speakers, view their language. When I first began work on this grammar, teenage speakers of the language who were learning English would often tell me that Pohnpeian, unlike English, had no grammar. By this, they meant *both* that it had no written grammar *and* that there were no rules for speaking the language. I no longer hear such comments.[18] As many linguists report, the prestige of a language can be enhanced by providing it with a reference grammar, and a dictionary.

Second, the grammar has been utilized by both learners and teachers in the teaching of Pohnpeian to speakers of other languages. Pohnpei is currently the site of the capitol of the Federated States of Micronesia, and consequently many outsiders are resident on the island. Because Pohnpeians encourage others to learn their language, and because they are supportive of such efforts, the College of Micronesia periodically offers a course in spoken Pohnpeian. Peace Corps volunteers coming to the island also receive instruction in the language. As a consequence, the language is utilized by both native and non-native speakers. Clearly, this is an important factor in maintaining the vitality of the language.

Third, the grammar has served as the basis for on-going efforts to teach the standard orthography of Pohnpeian, both in the form of workshops as well as in courses at the College of Micronesia. The grammar is useful for this purpose because it contains a six page appendix that lists and explains the recommendations of two Pohnpeian Orthography Workshops that were conducted on Pohnpei in 1972 and 1973. Each of these recommendations is summarized and references are provided to those sections of the grammar that describe the structural properties of Pohnpeian that prompted the recommendations. More importantly, it is this feature of the grammar that gave rise to a *community grammar* written *by* two Pohnpeians, *for* a Pohnpeian speaking audience.

16 The Ponapean Reference Grammar—xv + 393 pages—currently sells for $26 dollars and can be purchased by the Pohnpei Department of Education at a discount of 40%.

17 I began work on Pohnpeian in 1968 while a Peace Corps staff member.

18 It is likely, however, that many young speakers are unaware of the PRG's existence, even though it is still in print. Current attitudes about the language are clearly a consequence of attitudes shaped in the past.

5.2 WHAT IS A COMMUNITY GRAMMAR? A community grammar, as described by Michael Noonan (n.d.), is "a kind of reference grammar created for, and sometimes by, members of a linguistic community as an aid to establishing [or reestablishing] a language in the schools, for teaching the language to adults, [etc.]".

At the meeting in Tokyo that spawned this volume, I gave a paper entitled 'FINE Grammars for Small Languages'. FINE is an acronym for what I believe to be the essential properties of a community grammar. These are:

F = *focused*. A community grammar should be written for a specific purpose, responding to what in Peace Corps jargon was once called a 'felt need'—that is, a need that is *felt* in the community, rather than one that is merely *voiced*, or worse, *imposed* from the outside.[19]

I = *interesting*. The grammar should be constructed so that it will engage and entertain its readers.

N = *naturalistic*. The grammar should be based on real data or, at the very least, realistic data.

E = *educational*. The grammar should not only engage its readers, but instruct them as well. That is, the grammar should have either an overt or covert pedagogical function.

While I have since abandoned the use of this acronym (acronyms have a way of becoming intellectual straightjackets), it nevertheless remains a useful mnemonic.

The type of FINE or community grammar that I had envisioned at that conference (among many possible types) was one designed to teach speakers of Pohnpeian about the structure of their language as it bears on the conventions used in the standard orthography. Pohnpeian has, in fact, been written since the middle of the 19th century and most Pohnpeians can read and write their language, but there is, at present, a great deal of inconsistency in how the language is written. With the introduction of Pohnpeian into the school system, however, the Pohnpei Department of Education has become increasingly concerned that all teachers and students employ the standard system that was devised for the language in the early 1970s.[20] Inconsistencies in spelling by teachers obviously cause problems for children who are learning to read and write. Further, many teachers are insecure about writing Pohnpeian, primarily because they do not know or understand the conventions employed in the standard orthography.

At this point, let me slightly diverge to talk about what I mean by an 'orthography'. First, I should note that an orthography is not the same as a phonemic transcription, nor is it just an alphabet. Good orthographies (and linguists do not always produce good orthographies) address all areas of the grammar. The alphabet requires a solid understanding

19 As Peace Corps and other community development workers can testify, it is often very difficult to determine what a 'felt' need is. Extended contact and interaction with a community are usually required before one can make such a determination.

20 See Rehg 2004.

of the phonology of the language, as well as of phonological variation. Word division requires a good grasp of the morphology and, to a lesser extent, the syntax of the language. In addition, orthographies must address such matters as punctuation, capitalization, the treatment of loan words, and more. Above all, the designer of an orthography must have a good understanding of the dynamics of the culture in which the orthography is to be employed. Perhaps more than any other undertaking, it is essential that the development of an orthography be carried out as a community-based endeavor.

So, in what sense might a book focused on orthography be considered a grammar? Well, first it is a community grammar with a pedagogical function. It necessarily deals with phonology, morphology, syntax, and dialect variation. Further, it aspires to teach its users not only about the orthography, but about the structure of Pohnpeian as well.

My belief when I gave this talk in Tokyo was that such a community grammar would/could meet the FINE criteria I previously outlined. Such a grammar would be *focused* on a *felt* need in the community, namely to support and teach the standard orthography that has been endorsed by both church and state. It could be written, I believed, so that it would be *interesting*, it would focus on *natural* data, the kinds of mistakes that people commonly make in attempting to use the standard orthography, and it would be overtly *educational*.

6.0 AN EXACT REPLICA OF A FIGMENT OF MY IMAGINATION. The talk I gave in Tokyo took place in December of 2009. Six months later, in May of 2010, I went to Pohnpei, primarily to work with colleagues there on a second edition of the Pohnpeian dictionary. In the course of our work, I brought up the idea of developing a community grammar for the purpose of supporting the teaching of the Pohnpeian standard orthography. One of my colleagues, Damian Sohl, was a participant in the PLDP project previously mentioned and holds a BA in Linguistics from the University of Hawai'i. He was a co-author of the Pohnpeian dictionary and an assistant in the preparation of the grammar, a consultant to the Pohnpeian Orthography Workshop, and had previously served as the Pohnpei State Director of Education. My other colleague, Robert Andreas, holds an MA in Linguistics from the University of Hawai'i and is currently a Professor in the Division of Education at the College of Micronesia. Both have extensive experience teaching workshops and courses in Pohnpeian orthography, and both have been strong supporters of it. Consequently, I was confident that they would support the idea of developing such a community grammar.

Much to my astonishment, I discovered that they had already written almost exactly the kind of community grammar that I had envisioned. It was, in the words of Elizabeth McCracken (2008), an "exact replica of a figment of my imagination". While I was in frequent contact with both Sohl and Andreas during the time they were developing this grammar, neither had previously mentioned it to me. Our correspondence had focused on matters related to Pohnpeian grammar and lexicography. I was, of course, delighted that they had undertaken this project and, I must confess, reassured that my idea about what kind of community grammar the people of Pohnpei might want was on target.

6.1 THE STRUCTURE OF THIS COMMUNITY GRAMMAR. The community grammar developed by Sohl and Andreas contains 103 pages of material on Pohnpeian orthography,

divided into ten 'lessons', all written in Pohnpeian.[21] The target audience is primarily teachers and students at the College of Micronesia, but it is, in fact, an invaluable resource for any speaker of Pohnpeian interested in learning the standard orthography of this language. The structure of these lesson varies somewhat, but each typically provides (a) a statement of goals and objectives, (b) information about one or more of the conventions employed in the standard orthography, along with a discussion of relevant aspects of Pohnpeian grammar, (c) a list of technical terms used in the lesson, and (4) a self-test on its content.

These materials are overtly pedagogical. They are designed to serve essentially as a textbook and as a reference source for those wanting to master the standard orthography of Pohnpeian. They are appropriately designed to meet this need, but, in fact, they go well beyond this purpose. What is especially interesting about these materials is that they not only provide information about Pohnpeian orthography, but they also use the teaching of orthography as a vehicle for teaching a very substantial amount about the structure of Pohnpeian. The scope of these materials is such that any Pohnpeian completing them will have a relatively sophisticated understanding of the major grammatical features of his/her language.

Of course, a common problem in describing the grammar of a language that does not have a long tradition of grammatical description is the lack of suitable technical vocabulary. In the case of the Pohnpeian, this problem was dealt with in three ways—(1) by 'Pohnpeianizing' English words already known to most Pohnpeians, (2) by extending the meaning of Pohnpeian words to fill lexical gaps, and (3) by introducing new terms from English.

It is likely that most Pohnpeians will already be familiar with some grammatical terminology as a result of their schooling, which places a heavy emphasis on the teaching of English. Therefore, words like 'consonant', 'vowel', 'noun', 'verb', and 'sentence' are certain to be familiar to any Pohnpeian with a high school education. In these lessons, however, such words are spelled so that they reflect the way a monolingual speaker of Pohnpeian would pronounce them. That is, they are adapted to confirm to the constraints of Pohnpeian phonology. Examples are:

English	Pohnpeian[22]
consonant	*kansonan*
vowel	*pawel*
noun	*naun*
verb	*perip*
adverb	*adperip*
sentence	*sendens*

Damian Sohl reports that, during the first workshop in which he used these materials, the participants laughed when he said 'perip' rather than 'verb'. They were unaccustomed to hearing this word pronounced as it would be by a monolingual Pohnpeian speaker.

21 This grammar is currently unpublished and has been distributed to students in Xeroxed form. (See Sohl & Andreas n.d.)

22 For an explanation of the conventions of Pohnpeian orthography, see the PRG or Rehg 2004.

However, he reports that, by the end of the first day, everyone was using 'perip' without hesitation and without being self-conscious about it.

A second strategy to solve the terminology problem was to extend the meaning of existing Pohnpeian words. Sometimes this was done by using an English loan in combination with a Pohnpeian word to create a compound with a new meaning. For example:

English	Pohnpeian
base vowel	*pawel poad*
insert vowel	*pawel peidaid*

Poad in Pohnpeian means 'planted'. Therefore, a 'base vowel' is a planted or rooted vowel. An insert vowel is one that is *peidaid*, meaning 'transported.'

Other technical terms were created using only Pohnpeian words. An especially interesting set employing the word *pwuloi* follows. *Pwuloi* is a noun meaning 'the part of the stem between the joints, of cane-like plants (like bamboo)'. It can also be used to refer to the stanza of a song, and it is additionally used as a numeral classifier in counting sections of a stem from joint to joint, or stanzas in a song. Examples follow.

English	Pohnpeian
phone/speech sound	*pwuloin ngihl* section-of voice
suffix	*pwuloimwur* section-behind
locative phrase	*pwuloin lokaiahn wasa* section-of utterance-of place
temporal phrase	*pwuloin lokaiahn ansou* section-of utterance-of time
relative clause	*pwuloin koasoai idengek* section-of speech-of lean (against)

In the preceding examples, I have glossed *lokaia* as 'utterance' and *koasoai* as 'speech'; the actual meaning difference between these two words, however, is quite subtle. *Lokaia* generally refers to a speech act that lacks the formality and completeness of one called *koasoai*, although at present these two words are often used synonymously. I did not fully understand this distinction before reading this community grammar.

Third, new technical vocabulary was introduced from English. In some cases, such forms were 'Pohnpeianized'. That is, they were spelled to reflect how a Pohnpeian monolingual speaker might pronounce them; for example 'enclitic' is rendered as *enklidik*. In other cases, English spellings were retained, as for example 'demonstrative pronoun' and 'prepositional noun'.

6.2 THE VALUE OF THIS COMMUNITY GRAMMAR. It is certain that this grammar represents a valuable contribution to the community. The view of some Pohnpeians, rightly or wrongly, is that their language will be accorded respect to the extent that it mirrors the attributes of major languages like English. Consequently, they are very much concerned about 'standards' for the language, in both its spoken and written forms. In the case of spoken Pohnpeian, there are already complex and widely-accepted notions of what constitutes excellence. A proficient speaker of Pohnpeian, therefore, is one who is controls all levels of honorific speech, has an extensive vocabulary, commands all oratorical styles, etc. Comparable standards for written Pohnpeian, however, are still in the formative stage, but being able to spell Pohnpeian 'correctly', in accord with the rules of the standard orthography, is considered by some to be an essential foundation for developing such standards. The community grammar developed by Sohl and Andreas supports the teaching of the standard orthography, but, in and of itself, it also serves as an example of 'best practices' in writing the language. Perhaps more significantly, it represents the beginning of an indigenous grammatical tradition.

This community grammar is also of value to linguists and other students of the language. Because this grammar is written in Pohnpeian, it provides a rich source of textual material in a relatively new genre. It further demonstrates to linguists working with small languages that our reference grammars can be put to practical uses that serve community needs, but only if we insure that they are accessible to speakers of the language and that some speakers are trained to use them.

6. 3 COMMUNITY GRAMMARS FOR THE FUTURE. It is easy to envision still other types of community grammars that could play a significant role in advancing the aspirations that the people of Pohnpei have for their language and culture. One idea that was suggested to me by an educator there is the development of a community grammar to teach honorific speech, or *meing* as it is called in Pohnpeian. Control of this speech style is considered essential if one is to be viewed as a person of consequence. In the PRG, we noted "since not all speakers of Ponapean are able to use honorific speech with equal facility, command of this speech style is typically equated with sophistication, cultivation, and the ability to speak Ponapean well" (Rehg & Sohl 1981:359).[23] When I mentioned this idea to a younger speaker of Pohnpeian, who in all likelihood did not yet fully command honorific speech styles, he reacted extremely positively and assured me that if I charged $5 for the book, I could make lots of money![24]

Other types of community grammars were also suggested to me during my last visit to Pohnpei. These include a pedagogical grammar designed specifically for second language learners, a comparative grammar that would discuss differences and similarities among Pohnpeian and neighboring languages, a contrastive grammar that would compare English and Pohnpeian grammatical structures, and a historical grammar that would discuss the origins of Pohnpeian and how it evolved. It is easy to think of still others, all of which would likely gain an audience on the island.

23 See also Rehg 1998 for comments on the acquisition of Pohnpeian honorific speech.

24 While completing the final draft of this chapter, I received a telephone call from a Pohnpeian who currently lives in Honolulu telling me of his plans to write a 'manual of meing' and asking for my advice. Clearly, this is an idea whose time has come.

7.0 CONCLUSION: WHAT ARE LINGUISTS GOOD FOR? So, what are linguists good for? The defining mission of our discipline is the scientific study of language, but, as I hope to have illustrated in this chapter, our work, properly envisioned and presented, can also play a valuable role in language conservation efforts.

I am by no means holding up the Pohnpeian experience as a model for others. Each field situation is unique and must be approached with great sensitivity to the dynamics and aspirations of the community. I am also not suggesting that linguists abandon their current research to take up the task of sustaining minority languages. What I am instead proposing is that, even those who are most deeply committed to the development of linguistic theory might try to combine that work with the documentation and conservation of endangered languages. The simple fact is that there are more endangered and threatened languages than there are linguists to work on them. Connecting with these communities and working with such languages clearly has the potential to enrich all concerned. As Dobrin and Good (2009:629) have noted: "Linguistics could come to more closely resemble fields like medicine and economics, where interplay between theory and practice is welcomed in adding to their richness, and where 'applied' forms of work are not seen as belonging to a separate discipline."

References

Ameka, Felix K., Alan Dench, & Nicholas Evans (eds.) 2006. *Catching language: The standing challenge of grammar writing*. Berlin: Mouton de Gruyter.

Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working with Canadian indigenous communities. *Language Documentation & Conservation* 3(1). 15-50.

Dobrin, Lise. 2009. SIL International and the disciplinary culture of linguistics: Introduction. *Language* 85(3). 618-619.

Dobrin, Lise, & Jeff Good. 2009. Practical language development: Whose mission? *Language* 85(3). 619-629.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In: Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel (eds), *Essentials of language documentation*, 1-30. Berlin & New York: Mouton de Gruyter.

McCracken, Elizabeth. 2008. *An exact replica of a figment of my imagination: A memoir.* Boston: Little, Brown and Company.

Noonan, Michael. 2006. Grammar writing for a grammar-reading audience. *Studies in Language* 30(2). 351-366.

Noonan, Michael. n.d. *Describe everything! or Description is not the same as standardization*. https://pantherfile.uwm.edu/noonan/www/Describe%20Everything% 21.pdf (5 November, 2009).

Payne, Thomas E. & David J. Weber (eds.) 2006. *Perspectives on grammar writing*. Studies in Language 30(2).

Rehg, Kenneth L. & Damian G. Sohl. 1979. *Ponapean-English dictionary*. Honolulu: University of Hawai'i Press.

Rehg, Kenneth L., with the assistance of Damian G. Sohl. 1981. *Ponapean reference grammar.* Honolulu: University of Hawai'i Press.
Rehg, Kenneth L. 1998. Taking the pulse of Pohnpeian. *Oceanic Linguistics* 37/2.323-345.
Rehg, Kenneth L. 2004. Linguists, literacy, and the law of unintended consequences. *Oceanic Linguistics* 43:498-518.
Rehg, Kenneth L. 2007. Does Hawaiian have diphthongs? And how can you tell? In: Jeff Siegel, John Lynch & Diana Eades (eds.) *Language description, history and development: Linguistic indulgence in memory of Terry Crowley*. Amsterdam: John Benjamins, 119-131.
Sohl, Damian G. & Robert Andreas. n.d. Padahk en lokaiahn Pohnpei. Xerox.
Strange, Steven K., & Jack Zupko. 2004. *Stoicisim: Traditions and transformations*. Cambridge: Cambridge University Press.

rehg@hawaii.edu

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 69-89
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4585

5

Sounds in grammar writing

Keren Rice
University of Toronto

While there has been much written on writing grammars in recent years, relatively little has been written on the place of sounds and their patterning in grammar writing. In this chapter I provide an overview of some of the challenges of writing about sounds, and discuss the kinds of information on sounds that are generally included in grammars. I then address what a grammar might ideally include on the sounds of a language, advocating the inclusion of sound files to augment the usual topics, increasing both the scientific merit and the human value of the grammar.

What is the role of phonetics and phonology, or, more generally, of sounds and their patterning, in grammar writing? In this chapter, I address how sounds have been treated in grammars, what aspects of sound must be covered in a grammar, and what areas of sounds are seldom addressed in grammars and should be. In other words, I examine how much and what types of information about sound a well-balanced grammar should contain.[1]

This chapter is organized as follows. I begin with a brief overview of recent work on what a grammar is and the larger context for grammar writing today. Following these preliminaries, I review the attention paid to sound in recent work on grammar writing, as well as some of the challenges in writing the sound sections of a grammar. I then provide a brief historical overview of how grammars have presented sound over time. I close with a discussion of what might be the 'ideal' representation of sound in a grammar, and then discuss briefly the core of what information about sound needs to be part of a grammar. The person who is most interested in what to include in a grammar may well want to turn immediately to section 4, skipping some of the background discussion on the role of sound in a grammar.

Grammar writing is a broad topic; in this discussion I focus on writing grammars of languages that are undescribed or underdescribed and, for the most part, grammars of languages that are endangered. This is important to keep in mind, although much of what follows is perhaps relevant, no matter what the status of the language.

1. SOME CONTEXT. This section is divided into two parts. In the first part, I briefly introduce the recent literature on grammar writing to establish a notion of what a grammar is, and in the second part I discuss the audience for a grammar.

1.1. WHAT IS A GRAMMAR? It is useful to begin discussion of sound in grammar writing by providing the larger context of what is viewed as a grammar today. There has been considerable interest in grammar writing in recent years (collections edited by Payne and Weber 2005, Ameka, Dench, and Evans 2006, as well as an earlier book, edited by Graustein and Leitner 1989). In the introduction to Ameka, Dench, and Evans, Evans and

1 Many thanks to Carol Genetti, Toshihide Nakayama, and Noboru Yoshioka for very helpful comments. This work was funded by the Canada Research Chair in Linguistics and Aboriginal Studies.

ISBN: 978-0-9856211-4-8

Dench (2006:1-2) begin by examining what a grammar is and who the audience for a grammar is, as well as the kinds of contributions that a grammar can make. They define the goals of a description grammar and its potential audience – "Each grammar seeks to bring together, in one place, a coherent treatment of how the whole language works, and therefore forms the primary source of information on a given language, consulted by a wide range of users: areal specialists, typologists, formal linguists, historical linguists, and members of the speech communities concerned." They also describe the many challenges of grammar writing, including respect for the genius of the language balanced with general knowledge of how languages work, finding a balance between rigor and readability, and meeting the needs of a variety of audiences.

Thus, the writer of a grammar has many responsibilities. Balancing these is a tall order, and part of the goal of this chapter is to examine what this balance might be with respect to sounds – What are the essential features of a language in terms of sound? How do these interact with other aspects of the language?

1.2. THE LARGER SETTING FOR GRAMMARS: LANGUAGE, LANGUAGE LOSS, AND AUDIENCES. In recent years, with the recognition of the decrease in transmission of many languages, the importance of grammars has become more and more evident, for both the community of linguists and the community of speakers or would-be speakers of a language.

As noted above, in writing about the audience for grammars, Evans and Dench (2006:1) point out that a grammar is likely to be consulted by a wide range of users, including various types of linguists and members of the speech communities concerned. The last of these is probably a fairly recent addition to the list: grammars today are often viewed as serving the needs of linguists and also as playing a key role in language conservation, language revitalization, and language reclamation. In an chapter on grammars and the community, Mithun (2005:281) asks whether a single grammar can serve all potential users. She goes on to say that, whether it can or not, it is likely to be called upon to do so. Thus, whatever the goals of the grammar writer, those who are engaged in revitalization and reclamation work with whatever materials are available, making them a potential audience. Thus, in thinking about the presentation of phonetics and phonology, we must keep in mind the needs of the various audiences, with the knowledge that the audience might someday be one that is interested in revitalization or reclamation of the language.

2. SOUND IN WORK ON GRAMMAR WRITING: A BRIEF OVERVIEW OF RECENT LITERATURE. In this section I review the type of direct attention that is put to sound in the recent writing on grammars. The two recent books on grammar writing, Ameka, Dench, and Evans (2006) and Payne and Weber (2005), include little material that directly addresses issues around the presentation of phonetics and phonology, or addresses phonetic and phonological issues in any depth.

Ameka, Dench, and Evans (2006) contains articles on a variety of general topics – the art and craft of grammar writing, the roles of native and non-native speakers in grammar writing, cross-linguistic grammatography, linguistic typology, basic linguistic theory, the role of theory in grammar, the grammar-lexicon trade-off, field semantics, diachrony and synchrony, polylectal grammars, writing culture in grammar – as well as articles on some specific topics – word order, function words, converbs, 'disposal' constructions in Sinitic

languages, and the morpheme *ma-* in Tagalog. Payne and Weber (2005) includes general articles – contextualizing a grammar, grammar and the community, collective fieldwork, growing a grammar – and articles on specific topics such as from parts of speech to the grammar. No article in either volume specifically addresses issues of sound alone. In Ameka, Dench, and Evans, the only article that has anything specific to say about phonology is the one by Mosel, on grammatography. Her mention is brief: the basics of the sound system of a language and the orthography deserve a place in a grammar. Mosel also addresses where the presentation of phonology fits in a grammar.[2]

The articles in Payne and Weber (2005) are, by and large, general in focus. Mithun, in her work on grammars and the community, discusses how the sound system of Mohawk might be presented in a layered way, beginning with a list of distinctive sounds, symbols, orthography, and an example, moving to a description of the phonetic properties of stress, tone, and so on, with sections on intonation and perhaps history of transcription practices and cognates in related grammars. This is written as a reflection on layering, not on the content of the phonology section of a grammar.

Noonan (2005), in his contribution to Payne and Weber, discussed with several linguists what it was important to include within a grammar, and the phonologists and phoneticians he interviewed spoke to the types of things that should be said about sounds:

- Standard IPA symbols
- Detailed instrumental documentation accompanying descriptive statements
- A full description of segmental and suprasegmental contrasts and an explanation for arriving at them
- Description of distributional patterns of elements of the phonology
- Paradigms illustrating morphophonemic processes
- Where practical, audio and video recordings of various genres should be included

This is the most explicit discussion of the needs of a grammar in terms of sound in these books. We will return to these points.

Slightly earlier work on grammars is similar in having little to say directly about sound. A 1989 collection edited by Graustein and Leitner contains articles on a number of topics including grammar at the interface of language, linguistics, and users, and linguistic theories and grammar writing (linguistic pragmatics, functional grammar, cognitive linguistics, modern Praguian linguistics). Only one article, by Lehmann, notes that a grammar includes the phonology with its interfaces to phonetics and orthography. This book, like Payne and Weber (2005), is concerned largely with general issues in grammar writing and not with particular areas.

In general, then, grammar writing has received good attention in recent years. The work on this topic often focuses on issues such as form and function, the empirical founda-

2 Cristofaro (2006: 138) speaks obliquely about phonology in her contribution to Ameka, Dench, and Evans (2006) in the following statement: "… it was not uncommon for grammars written until about the '80s to privilege phonology and morphology over syntax. Thus, several grammars written in that period have long and detailed sections about noun and verb structure, while the space devoted to sentence structure is comparatively limited." Cristofaro's major interest is in morphological and syntactic typology.

tions for a grammar, the cultural context of a grammar, the contributions of grammar to linguistic typology, community use of grammars, and the like. In none of these works is there a detailed reflection on the role of sound in a grammar, what the issues are, and how to work through these issues.

One might ask why this is the case. Why are there discussions of topics such as semantics and diachrony, but not of the presentation of sound? I turn to some speculation on this topic next.

3. WHY LITTLE REFLECTION ON SOUND IN A GRAMMAR? As discussed in the previous section, overall it appears that, in recent work on grammar writing, sound has received little focused attention. Why might this be the case? Is it because there is no debate about what phonology consists of? Is it because the issues around sound have not been thought through recently? Is it because morphology, syntax, semantics, and pragmatics are considered to be more important than sounds? Is it because, as Noonan (2005:312) notes, there are "generally lower standards of training of the field linguist in phonology and, in particular, phonetics"? There are probably a number of reasons and I address a few of them here, looking at the traditional goals of grammars and at issues in representing sound on the page.

3.1. THE DEFINITION OF GRAMMAR. A review of the definition of the word 'grammar' is itself instructive in understanding why issues of sound have received less attention than other areas in work on grammar writing. Definitions of the word 'grammar' tend to refer to word formation and sentence structure. In a search for definitions of 'grammar' (http://www.google.ca/search?hl=en&biw=1356&bih=728&defl=en&q=define:grammar&sa=X&ei=BcuETfegFMmY0QGulaXWCA&ved=0CBUQkAE; accessed 19 March 2011), of the definitions that are relevant, most of them are defined along the following lines: 'the branch of linguistics that deals with syntax and morphology, and sometimes with semantics', 'the logical and structural rules that govern the composition of phrases, sentences, and words'. The Wikipedia discussion of grammar includes phonetics and phonology, but in a secondary use – "In linguistics, *grammar* is the set of structural rules that govern the composition of sentences, phrases, and words in any given natural language. The term refers also to the study of such rules, and this field includes morphology, syntax, and phonology, often complemented by phonetics, semantics, and pragmatics" (http://en.wikipedia.org/wiki/Grammar, accessed 19 March 2011).

The use of the term grammar to refer to morphology and syntax is reflected in the fact that, at least some parts of the world, grammar and phonetics were considered separate areas for some time. For instance, until 1971 University College London had two departments, one of phonetics and another of general linguistics; these amalgamated in 1971 (http://www.phon.ucl.ac.uk; accessed 19 March 2011).

3.2. THE GOALS OF TRADITIONAL GRAMMARS. An important goal of traditional grammars was to describe a language in order to assist in reading that language. For instance, a classic grammar of Old English, by Mitchell and Robinson (1992), notes that the grammar is good for "those wishing to acquire a reading knowledge of the language. But potential specialists in phonology should find it a help in their preliminary studies of

the essential grammar;" obtaining reading knowledge of the language was clearly viewed as primary. In an earlier grammar of Old English, Wright and Wright (1925), the authors say in the preface "… we should strongly recommend the beginner not to work through the phonology at the outset … In fact, it is in our opinion a sheer waste of time for a student to attempt to study in detail the phonology of any language before he has acquired a good working knowledge of its vocabulary and inflexions." Beyond a chapter on orthography and pronunciation that includes information on vowels, consonants, and accentuation, the phonology in this grammar deals largely with comparative and historical issues.

With a focus on written languages and a definition of grammar that encompasses morphology and syntax, the stage was set for grammars of unwritten languages to focus on these areas, with limited attention to sound.

3.3. LINGUISTIC CHALLENGES OF UNWRITTEN LANGUAGES. Turning to languages without a history of writing, the substance of phonology presented a strong challenge for early researchers. Goddard (1996:17) writes about early contact of Europeans with languages of North America, noting the struggles that these languages presented in terms of their sounds:

> From the beginning visitors who came into contact with American Indians recorded individual words and word lists. All early recorders struggled with the problem of writing unfamiliar sounds with the imprecise alphabets of standard European languages. This problem of phonetic accuracy remained until a comprehensive scientific understanding of phonetics emerged, beginning in the last third of the nineteenth century. Before there was a general science of phonetics, students of language had no way of accurately describing and hence understanding how sounds were produced by the organs of speech, and hence even when an observer learned to recognize a new sound there was no way of defining a new phonetic symbol for it or of otherwise communicating clearly to others the nature of the sound. Thus there was little effective cumulative knowledge about the sounds used in the languages of the world.

Boas, in his introduction to the *Handbook of American Indian Languages*, reinforces the difficulties with sounds. In this 1911 publication, Boas found it necessary to define the core of phonology as consisting of articulate speech, or sounds produced with the larynx, oral cavity, tongue, lips, and nose, and he further comments that it is important to recall that languages have a definite and limited number of sounds that is never excessively large.

3.4. INTERPRETING TRANSCRIPTION. Even with the development of the science of phonetics that Goddard mentions, writing and interpreting an unfamiliar language presents challenges. One comes from the interpreting of the transcription system. While the International Phonetic Alphabet is designed to give a unique symbol to each sound found in the languages of the world, in practice, full details are often not given in grammars, and may be elusive to the listener for some time. For instance, on seeing the symbol [u], even someone trained in phonetics might not know how high and how rounded this vowel is. Similarly, the symbol [t] is often interpretable as, for instance, either a dental stop or an alveolar stop,

and some might not hear a difference between these. Thus, it is not uniformly the case in practice that a symbol is uniquely interpretable: the transcription itself is an abstraction.

3.5. HEARING SOUND – WHOSE EARS? Another reason that representing sounds is a challenge is that the understanding of sounds is filtered through the hearer's native language. This was noted at least as long ago as Boas (1911:16-17), who remarked:

> It has been maintained that this is not a characteristic found in more primitive types of languages, and, particularly, examples of American languages have often been brought forward to show that the accuracy of their pronunciation is much less than that found in the languages of the civilized world. It would seem that this view is based largely on the fact that certain sounds that occur in American languages are interpreted by observers sometimes as one European sound, sometimes as another. Thus the Pawnee language contains a sound which may be heard more or less distinctly sometimes as an *l*, sometimes an *r*, sometimes as *n*, and again as *d*, which, however, without any doubt, is throughout the same sound, although modified to a certain extent by its position in the word and by surrounding sounds. … This peculiar sound is, of course, entirely foreign to our system; but its variations are not greater than those of the English *r* in various combinations, as in broth, mother, where.

3.6. SUMMARY. The factors identified above, both linguistic and social, and undoubtedly many others, might be expected to make sound an area ripe for reflection. Yet this has not occurred. In the next section, I survey a number of grammars, largely of unwritten languages of North America, to see how the traditions around defining phonology have emerged in the past century.

4. A BRIEF SURVEY OF WHAT IS INCLUDED IN PHONOLOGY: EARLY DAYS. In order to establish what is considered essential in the presentation of sound in a grammar, I undertook a brief survey of a number of grammars, reviewing the sections called phonology, phonetics and phonology, sounds, or something similar. I selected grammars largely, but not entirely, of North American languages. This survey is cursory, and it is difficult to know if the findings would hold if a larger and broader set of grammars were examined. Nevertheless, I think that it is worthwhile to include the survey as it provides us with some notion of what has been taken to be phonology over some time period and how this has evolved.

I began with an early grammar, by Petitot (1876). This is a rather unusual grammar, including detailed information on three Athabaskan languages of northern Canada, plus scattered information on other related languages in the area. It is part of the introduction to a dictionary, and the dictionary forms the bulk of the book. The grammar includes detailed discussion of morphology, establishing paradigms and comparing the different languages. The presentation of phonology is brief: Petitot includes the alphabet that he uses and a description of how the sounds are made.

Beyond Petitot, I began the survey with Sapir's 1912 grammar of Takelma (Takelman). This is an early grammar by Sapir, written as his thesis. The grammar is divided into sections, with the discussion of phonology occupying sections 2 through 24. Sapir begins

by introducing Takelma phonology and comparing it with that of geographically close languages. He describes the vowels, comparing the pronunciation of Takelma vowels with those of English, and he examines phonological processes involving vowels, such as vowel-glide alternations, hiatus resolution, *u* dissimilation, and *ɪ* umlaut.

Sapir also discusses speech effects involving vowels. For instance, he notes that quantity depends on factors such as speech rate and placement of stress-accent, with vowels reducing in quantity when stress-accent is lost, but short vowels sometimes lengthening "when dwelt upon for rhetorical emphasis" (1912:13). In discussion of stress and pitch accent, he notes the difficulties of determining which syllable is assigned stress-accent in uninterrupted speech. He uses musical notation to show tone levels.

Sapir's discussion of consonants is likewise detailed, including pronunciation and positional constraints on consonants and consonant clusters. He also examines phonological processes such as dissimilation and epenthetic *h*.

Sapir's Takelma grammar contains the core of what continues to be required of the phonology of grammars – discussion of the sound system, with attention to phonemes, allophones, and distributional constraints, as well as discussion of prosodic characteristics and processes. Sapir sought ways to provide a visual representation of speech, both through the use of a standard transcription system and by using musical notation to indicate tones.

Haas (1940), in a grammar of the isolate language Tunica (her dissertation), provides a detailed survey of the phonology. She includes phoneme charts along with descriptions of Tunica sounds as compared with English. She distinguishes syllable types, noting the existence of both stressed and unstressed syllables, and she identifies what she calls phonomechanics, or phonological processes (vocalic contraction, assimilation, syncope).

While this is just two grammars, the information on phonology found in the Takelma and Tunica grammars forms the core template for the phonology in the grammars that I surveyed. The discussion of phonology includes a list of sounds (consonants, vowels, prosody), their pronunciation, phonotactics, and discussion of allophones and allomorphs.

In surveying later grammars, these core components remain. I looked Broadbent's 1964 grammar of Southern Sierra Miwok (Utian) and Barker's 1964 grammar of Klamath (Plateau Penutian), both published in the University of California Publications in Linguistics series. Broadbent includes discussion of consonants and their positional variants as well as vowels, addressing their distribution and variation in quality. She introduces the syllable canon and stress. She discusses intonation and juncture, provides a phonological definition of the word, and discusses morphophonemics. I found her remarks of individual variation to be of particular interest (Broadbent 1963:13): "The phone [s̲] occurred only in the speech of Chief Leeme. The alveolar variant appeared only in forms said to represent the Yosemite dialect, or when the informant was slightly inebriated. Castro Johnson, who lived in Yosemite for several years as a young man, accepted such forms as characteristic of Yosemite speech. Other informants, however, said that they did not represent Yosemite or any other Southern dialect, saying that the alveolar spirant was a Central Sierra feature. Only Chief Leeme claimed to speak the Yosemite dialect; other informants referred to their memory of the speech of undisputed Yosemite individuals, now deceased. If this variable phone was present in Southern Sierra, then, it occurred only in the Yosemite dialect, and its presence there is disputed by the informants currently available. In other dialects, it is

regularly replaced by /h/." The recognition of variation of various types is another important aspect of phonology.

Barker, in the Klamath grammar, covers much the same topics: he introduces symbols, discusses consonants and their variation, vowels, pitch, stress, and juncture. He presents what he calls anomalous phenomena, and also provides alternative analyses. Barker (1964:48) also comments on other aspects of the material he gathered, noting for instance that "Phenomena such as stuttering, swallowing, coughing, and hesitation vowels are frequent on the tapes." He recognizes voice qualifiers in the texts (1964:49) – "falsetto utterances for little cute characters, deep bass utterances for older and more respected figures, growled utterances, whispered utterances, and many other varieties" and further notes the use of "Extra vowel length for emphatic purposes is characteristic of Klamath. It is an added device for *narrative style*. It may occur with any stressed vowel and may be of any duration. It may have unusual pitch contours, such as wavering, ululating, etc." Thus a focus on phonological aspects of performance was important to Barker in addition to more narrowly construed phonological analysis.

These grammars set the stage for later grammars in terms of what is required in the phonology. They include aspects of sounds that can be recorded on paper, including contrastive sounds, allophones, morphophonemics, and prosody, often both at the word level and beyond the word. There is also discussion of variation and of different speech styles. It is interesting to note that many of the grammars of this time period form the basis for the teaching of phonology as it became known in the 1970's, with an emphasis on word-level phonology – the sound system, allophones, and morphophonemics occupied the attention of phonologists in this time period, with less attention to phonology above the level of the word.

5. AN ASIDE: 'BEST PRACTICES' GUIDELINES FOR GRAMMAR WRITING. Perhaps partly due to the activity around grammar writing over the previous decades, Comrie and Smith (1977:5), in introducing the Routledge Descriptive Grammar Series, aim to provide a standard framework for the series to serve as "catalyst in the elicitation of all information that could be of interest for theoretical work …". They note that such a framework is useful, but should not be interpreted as a straightjacket.

With respect to phonology, Comrie and Smith (1977:9) write:

> In the section of phonology all examples should be accompanied by the relevant phonemic or phonetic transcription (in, respectively, obliques and square brackets) in terms of the IPA phonetic alphabet. In sections not dealing specifically with phonetic detail it may be possible to use an adaptation of the IPA system for typographic convenience (for instance, by using š rather than ʃ, …). Any departures from the IPA system should, however, be made quite explicit and cleared with the editors in advance.

Comrie and Smith place phonology as the third section of a grammar, after syntax and morphology and followed by lexicon and basic vocabulary. [This has not been very well accepted in the practice of grammar writing; the chapters on phonology generally come

before those on morphology and syntax, although phonology is often also addressed in the presentation of morphology and discussion of phonology in discourse may follow.] In the section on phonology, they call for glossing and using IPA symbols, and they provide a list of descriptive articulatory features to use with respect to place of articulation, manner of articulation, laryngeal features, and so on.

In summary, Comrie and Smith (1977:58-65) propose that the following phonological information be included in the grammar:

- Sections on phonological units (segmental), including allophony, phonetic realization, restrictions with respect to word classes and phonotactics.
- Discussion of phonotactics, including positional restrictions, sequence restrictions (both adjacent and long distance), syllable shape and restrictions, and word class restrictions.
- Discussion of suprasegmental phonology, including length, stress, pitch, intonation, with discussion of distribution, tactics, processes, etc.
- Presentation of morphophonology, both segmental and suprasegmental.
    - Segmental: assimilation, dissimilation, other alternations, metathesis, coalescence, deletion, insertion, reduplication
    - Suprasegmental: changes in stress and tone under morphological processes

Most of the presentation on phonology is contained in this section, with a few references to phonology in sections of the outline on morphology and syntax.

Another grammar guideline, this one from the 1990's, is for the short-lived Cambridge University Press Grammar Series that was edited by Dixon and Rice. In terms of phonology, these guidelines included the following.

- Consonant and vowel phonemes in tabular array, with description of phonetic realizations including allophones and environments and dialect differences; IPA unless a good reason
- Labels for tables, details
- Explicit information on phonotactics, stress, tone, segmental features functioning prosodically, etc.
- Intonation marking commands, polar questions, content questions, etc.
- Criteria for defining word (phonological, grammatical)

These two sets of guidelines are similar, reflecting what we have seen in the grammars reviewed.

6. A RETURN TO THE SURVEY: A FEW MORE RECENT GRAMMARS. The more recent grammars develop the foundations laid out in the earlier grammars. The major changes in grammars come because of both technological and theoretical developments. In terms of technology, it has become increasingly possible to do phonetic analysis. This allows not only for more careful work on phonetics, but also for better work on phonology above the level of the word. Linguistic theory has also developed, paying more careful attention to

the relationship between phonetics and phonology, to phonology above the level of the word and to language variation, among other topics.

Here I look at grammars of languages spoken in parts of the world other than North America. Chelliah (1997), in her study of Meithei (Tibeto-Burman), discusses the standard phonological topics (consonants and vowels with their distribution and variation, syllable structure, tone, lexical rules, post-lexical rules). In addition, Chelliah includes pitch tracks in order to compare vowels of different tones. Sapir, as noted earlier, used musical scores to show tones in the Takelma grammar, so the need for a representation of tone has long been recognized, but the technological developments of recent years make this easier than it had been in the past.

Aikhenvald (2003), in a grammar of Tariana (Arawak), includes an extensive section on phonology: segmental phonology, syllable structure, stress, the nature of the phonological word and evidence for it, phonological processes, prosodic classes of morphemes, pause marking, phonological phrase, and intonational phrase. Aikhenvald clearly goes beyond the word level in looking at phrasing. It is interesting that she has incorporated phonology of higher structural levels, but in her detailed discussion of discourse organization, she gives rich information about sentence-linking, among other topics, but does not discuss phonological issues relating to discourse. Aikhenvald's careful attention to the different types of words perhaps reflects discussion on this topic in the theoretical literature. Most striking about the presentation of phonology in the Tariana grammar is the discussion of phrasing beyond the level of the word.

Dixon (2004) is an award-winning grammar of Jarawara of Southern Amazonia (Arawá). The contents of the sections on phonology are by now familiar – vowels, consonants, historical development, phonotactics, loans, stress, grammatical and phonological word, phonological rules. It is interesting to note that this grammar was awarded the Bloomfield Book Award by the Linguistic Society of America in 2006, with the following citation.

> R. M. W. Dixon's *The Jarawara Language of Southern Amazonia*, written with the assistance of Alan R. Vogel, is an invaluable record of a language in serious danger of extinction. The complexities of the language are unraveled with a clarity and insight that allow the reader to share in what the author describes as 'the intellectual pleasure of working out such a magnificent system'. (http://www.linguisticsociety.org/content/leonard-bloomfield-book-award-previous-holders)

While definitely worthy of this award, the phonology section is presented in great depth but is at the same time quite traditional in nature.

Genetti's 2007 grammar of Dolokha Newar (Tibeto-Burman) is the most recent grammar that I examined. This is another award-winning grammar, receiving the inaugural Gabelentz award from the Association for Linguistic Typology in 2010. The grammar includes the standard: consonants, vowels, processes, phonotactics, syllable structure, word structure, stress. In addition, it contains detailed information about prosody, with discussion of intonational units, phrasal accents, terminal pitch contours, and units about the level of the word. Genetti (2007:89) notes that "… prosody is one of the central systems by which speakers parse and organize connected speech. It is used both to break the speech

into manageable chunks (intonation units) which are easily processed cognitively. It is also used to highlight and background particular units, and particular words within those units. And, crucially, it is used as a 'signpost' which provides cues to the hearer about the relationships between units, as well as whether or not the material constitutes embedded direct quotation. However, the signpost function does more than simply provide cues to the hearer. It also allows for higher level prosodic structuring, as speakers use transitional continuity to combine single intonation units into structured groups. … There is one other important function of prosody which I am not able to address, that of conveying affect, or the emotional state or attitude of the speaker."

Genetti (2007:485) provides detailed discussion of the relationship between prosodic and syntactic structuring: "It is at the sentence level that one can witness the interaction of the clause-combining strategies … and the genius of the design principles that form the basis of the grammar. … the syntactic structuring … gives a partial view of how speakers are segmenting the speech stream … and relating those units …. Simultaneous to the syntactic structuring of speech is the prosodic structuring of speech. .… examining the interaction of the syntactic and prosodic levels allows us greater insight into how speakers simultaneously utilize these distinct domains in the formation of sentences and the construction of narrative." She presents diagrams to indicate prosodic phrasing; an example is given below.

… ∧	daNga	par-ai	ju-	ju-eni	"lo	ba#!bu.	∧
	astonishment	feel-BV	be-FS	be-PART	EXCL	baby	
… ∧	thijin	u	anaut3ha#		kha#	khoN-gu. __	
	1pINC.ERG	this	strange		matter	see-1pPST	

He felt astonished: "Lo baby! We saw a strange thing…" 491

This grammar thus integrates aspects of sounds fully, both contextualizing the importance of phrasing and making the reader broadly aware of its importance not only at the word level but at higher levels as well.

7. INTERIM SUMMARY. All of the grammars that I reviewed include something that we can call phonology. When we move away from the grammars based on written languages (Old English) to those on languages without a written tradition, discussion of sounds is present in some form or another. The earliest grammar that I surveyed, Petitot, presents the system of sounds. By Sapir, sounds had come to include not just segments but also prosody, and the topic of variation in both particular sounds and discourse context became important. More recently, sound at a level larger than the word has been discussed in more detail.

As noted earlier, changes in the treatment of sounds in grammars likely reflects different developments within the field. On the technological front, linguists have been keen to record from the moment this became possible, and, with the development of the ability to do acoustic analysis, at least some linguists have included acoustic representations of sounds in grammars. With programs such as Praat, many grammars now include some spectrograms and pitch tracks. The better understanding of variation brought about through sociolinguistic work has allowed for a deeper study of variation. Phonological

theory has allowed for different ways of talking about sounds and for recognition of the role of sound throughout the grammar. At the same time, grammars have changed phonological theory, with aspects of language hitherto unobserved accommodated in the theory. The increased work on typology makes linguists aware that appropriate data on a wide variety of languages is required to answer important questions. Evans and Dench (2006:16) note that in semantic fieldwork "Recent advances … have begun to give us better techniques for tackling these problems" (production of good meaning-based grammars). Just as with semantics, both the scope and methods of phonology have evolved.

It is worthwhile to close this section on the increasing recognition of the importance of sound with a quote from Dixon (1994:299) (quoted from Mosel 2006:63):

> The most important point is that a language can only profitably be studied as a whole. One must recognize and distinguish different levels of structural organization – phonological, morphological, syntactic, semantic, discourse and pragmatic – but each of these continuously interrelates with the others.

Phonology is definitely a level that interrelates continuously with all other levels, and to study the other areas without reference to sound has become increasingly unacceptable as the methods have allowed for this study.

I ask next if we have reached the point that we can say we know what the phonology must include, or are there still strides to be taken.

8. BEYOND THE CURRENT PARADIGM. Do we stop here, saying that we are satisfied with what is represented in a typical grammar in terms of sound? The heart of the study of phonetics and phonology is about sounds, sound systems, pronunciation, interaction of sounds, variation in sounds, and patterning of sounds at all levels from the morpheme to the word to the phrase to discourse. We write about sounds, and represent sounds through symbols on the page, but, with rare exception, we do not represent sounds themselves in a grammar, only approximations through transcription and acoustic representation. Today we have the tools to represent sound more directly, through recordings of the sounds themselves.

Before turning to sound itself, it is worthwhile to review briefly the value of acoustic representations of sounds. One way of representing sounds more directly than transcription is through the use of spectrograms, pitch tracks, and the like. This in itself is very useful: it gives an accurate picture of a sound, helping to deal with the issues of perception noted earlier as well as with issues of reliability and accountability. Acoustic representations require a depth of knowledge to interpret, and they remain a representation of sound rather than sound itself.

Why might a more direct representation of sound in a grammar be of value? I would like to look at this from two perspectives, first the perspective of the linguist and second the perspective of the community of speakers. I begin with the linguist. However, before turning to the value of including sound in a grammar, an important caveat is in order. There are individuals and communities who are happy to work with a linguist, but who do not want their recordings made publically available. Whatever the merits of including sound in a grammar, these are overridden by these ethical issues.

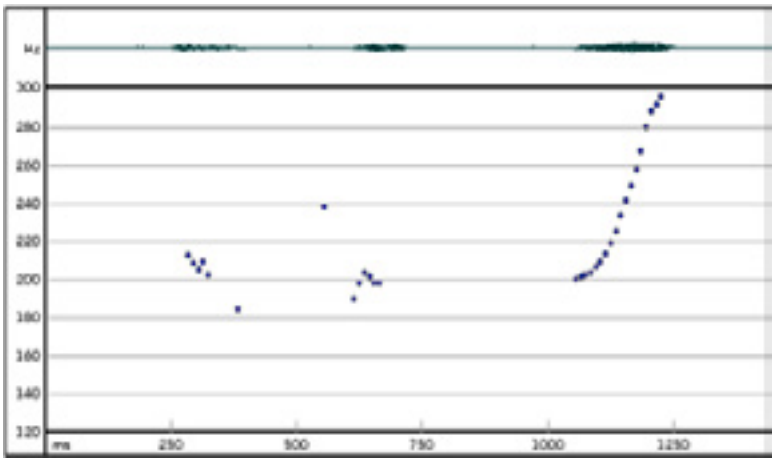
8.1 SOUND AND THE LINGUIST. As discussed in section 2, using words to describe a sound does not necessarily call up the identical sound to all readers. Recordings themselves would allow the reader to hear the sounds directly.

In many languages, there are many sounds that are difficult for a non-native speaker of a language to distinguish. Suppose that two morphemes are distinguished solely by two very similar sounds. If these sounds are conflated by the linguist, not only is the sound system itself misrepresented, but there are potential implications for the morphology as well, with possible misanalysis of two or more morphemes as one. (An alternative analytic problem can arise, with misanalysis of one morpheme as two if allophones are not recognized as such; recordings are not particularly helpful in sorting this out as it is an issue related to analysis rather than to form.) If sound were available, it might be possible to correct such a misanalysis.

Some sounds are particularly difficult to deal with. Tone is a notoriously challenging area, as are other aspects of prosody. Pitch tracks are of enormous value in seeing what tones look like, but they cannot tell most people just what they sound like. See Remijsen (2011) for discussion.

In transcribing, we tend to come to an analysis of what is phonemic and what is allophonic and then adopt a phonemic transcription system, with remarks on allophones and other variation in the section on phonology. While this is an appropriate analytic strategy, there are circumstances under which important information might be lost. As an example, in many dialects of Dene (Slavey; Athabaskan), the palatal glide [j] and the voiced alveopalatal fricative [ʒ] appear to be in free variation in some environments. In the grammar of Slave (Rice 1989), I comment on the variation; in texts in the grammar, I level the variation between these sounds. However, I have a suspicion that remains untested that there is a contextual difference involved in choosing one or the other of these sounds, with the fricative occurring when something is new information and the glide otherwise. It is not possible to determine whether this suspicion is supported based on the transcriptions (or whether other factors might be involved in the variation), as the difference was leveled out; it would be possible to study this systematically if oral texts were part of the grammar. Such situations are relatively common: variation is noted, but not transcribed beyond the discussion about variation. It is then not possible later on to follow up on the variation to see if there are any linguistic factors that might control it.

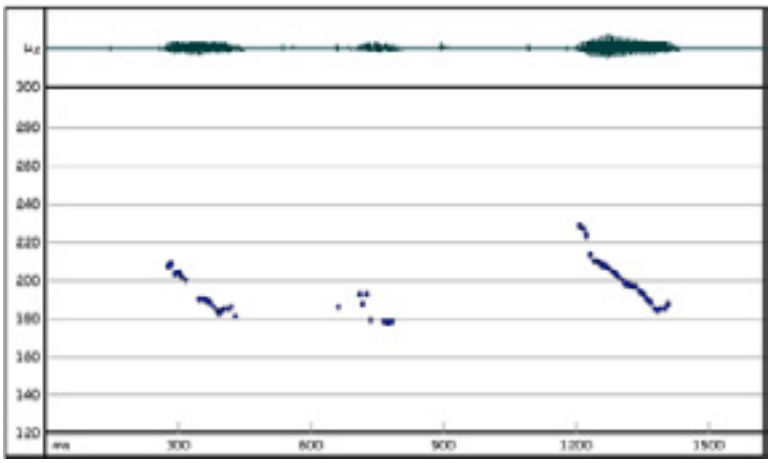
The study of sound above the level of the word is also difficult to represent on the page, and it remains relatively unusual to find good discussions of sound at this level beyond intonation and some sandhi phenomena. The study of sounds above the word level would be greatly enhanced if sound were available. This is partially addressed through spectrograms. For instance, consider the three pitch tracks below, from Holton (2005), an article on Tanacross Athabaskan. These show the pitch contours for different phrase types – a yes/no question, a declarative, and a content question. While the differences between them are clear, and the inclusion of pitch tracks in a grammar is of great value, just how they translate to speech is not necessarily easily determined.



	¬u>g	ièh	-t'e!>T
lexical tone	L		H
intonation			H* H%

"Are you frying the fish?"

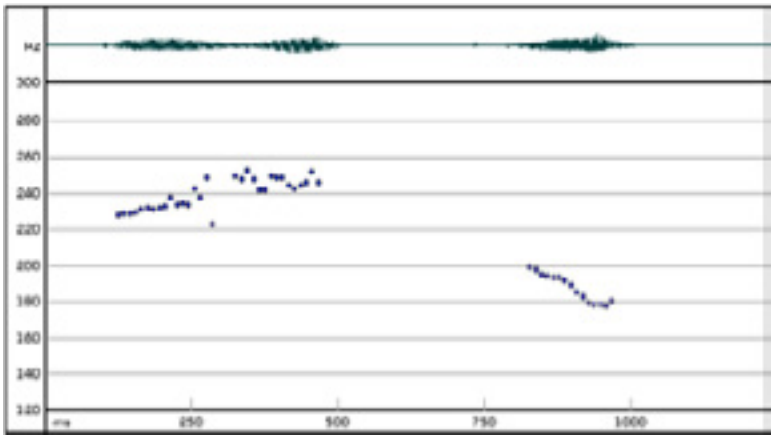
FIGURE 2: Pitch track for yes/no interrogative with high tone stem



	¬uÚ>g	Ek	t'e!>T
lexical tone	L		H
intonation			H* L%

"I'm frying the fish"

FIGURE 5: Pitch track for declarative with high tone stem



	diÚ>	i@7h	t'e!>T
lexical tone	H		H
intonation			H+L* L%

"What are you frying?"

FIGURE 7: Pitch track for content interrogative with high tone stem

The spectrograms are very valuable, but it is probably the rare person who can actually 'hear' a spectrogram.

Noonan (2005:354) speaks as a linguist about the standards that are required of a grammar. From the perspective of language loss, he talks of the responsibility of the grammar writer to set their standards high: "… we should be aware that when we are writing grammars of those languages which will likely be moribund or extinct by the end of the century – that is, the great majority of the world's languages – that we are writing for the ages. So, we must make sure that what we are doing reaches for a very high standard." This high standard includes the points reviewed in this chapter and others: systematic description of the sounds of the language, their pronunciation, allophony, distribution, variation; relationship to orthography; phonological processes not recognized in the orthography; patterning of sounds with respect to morphology, syntax, discourse; variation; prosody at the word level and above; how sound interacts with information structure; instrumental accompaniment. And, finally, Noonan (2005:365) adds "Where practical, audio and video recordings should be made of various language genres."

Thus, from the perspective of the linguist, the inclusion of sound would augment and support phonological analysis in many ways, as well as providing for work that meets the highest of standards.

8.2. SOUND AND THE COMMUNITY. Sound is also important from a second perspective. As discussed in section 1.2, whoever the audience for a grammar is perceived to be when the grammar is written, that audience today often includes users who were not necessarily expected to be interested in a grammar, namely members of a community who are interested in language revitalization and reclamation. Parsons-Yazzie and Speas (2007:17), writing about Navajo in a grammar for Navajo speakers and learners, stress the importance

of pronunciation, saying "It is vital that you realize pronunciation is extremely important to the Navajo language." They add: "It is virtually impossible to learn a language by reading and memorizing material in a textbook. You must use the language to communicate! Practice with your classmates, but also seek out fluent speakers of Navajo and talk with them .… We know that you will succeed if you persist."

Rice and Saxon (2002:130), writing about representing variation in a dictionary, say "… invaluable information would be lost from a story's telling if the pronunciation variants that the storyteller used were washed over by means of standardized spellings. In the Western tradition, on the contrary, the written text is taken as primary and authoritative in almost all contexts."

Transcription of a language, while important, is nevertheless an abstract representation on the page, not fully interpretable, with something generally lost in the translation from sound to paper. One might even say that transcriptions, while serving a very important function, take some life out of the language. Acoustic representations offer an improvement, but are still an abstraction. While sound accompaniment to a grammar was difficult in the past, current technology has made it reasonable, assuming that ethical conditions are met.

9. WHAT TO INCLUDE? For the graduate student seeking to write a grammar of a previously undescribed language for their dissertation, the demands of what the ideal section on sounds would include could well seem unapproachable. In this section, I briefly review some of the topics that I think must be included, and then raise a few specific questions about sounds that the grammar writer will likely need to think about. This section does not go beyond previous sections in what belongs in the phonology section of a grammar. I hope that the survey presented in section 4 points to the need to discuss inventories, phonotactics, allophony, morphophonemics, and phrase-level phonology, and that some acoustic material and sound would greatly enhance a grammar. To me, there is no substitute for reading grammars to help decide what should be in a grammar, and sounds are no exception to this. For what is considered appropriate in terms of sound today, a good starting point might be the recent PhD theses that have won awards as outstanding grammars from the Association for Linguistic Typology (http://www.linguistic-typology.org/awards.html; accessed 4 January 2012) and the Society for the Study of Indigenous Languages of the Americas Mary Haas award (http://www.ssila.org/; accessed 4 January 2012). There are also a number of questionnaires that might be of value in thinking about what is needed in phonology; several are available at http://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaires.php (accessed 4 January 2012). With the questionnaires, as with the grammars, the user must exercise their own judgment as to what is appropriate for their particular circumstances and for the language under study.

What *must* be included in a section on sounds? Culled from the survey and guidelines by Comrie and Smith, Dixon and Rice, and Noonan, as well as Bowern's textbook on phonology (2007: 70-71), I offer the following.

- Presentation of segmental inventories, together with articulatory descriptions and discussion of allophones and variation, with careful exemplification.
- Presentation of phonotactics and syllable structure, with careful exemplification, including discussion of any morphological factors that affect the distribution

of sounds (e.g., there might be a richer inventory in stems than in affixes). In discussion of phonotactics, it is important to talk about positional inventories, including any differences in inventories that might exist depending on prosodic position.

- Presentation of suprasegmentals, including tone, stress, and intonation, with careful exemplification, augmented with pitch tracks.
- Presentation of phonological rules, with careful exemplification and motivation.
- Phonological analyses are not always as clean as one might like – for instance, it is often difficult to decide on what is phonemic and what is not in a language and there might be sounds that are of very limited occurrence– and these types of complexities should be addressed.

While the above topics largely concern word-level phonology, at least the basics of phonology above the word should be included in a grammar. This could include segmental effects such as sandhi, and suprasegmental effects – groupings of words into phrases and intonation in different sentence types are two important topics. Again, pitch tracks will be extremely helpful here.

The above is a very broad sweep, and there are many particular questions to consider. I pose some here, with brief discussion. As noted above, there is no substitute for reading grammars to come to a sense of how others have addressed these questions, and others.

How much articulatory detail is required in descriptions of sounds? This probably depends on the sound. For coronal sounds in particular, it is probably worthwhile to be as explicit as possible about how the sound is made as there is considerable variation cross-linguistically. For instance, as noted earlier, the symbol *t* is used to represent a stop at either a dental or an alveolar place of articulation, so it is important to give details about what the place of articulation is (assuming that one can be determined). Rhotics should be described in as much detail as possible, as the symbol *r* is used in many different ways. Laryngeal features of stops and affricates should be spelled out – taking *t* as an example again, this symbol is used to represent the expected unaspirated stop but, in many cases, it is the symbol used to represent an aspirated stop, especially when there is not a phonemic unaspirated stop in the system. These descriptions should be given in articulatory terms. There is one type of description to avoid. It is very tempting to make a statement that sound *x* is like a sound in some other language, as, for instance, Sapir did in the grammar of Takelma. These kinds of statements are frustrating for the user – the user might not know that other language and, even if they do, they might not know the dialect that the person is using. IPA provides a kind of standard that, at least for linguists, should give a reasonable idea of what the sound is like. For community users though, IPA can present a challenge. Thus, for both academic and community users of a grammar, there is nothing like sound files!

Related to questions of description of sounds are questions of use of orthography. Some linguists are insistent that examples in a grammar must be written using IPA. If there is an accepted orthography for a language that is fairly phonemic in nature, I myself see nothing wrong with using that system, with careful note of the relationship between the orthographic symbols and IPA, and reminders of relationships as appropriate, through, perhaps, the use of both orthography and IPA at relevant points in the grammar. When there

is variation in pronunciation, it needs to be thought through carefully how to present this, as the orthography may well reduce that variation in the interest of a standard. Dictionaries often present both orthography and pronunciation, and such a system could be used.

Also related to the degree of detail in a description of sound is the use of acoustic material. What might it be helpful to include? At the word level it might be useful to present information about voice onset time in stops as languages vary considerably in this, information about duration of consonants (especially about consonants such as labiovelars compared with labials or velars and about geminates as compared to singletons), information about the duration of vowels (short vs. long vowels, lax vs. tense vowels; phonologically long vowels as opposed to phonetically long vowels; epenthetic vowels as compared to underlying vowels of the same quality), and information about tones. Scatterplots showing the range of variation in vowels in a particular environment can also be useful to the reader in understanding the range of variation. At the phrasal level, pitch tracks can be extremely helpful is representing intonation patterns.

There are many other sorts of questions to grapple with. How many examples should be included? There should be sufficient examples to show the sound contrasts that are found, the positions in which the sounds are contrastive, allophony, and variation. With variation it is valuable to identify whether the variation is found across speakers, or whether within speaker variation is present as well, and, if the information is available, discussion of how common the variants are is of value. In exemplifying processes, I think it is important to provide as full data sets as possible. For instance, in a field methods class one year, we studied a language where, in vowel-vowel sequences, one of the vowels deleted. It was important to find data to illustrate all possible vowel-vowel combinations in order to see what happened to each one; a statement with just a few examples was not, we agreed, appropriate as the reader would not know if we had actually found the data to test each one. In such a case, if there were sequences that were absent for some reason, it would be important to comment on that as well. We also found variation in some cases in how a particular sequence was resolved; this too requires comment.

Another important question to consider is the type of formalism use. The goal is to be as clear as possible. If the formalism increases the clarity, it is appropriate. However, formalism for the sake of formalism is not such a good idea – a grammar is meant to be a contribution that lasts over time, and formalism tends to be much more transitory. There are some cases of formalism that I find very useful. Valentine (2001), in his grammar of Nishnabemowin, an Algonquian language, shows how vowel deletion works through the use of metrical trees in a way that is clear and illuminates the process. Complex rules, on the other hand, do not generally provide insight, and a clear description, together with comprehensive data, usually is more helpful.

Again, there is no substitute to studying grammars to determine what the content of a grammar should be with respect to sound systems. The description should provide the reader with the core information about what it is that the language under discussion is all

about. This, supplemented with sound files, should bring the language alive in the minds, and ears, of the reader.[3]

Why so much, the person interested in morphology and syntax might ask. Isn't it sufficient to give sound charts and examples of contrasts and their distribution and to discuss phonotactics, syllable structure, and introduce morphophonemic processes? The problem with leaving the phonology sketchy in this way is that, as the quote from Dixon (1994: 229) given earlier makes clear, the components of language interact, and without firm grounding in the sounds of the language, the language is reduced to language on the page, not language in the real world.

10. SUMMARY AND CONCLUSIONS. The role of sound in grammar writing has evolved as the field has developed: grammars increasingly attend to variation, to phonology above the level of the word, and to phonetic detail. Acoustic representations provide further information than transcription in understanding details of pronunciation. Yet the very representation of a system of sounds on the page is problematic in being an abstraction rather than the sounds themselves. From the perspective of linguists, the use of these representations, as invaluable as they are, raises issues around verifiability, accountability, and scientific rigor, as linguists have long been aware; the inclusion of sound in addition to transcription and acoustic material helps to address these issues. From the perspective of a community, there are issues of abstractness, and a lack of a kind of reality as the language is transferred to the page.

Enhancing the presentation of sound, both by describing the role of sound at all levels and by making sound available, will allow a grammar to better meet the needs of a linguist, leading to higher quality description. It will allow for better studies of areas such as phonetic typology and the role of prosody in information structure. Enhancing the phonology will also better meet the needs of the speaker/heritage learner, with the language becoming 'real' through the inclusion of sound, just as it becomes real through the use of real examples, drawn from texts, conversations, and other natural speech. Sounds can include examples of different syllable shapes, examples of sounds contrasted with other sounds, examples of sounds in context, examples of connected speech, examples of different speakers. Coupled with time-aligned transcription/orthography, and video when feasible, a grammar would present a richness that is unprecedented.

I have advocated that, in addition to the usual information included about phonology in a grammar– phonological inventory and realization, with careful description; phonotactics; allomorphy; extended to levels beyond the word, including segmental and prosodic properties, and so on; there be an extension to include sound. Such a grammar would be of both scientific merit and human value.

3 Talk of sound files raises what can be a complex question. While sound files without any background noise might be the ideal in some ways, in reality, it is often very difficult to make recordings without a rooster crowing, a dog barking, the radio playing, a baby crying, rain on the tin roof, and so on. The sound is valuable even if the conditions are not ideal.

References

Aikhenvald, Alexandra. 2003. A grammar of Tariana. *Cambridge Grammatical Descriptions*. Cambridge: Cambridge University Press.

Barker, M.A.R. 1964. *Klamath grammar*. Berkeley: University of California Publications in Linguistics 32.

Boas, Franz. 1911. *Introduction. Handbook of American Indian Languages*. Vol. 1. Bureau of American Ethnology, Bulletin 40. Washington: Government Print Office, Smithsonian Institution, Bureau of American Ethnology.

Bowern, Claire. 2007. *Linguistic fieldwork: a practical guide*. New York: Palgrave Macmillan.

Broadbent, Sylvia M. 1964. *Southern Sierra Miwok language*. Berkeley: University of California Publications in Linguistics 38.

Chelliah, Shobbhana L. 1997. *A grammar of Meithei*. Berlin: Mouton de Gruyter.

Comrie, Bernard & Norval Smith. 1977. *Routledge Descriptive Grammar Series*. Lingua Descriptive Studies: Questionnaire. Lingua 42. 1-71.

Cristofaro, Sonia. 2006. The organization of reference grammars: A typologist user's point of view. In Felix K. Ameka, Alan Dench, & Nicholas Evans (editors). *Catching language. The standing challenge of grammar writing*. 137-170. Mouton de Gruyter.

Dixon, R.M.W. 2004. *The Jarawara language of Southern Amazonia*. Oxford: Oxford University Press.

Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Genetti, Carol. 2007. *A grammar of Dolakha Newar*. Berlin: Mouton de Gruyter

Goddard, Ives. 1996. The description of native languages of North America before Boas. In Ives Goddard (editor). *Handbook of North American Indians*. Volume 17. 17-42. Languages. Washington: Smithsonian Institution.

Graustein, Gottfried & Gerhard Leitner. 1989. *Reference grammars and modern linguistic theory*. Tübingen: Max Niemeyer Verlag.

Lehmann, Christian. 1989. Language description and general comparative grammar. In Gottfried Graustein & Gerhard Leitner (editors). *Reference grammars and modern linguistic theory*. 133-162. Tübingen: Max Niemeyer Verlag

Haas, Mary. 1940. Tunica. Extract from *Handbook of American Indian languages* 4. New York: J.J. Augustin. 1-143.

Mitchell, Bruce & Fred C. Robinson. 1992. *A guide to Old English*. Fifth edition. Blackwell.

Mithun, Marianne. 2001. Who shapes the record: The speaker and the linguist. In Paul Newman & Martha Ratliff (editors). *Linguistic Fieldwork: Essays on the Practice of Empirical Linguistic Research*. 34-54. Cambridge: Cambridge University Press.

Mithun, Marianne. 2005. Grammars and the community. In Thomas Payne & David Weber (guest editors). *Studies in Language* 30.2. 281-306.

Mosel, Ulrike. 2006. Grammatography: The art and craft of writing grammars. In Felix K. Ameka, Alan Dench, & Nicholas Evans (editors). *Catching language. The standing challenge of grammar writing*. 41-68. Berlin: Mouton de Gruyter.

Evans, Nicholas & Alan Dench. 2006. Introduction. In Felix K. Ameka, Alan Dench, & Nicholas Evans (editors). *Catching language: the standing challenge of grammar writing*. 1-39. Berlin: Mouton de Gruyter.

Noonan, Michael. 2005. Grammar writing for a grammar reading audience. In Thomas Payne & David Weber (guest editors) *Studies in Language* 30.2. 351-365.

Parsons Yazzie, Evangeline & Margaret Speas. 2007. *Diné Bizeed bínáhoo'aah/Rediscovering the Navajo Language. An introduction to the Navajo Language*. Flagstaff, Arizona: Salina Bookshelf Inc.

Petitot, Emile. 1876. F. S. 1876. *Dictionnaire de Langue Dene-Dindjie, dialectes Montagnais ou Chippewyan, Peaux de Lievre et Loucheux, etc*. Paris: E. Leroux. 367 pp

Remijsen, Bert. 2011. The study of sound stands to gain from a greater use of sound data in training and in publishing. Talk presented at the Berkeley Tone Workshop. February 2011.

Rice, Keren. 1989. *Grammar of Slave*. Berlin: Mouton de Gruyter.

Rice, Keren & Leslie Saxon. 2002. Issues of standardization and community in Aboriginal language lexicography. In William Frawley, Kenneth C. Hill, & Pamela Munro (editors) *Making dictionaries. Preserving Indigenous languages of the Americas*. 125-154. Berkeley: University of California Press.

Sapir, Edward. 1912. The Takelma language of southwestern Oregon. Handbook of American Indian Languages. Bulletin 40. Bureau of American Ethnology. pp. 1-296. Reprinted in Golla, Victor (editor). 1990. *The collected works of Edward Sapir VIII: Takelma texts and grammar*. Berlin: Mouton de Gruyter.

Valentine, J. Randolph. 2001. *Nishaabemwin reference grammar*. University of Toronto: University of Toronto Press.

Wright, Joseph & Elizabeth Mary Wright. 1925. *Old English grammar*. Third edition. London: Oxford University Press.

rice@chass.utoronto.ca

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 91-108
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4586

6

Toward a balanced grammatical description

Thomas E. Payne
University of Oregon and SIL International

The writer of a grammatical description attempts to accomplish many goals in one complex document. Some of these goals seem to conflict with one another, thus causing tension, discouragement and paralysis for many descriptive linguists. For example, all grammar writers want their work to speak clearly to general linguists and to specialists in their language area tradition. Yet a grammar that addresses universal issues, may not be detailed enough for specialists; while a highly detailed description written in a specialized areal framework may be incomprehensible to those outside of a particular tradition. In the present chapter, I describe four tensions that grammar writers often face, and provide concrete suggestions on how to balance these tensions effectively and creatively. These tensions are:

- Comprehensiveness vs. usefulness.
- Technical accuracy vs. understandability.
- Universality vs. specificity.
- A 'form-driven' vs. a 'function-driven' approach.

By drawing attention to these potential conflicts, I hope to help free junior linguists from the unrealistic expectation that their work must fully accomplish all of the ideals that motivate the complex task of describing the grammar of a language. The goal of a description grammar is to produce an esthetically pleasing, intellectually stimulating, and genuinely informative piece of work.

1. INTRODUCTION. Scholars who attempt to write linguistic grammars of underdocumented languages strive to accomplish several worthy goals, many of which seem to conflict with one another.[1] Tension, discouragement and paralysis may arise as the author of a grammar attempts to fulfill unrealistic expectations of what a grammar 'should' be. The larger purpose of the present chapter is to expose some of these conflicts, and thus in some measure to free novice linguists from the unrealistic assumption that their work must fully accomplish all of the ideals that motivate the complex task of describing the grammar of a language. Among the many tensions faced by grammar writers, the following stand out as being particularly perplexing:

a. Comprehensiveness vs. usefulness.
b. Technical accuracy vs. understandability.
c. Universality vs. specificity.
d. A 'form-driven' vs. a 'function-driven' approach.

1 By 'linguistic grammar' I mean a grammatical description of a language based on principles of the science of linguistics. A major defining characteristic of such a grammar is that assertions about grammatical patterns are based on empirical evidence rather than on authority or tradition. Most linguistic grammars are reference grammars, but other types of grammars, e.g., pedagogical grammars or school grammars, may also be based on principles of the discipline of linguistics.

ISBN: 978-0-9856211-4-8

The descriptive linguist must balance all of these tensions (and more!) in an esthetically pleasing, intellectually stimulating, and genuinely informative package. For this reason, writing a linguistic grammar is definitely an art as well as a science.

2. WHAT IS A GRAMMATICAL DESCRIPTION ANYWAY? Before looking at the specific tensions that writers of grammatical descriptions face, I would like to spend a few paragraphs considering the question of what it is we are creating when we write a linguistic grammar. I believe it is important to always keep the larger view in mind as we attempt any large task, such as writing a grammar.

2.1 A GRAMMATICAL DESCRIPTION IS A COMMUNICATIVE ACT. Sometimes grammar writers tend to forget that a written grammar is an act of communication (see Payne 2007 for elaboration of this idea). The writer has important knowledge to share with a particular, interested audience. The description will succeed or fail to the extent that it communicates that knowledge in a way that the intended audience is able to appreciate and incorporate into their own cognitive frameworks.

Like any communicative act, a grammatical description has several characteristics. First, all communicative acts have a 'speaker'. The speaker in the case of a linguistic grammar is the descriptive linguist. Each linguist has particular interests, goals, personality and background, all of which contribute to the unique characteristics of the grammar. For example, some linguists are very analytic and detail oriented, while others are more 'global' in their approach to life in general, and grammatical description in particular. Detail-oriented individuals may relish the nuances of phonological and morphophonemic variation, but find syntax and pragmatics overwhelming to deal with at a level of detail that they are comfortable with. More global thinkers, on the other hand, may delight in making grand claims about discourse, syntax and information flow, but have little patience with allophonic variation. Grammars written by different writers will reflect these kinds of individual predilections. There is no one way of describing the grammar of a language, just as there is no one way to accomplish any speech act, such as apologizing, sermonizing, encouraging or proposing marriage.

Second, a grammatical description is simultaneously underinformative and overinformative. The terms 'impoverished' and 'exuberant' respectively were used by Alton 'Pete' Becker (see, e.g., Becker 1979) to describe these characteristics of communicative acts in general. A text itself cannot explicitly 'encode' all the information necessary for it to be understandable. Much important information is necessarily left implicit, to be inferred by the audience (see also Grice 1975, Sperber & Wilson 1995 and the other literature on the ostensive and inferential nature of human communication). At the same time, particularly salient information must be highlighted in special ways, and referred to multiple times throughout the text to keep it from simply blending into the background. This 'texture' of highlighting new, important or asserted information, and downplaying old, background, and presupposed information is a feature of all successful communicative acts, and should be a feature of linguistic grammars as well.

Finally, a grammatical description takes place in a context. Any communicative act makes assumptions about the audience, and attempts to engage them 'where they are'. Pragmatics is the study of language use in context. Many of the important concepts in

pragmatics refer to assumptions that people make about their interlocutors when engaging in communication, e.g., what the audience already knows ('given' information), what they are thinking about at the moment ('activated' information) and what is important to them ('newsworthy' information). The audience for a linguistic grammar is a community of linguists (often a dissertation committee) who are intensely interested in the details of individual languages, but who may not know much about the particular language being described. For this reason, background assumptions about general linguistic concepts can be assumed or treated lightly—the text of an academic grammar need not provide a basic course in linguistics. However, it must explicitly and carefully highlight the data and perspectives that the author wishes to add to the general body of knowledge. Of course there are many kinds of grammatical descriptions, including school grammars, pedagogical grammars and others. The focus of the present chapter is on reference grammars. The context for this type of grammar is usually a particular areal or language family tradition that the grammar writer is a part of, often in a graduate degree program. Other types of grammars will incorporate different assumptions about the backgrounds, interests, and states of mind of their readers.

All too often grammar writers tend to forget that a grammar is a communicative act. Linguists, of all people, should be aware of the properties of communicative acts in general, and should be able to apply this awareness to their own work whenever appropriate, as described above. For some reason, however, grammar writers often view their work as a schematic diagram of the categories and patterns that constitute the internal (subconscious) grammar of a language, or as a logical machine that 'generates' or 'sanctions' grammatical structures. While schematic diagrams and logical machines can be useful for certain purposes, they are very different kinds of objects than most human discourse. Therefore they tend to be unsatisfactory as communicative acts.

2.2 A GRAMMATICAL DESCRIPTION IS A WORK OF NON-FICTION LITERATURE. Perhaps I am old fashioned, but I still believe that the best model for descriptive grammars is a book that will eventually be available for use in libraries, and on night tables of many interested individuals. I realize, of course, that there are now many other possible models, primarily due to the explosion of computational technologies in the past few decades. History may eventually prove me wrong, nevertheless, I still recommend that grammar writers use the 'non-fiction literature' model for their grammatical description, rather than any of the current more computationally oriented models, such as annotated databases, hypertext documents or expert systems. Even a grammar that is published on the internet, I believe, should have characteristics known to hold of good non-fiction literature. In particular, it should be a coherent whole with a beginning, middle and end (rather than just a system of cross-references to fragmented topics, as so many technical manuals or 'help systems' tend to be). It should also be interesting and engaging to the intended audience. Toward this end, I would like to highlight the following desirable characteristics.

A grammatical description should be generously illustrated with examples. Even if no one remembers or accepts the technical analyses in a grammar, future readers should still be able to use and interpret the data. Linguistic data (not theoretical points or frameworks) constitute the subject matter of a linguistic grammar. Consequently data should always be in primary focus. Data may be presented in charts, e.g., paradigms of verb forms or pronoun

systems, examples elicited in imagined contexts, examples extracted from actual contexts, or entire transcribed and analyzed texts. A good linguistic grammar will employ all of these methods of presenting data, as they all have their particular functions in providing a full and satisfying portrait of a language.

A grammatical description should start with simple topics, and consistent patterns and gradually introduce complexity and irregularity. I find there is a tendency for some grammar writers to state a pattern, and then immediately give the exceptions to the pattern, sometimes even before any straightforward examples! Perhaps this is because grammar writers are so aware of the complexities that the regular, simple and straightforward facts seem like 'lies'. It feels wrong to state a generalization knowing all along that there are many exceptions, contradictions and variations. However, I must implore grammar writers to put themselves in the shoes of their audience for a moment. Exceptions and variation do need to be presented in due course. But if there is a pattern worth stating, it is worth giving the audience time and opportunity to digest it, and to incorporate it into their own emerging 'image' of the language before they are exposed to a complex range of exceptions. The same principle applies to the order in which topics are presented. Most good linguistic grammars start with 'lower level', relatively regular topics—phonology, morphology, noun phrase structure—and work their way 'up' to more complex topics, such as verb phrase structure, clause structure, clause combining, etc. This, in general, is a reasonable approach, though see section 7.2 below for a caveat regarding a strictly structural organization of a grammar.

Most of the exposition in a grammatical description should be in clear prose, rather than complex diagrams, charts and formulas. As mentioned above, formulas and diagrams can be helpful to a grammar writer in clarifying thoughts, and presenting knowledge in a precise way. However, over-reliance on formulas and diagrams can obscure rather than elucidate knowledge. This is especially true when the formulas and diagrams stem from a particular theoretical tradition. Theoretical traditions in linguistics are notoriously short-lived. What is currently in vogue will tomorrow be anachronistic. Much good linguistic work of past decades remains largely inaccessible to modern scholars simply because the frameworks employed have gone the way of the dinosaurs.

In summary, an important part of writing a 'balanced' grammatical description is keeping the big picture in mind. Only if we keep in mind what we are trying to do, can we do it in a way that is communicative and enjoyable to read. Grammar writers will do well to remember that the grammar they are producing is an act of communication and that it is a work of non-fiction literature.

3. COMPREHENSIVENESS VS. USABILITY. All field linguists want to write a comprehensive grammatical description. Field linguists typically have vast knowledge of the language they are describing, and it pains them not to express ALL of that hard-earned knowledge between the covers of the written grammar. However, a completely comprehensive grammatical description has never been written, and if it were it could never be published. Valuable knowledge buried in reams of explicit but relatively minor detail becomes hard to recognize, and virtually useless. Even a thousand page tome cannot represent all the categories and habitualized patterns that make up the grammar of everyday talk. Therefore, all grammars must be less than comprehensive in order to be usable, and hence attractive to potential publishers and readers.

I have known many fieldworkers who have felt paralyzed when it comes to describing the grammar of a language because they don't know where to start. They can't say *anything* until they can say *everything*. Thus, the drive to be comprehensive also works against the need to organize one's work in a way that is clear and understandable.

It has been observed by many researchers (e.g., Grice, 1981, Sperber & Wilson 1995 *inter alia*) that utterances are only partial representations of speaker intentions. Much of what is communicated via language is understood via inference. This is the normal way that humans go about the business of communication, and applies as well to a grammatical description, as I hope to show below. Sometimes saying too much can actually detract from the communicativity of a speech act. Here's an example from an actual conversation:

(1) He's holding her hand the whole time across the table.

A lot of detail has been left out of this utterance. For example, most people have two hands, yet the sentence does not mention WHICH of 'her hands' the subject is holding. In fact, if the speaker did specify 'her right hand' it may be a potential distraction. The hearer may legitimately wonder why the speaker is mentioning her right hand. There must be some relevance to that detail, and the hearer, being a cooperative conversationalist, would try to identify the relevance of the right, vs. left hand.

Something similar is true in grammar writing. Saying too much may not only bury relevant information, but may actually confuse readers. For one small example, consider the issue of word classes, or 'parts of speech'. The importance of word classes has seldom been questioned in discussions of what should be included in a grammatical description of a language. How can you even begin to describe a language if you don't have a clear idea of what the building blocks of that language are? Every grammatical description must at least make mention of nouns and verbs, and probably adjectives, adverbs and some kinds of particles as well. Yet, with a little reflection, it becomes clear that classes such as 'Noun' and 'Verb' are no more than convenient approximations, rather than absolute categories. They are imprecise generalizations that help readers understand something important about a language, but which do not directly correspond to fixed categories in even one language.

If you investigate the grammatical properties of a number of words, you soon find that the lexicon of any language is not divided into clear, mutually exclusive classes. There are in fact very good examples of Nouns and very good examples of Verbs, but many subtly different sub-classes that fall somewhere in between. Each subclass possesses a 'cluster' of grammatical properties that may or may not have any logical coherence. Consider an English word like *slurping*. Is this a noun or a verb? Well it can take a possessor, *his slurping bothers me*, but it doesn't easily occur with quantifiers, or certain other noun modifiers: ??*His many slurpings bother me*, ??*His much slurping bothers me*, ??*His fast slurping bothers me*. So *slurping* has some but not all properties of Nouns. On the other hand, it also has some properties of Verbs; it can take a direct object, as in *His slurping the soup bothers me*. It can also be modified with adverbial modifiers *His quickly slurping the soup bothers me*. Yet it doesn't take a nominative case subject, and cannot inflect like a Verb **He slurpings whenever he eats soup*. This particular cluster of properties cannot be attributed to some subclass, such as abstract nouns or nominalized verbs. For example, clear cases of abstract nouns cannot be modified by adverbials (**his truly sincerity*). Furthermore, some

nominalized verbs with *-ing* take plurals more easily than others: *his many failings* vs. ?*his many eatings*. This fact makes *failing* slightly more 'nouny' than *eating*.

These subtle differences among the behaviors of various forms are probably not available to the fieldworker faced with thousands of forms, each potentially exhibiting a cluster of from zero to about 10 grammatical properties. To exhaustively categorize every word according to the particular cluster of Noun and Verb properties it exhibits is a potentially never-ending task, and may actually be distracting to readers. Therefore the concepts of Noun and Verb stand as imprecise approximations that nonetheless are *precise enough* to be useful in expressing important grammatical concepts (similar to the way *her hand* is precise enough a reference form in example 1). This is but one example of how being too comprehensive can actually make a grammatical description less communicative.

4. TECHNICAL ACCURACY VS. UNDERSTANDABILITY. Accuracy is definitely a value in grammar writing. For this reason, many formalisms and abbreviatory systems have developed over the years as linguists have attempted to make their work as precise as possible. The problem is that formalisms (like language structures themselves) arise within particular communities, and are refined by generations of scholars and their graduate students in PhD dissertations, monographs, and research articles. Readers who lack a background in the specific analytic tradition employed by the grammar writer are likely to be mystified and put off by an over-reliance on formalism and theory-specific terminology. While formalisms and other analytic techniques may increase precision, they often do so at the expense of understandability of the text to future generations.

Furthermore, even the most elaborate mathematical formalisms are still not completely precise. As mentioned several times throughout this chapter, language users employ conventionalized categories and patterns in all kinds of creative ways to communicate unique and nuanced ideas. Is there a 'rule of English grammar' that can explain the structure of the following actual communicative exchange?

(2) A. That boy is silly.
B. He's not silly. He just be's silly when he's around girls.

Certainly speaker B (a 12 year old Anglo-American girl) had a rule in her grammar that made her response reasonable and communicative. And I venture to guess that most native English speakers will find B's response coherent and interpretable, even if they would never use it themselves (or at least would never admit to using it). But is it a rule of 'English' (whatever that is), or simply a quirky 'error' on the part of a less-than-fully-competent speaker? It is my contention that bending conventionalized patterns, and employing them creatively in new and unusual ways is the normal way that people communicate with one another. This is not bizarre, exceptional or erroneous use of language, as studies in corpus linguistics are beginning to show us.

If a grammar writer thinks that all such creative usages need to be incorporated into the written grammar, the task will never be complete. In the case of languages that lack a written tradition, it is especially difficult to determine which usages that appear in natural text are part of 'the Grammar of the Language' and which can be chalked up to individual creativity, performance error, or just plain confused thinking. Of course these distinctions

are not absolutely clear, and even fully competent native speakers will not necessarily agree. Therefore, it becomes another judgment call (a call for 'balance') on the part of the writer as to how 'accurate' one should be about describing the patterns of usages found in natural texts.

5. UNIVERSALITY VS. SPECIFICITY. Each language exhibits features common to all or many other languages, as well as features unique to that particular language. While grammar writers want their work to be usable and understandable by linguists working in other language traditions, and those studying universal characteristics of Language, they also want to highlight the unique and wonderful characteristics of the particular language they have spent so much time learning and analyzing. Often the concepts and terminology that have arisen in other language traditions do not seem to match the categories of the language being described very well, and so one is tempted to devise new and unique terms to describe these new and unique categories. Of course, the more new concepts and terminology are introduced into the written grammar, the more difficult it becomes for readers from other traditions to appreciate.

On the other hand, a language may exhibit a feature that is so distinct from what has been described in previous literature that a new term is necessary. If this is the case, the grammar writer must take care to define the new concept very carefully, and highlight the fact that this is truly new knowledge. For example, at present Doris Payne is grappling with the issue of how to label two tonally marked 'case forms' in Maasai. Let's call them 'Form A' and 'Form B' for now. Form A is the citation form for nouns, and occurs when a noun (of any grammatical relation) occurs before the verb (examples 3a, d, and e) or when an Object noun comes after the verb (3c and d). Form B is used for Subject nouns that come after the verb (whether they are subjects of transitive or intransitive clauses, ex. 3b, c and e):

(3) a. Ɔl-múrránì o-ipid-ó. 'The warrior (FORM A) jumped.'
MSG-warrior.FORM.A 3-jump-PF

b. É-ípíd-ó ɔl-mʊrraní. 'The warrior (FORM B) jumped.'
3-jump-PF MSG-warrior.FORM.B

c. É-tóósh-ó ɔl-mʊrraní ɔl-ásúráí. 'The warrior (B) hit the snake (A).'
3-hit-PF MSG-warrior.FORM.B MSG-snake.FORM.A

d. Ɔl-múrránʊ o-toosh-ó ɔl-ásúráí. 'The warrior (A) hit the snake (A).'
MSG-warrior.FORM.A 3-hit-PF MSG-snake.FORM.A

e. Ɔl-ásúráí é-tóósh-ó ɔl-mʊrraní. 'The warrior (B) hit the snake (A).'
MSG-snake.FORM.A 3-hit-PF MSG-warrior.FORM.B

The question is, how do we label these case forms in a way that genuinely helps readers who work in other language traditions understand the forms, while at the same time highlighting the special characteristics of Maasai? Some options one might consider include:

	Form A	Form B
1.	Object	Subject
2.	Accusative	(Marked) Nominative
3.	Absolutive	(Marked) Nominative
4.	Absolutive	Subjective
5.	Form A	Form B

Options 1-4 are all based on terminology from other established areal traditions, and all can be misleading for one reason or another. Option 1 (Object, Subject) is misleading, since Form A marks Subjects when they appear before the predicate (and in certain other contexts, such as predicate nominals). Option 2 is similarly misleading. Options 3 and 4, are also misleading in that the term 'Absolutive' is usually employed in opposition to the term 'Ergative'. However, in Maasai, there is clearly no Ergative case. The term 'Absolutive' in these options makes reference to the fact that 'Form A' is the 'naming form', or 'citation form', i.e., the form that speakers naturally revert to when a noun occurs outside of any grammatical context.

Option 5 is somewhat of a cop-out, since the labels 'Form A' and 'Form B' make no reference whatsoever to familiar linguistic categories. Such terms are occasionally called for, when categories are so unusual that entirely new terms are needed to refer to them. Once a good friend of mine, David Watters, did a discourse study of two verb forms in Kham in which he underlined all instances of one form with a red pencil and all instances of the other with a blue pencil. After dealing with this analytic technique for some time, he found himself naturally referring to 'red verbs' vs. 'blue verbs', and he developed a rather sophisticated analysis of the functions of these forms in texts. Since the uses of these two forms did not seem to correspond to any previously established categories in the general linguistics literature, he continued to use the terms 'red verb' and 'blue verb' to gloss and refer to the two forms in his initial write-ups. This solution 'worked' for David because he was able to provide content for otherwise grammatically meaningless terms in the process of doing his analysis. Eventually, after studying linguistic work in related languages, David found more 'linguisticky' labels for these forms—'conjunct' and 'disjunct'. These communicated well enough for those familiar with the literature on Tibeto-Burman languages, but still had to be explained in detail for general readers. Many such terminological quirks have resulted from similar decisions that linguists have made in the process of developing grammatical terminology for particular languages. Some examples that come to mind include 'heavy' vs. 'light' vowels, 'strong' vs. 'weak' conjugations, 'soft' vs. 'hard' consonants, and so forth.

6. A FORM-DRIVEN VS. FUNCTION-DRIVEN APPROACH. Every language is a formal, structural system that arises in a human community in response to communicative needs. Every structural piece of a language has both a formal and a functional dimension. Thus, a grammatical description may be organized according to forms—giving the function or functions of each form in turn—, or it may be organized according to functions—giving the form or forms used to accomplish each function in turn. The outlines of most grammatical descriptions to date seem to combine these two perspectives to one degree or another, but

largely without a principled reason for the division. Many problems arise when grammar writers fail to clearly identify forms and functions independently of one another, and therefore mix form-driven and function-driven description in a haphazard manner. For this reason, another area that requires careful balance in grammatical description is between form-driven and function-driven components.

The formal and functional dimensions of linguistic units are closely linked, but need to be identified independently of one another. For example the tool we call a 'screwdriver' is named for one of its functions—driving screws. What it IS is a thing designed specifically for driving screws. However, driving screws is only one of its possible functions. It can also be used for opening paint cans, scraping dirt out of tight corners, as a pointer in an academic lecture, or any number of other functions. The form of a screwdriver is thus logically distinct from any particular function that it might fulfill. It does not cease to be a screwdriver and suddenly become a can opener just because someone uses it to open a can.

Something similar is true of linguistic structures. For a simple example, consider the following expression:

(4) the boy who puts them in his basket

This seems like a noun phrase modified by a relative clause. Most grammar books describe a relative clause as a clause that modifies a noun, and indeed that is probably the major function of most structures that are called relative clauses in the linguistics literature (see, e.g., Keenan 1985, Payne 1997 and many others). However, this structure can also serve quite a different function, as in the following extended example from an actual text:

(5) There's a man in the tree picking pears, and a boy on the ground with a basket. The man throws the pears down to the boy *who puts them in his basket*.

In this example the 'relative clause' *who puts them in his basket* cannot be said to 'modify' the head noun, boy, either restrictively or non-restrictively. Rather, this clause actually asserts a sequential event. First the man throws the pears to the boy, then the boy puts them in his basket. These are two 'foregrounded' events in the event structure of the text (according to, e.g., Hopper & Thompson 1980). So this is one case (out of dozens or hundreds that could be provided) of how structures that primarily serve one function can be used to fulfill other functions.

6.1 ADVANTAGES OF A FORM-DRIVEN DESCRIPTION. Most of the grammars of underdocumented languages that have appeared in the last hundred years or so have been primarily form-driven. There are many good reasons why this is the case, including the following:

A form-driven description is relatively easy to outline. Linguistic structures tend to be more categorical than linguistic functions. That is, form 'discretizes' (makes into distinct categories) open-ended functional 'space'. For this reason, it is easier (though not a simple matter by any means) to identify particular forms, and situate them in the outline of a grammar, than it is to do the same with functions.

A form-driven description is consistent with the way many students and teachers view 'grammar'—a list of structural facts expressed as rules. Perhaps unfortunately, traditional approaches to first and second language teaching and linguistics have evolved with an emphasis on structures, often to the exclusion of functions. Structures, such as nouns, verbs, phrases, clauses, etc. are familiar (if not beloved) to students and teachers, and therefore a structure-driven outline 'resounds' with expectations of what a grammar 'should be'.

A form-driven description can be very clear. It is relatively easy to identify structures —identifying functions is more challenging. This point is closely related to the other two. Once you have a form-driven outline, the task of actually writing the grammar becomes a matter of 'marching through' the outline. Each structural topic can be treated in its own autonomous section, and need not necessarily be influenced by other sections.

6.2 DISADVANTAGES OF A FORM-DRIVEN DESCRIPTION. In addition to the advantages listed above, there are several disadvantages to a strictly form-driven outline, including the following.

A form-driven description can be boring. Why is it that everyone (that is, normal people rather than linguists or grammarians) seems to hate grammar? Grammar is what allows people to communicate with one another, and everyone loves to communicate! One reason for this strange phenomenon, I believe, is the way grammar is conceptualized and taught (see above). Somehow we have gotten the idea that 'grammar' consists of a list of impenetrable formal rules that must be memorized by rote. It is a logical 'machine' consisting of structural parts that have no necessary relation to real life. This conceptualization has worked its way into first and second language grammar classes and even into the linguistics literature. A grammatical description based on this 'grammar as machine' metaphor tends to be dry, boring and difficult to relate to the concerns of real people because it fails to take into account the fact that a written grammar is an act of communication. Payne (2007) is largely an argument against this conceptualization.

A form-driven description emphasizes idiosyncratic facts concerning the formal structures of the individual language, making it more difficult to compare the language to typologically very different languages. Languages are similar in their functions, but quite different in the structures they employ to accomplish those functions. A strictly form-driven description need not relate the language being described to other languages at all, because the universal need to communicate is not in focus. With the rise of typological linguistics, universal properties of human languages have become more of a focus than the sometimes idiosyncratic structural facts of particular languages.

A form-driven description can misrepresent or fail to represent 'functional systems' that span more than one word class or level of structure. Even as a form-driven description can obscure similarities from one language to the next, it can also obscure functional systems within one language. For example, the 'tense system' of English spans at least two levels of grammatical structure:

(6)	You mock my pain!	Present	zero
	You mocked my pain!	Past	morphological
	You will mock my pain!	Future	analytic

In a form-driven grammar, present and past tense would be treated in the word-level chapter while future tense would be treated in the phrase level or syntax chapter. Thus the notion of a unified 'tense system' would be lost.

6.3 ADVANTAGES OF A FUNCTION-DRIVEN DESCRIPTION. The advantages and disadvantages of a function-driven grammatical description are largely the mirror images of the disadvantages and advantages of a form-driven description described above. Nevertheless, I will list these briefly in this subsection and the following.

A function driven description acknowledges the common sense fact that language serves a purpose—namely communication.

A function driven description brings together different structural pieces that conspire to accomplish ranges of communicative functions ('functional systems').

A function driven description makes comparison among typologically distinct languages more possible.

6.4 DISADVANTAGES OF A FUNCTION-DRIVEN DESCRIPTION. A function driven description can be hard to outline. Functions are not discrete and categorical, and therefore it is challenging to identify them, and organize them into a coherent outline.

A function driven description tends to make typologically distinct languages seem more similar to each other.

A function driven description can be 'open-ended' in that almost anything can serve almost any function, given enough context. For a simple example, a 'passive' construction is often defined functionally as one that 'downplays' an AGENT and 'upgrades' a PATIENT. Well if that is the definition of passives, then would the following English sentences all be passives?

(7) The glass broke. (Downplaying the AGENT who broke the glass.)
Some guy broke the most beautiful vase in the world.
These jeans wear easily.
John underwent surgery.
As for okra, I can't stand it.
Okra is what I can't stand.

These all can be construed as somehow 'downplaying an AGENT' and/or 'upgrading a PATIENT'. Without clear structural guidelines, there is no way to constrain the range of sentences that might be construed as fulfilling a given function. For this reason, I would like to argue for a balanced formal/functional approach to linguistic description, to which I turn in the next section.

6.5 THE SOLUTION: A BALANCED FORMAL-FUNCTIONAL APPROACH. The type of grammatical description that I would like to recommend is one that employs a form-driven approach for those areas of grammar that are the most controlled, systematic and rule-dominated, and a function-first approach for those areas that tend to cross-cut structural levels. The controlled, systematic and rule-dominated parts of language include:

a. Phonology (excluding intonation)
b. Morphophonemics
c. Inventory of derivational morphology (which derivational categories apply to which roots, etc.)
d. Inflectional inventory (determining the range of inflectional possibilities for person and number 'agreement' and case marking)
e. Pronoun inventory (isolating the entire set of free pronouns or pronominal clitics)
f. Lexical inventory (acquiring the words for a large number of culturally significant things and activities)

Notice that in this section there is an emphasis on obtaining inventories of various forms. In many cases, this kind of information is best obtained via direct elicitation. This is because languages typically employ a small number of forms in text, though many more forms are logically possible. Full paradigms are seldom constructable based on data that appear in natural texts alone. For example, a declarative sentence with a second person subject is very rare in texts, because people don't often inform other people concerning activities of the person spoken to, e.g., You are baking bread. Questions are much more natural in such a context. Nevertheless, a description of the language would be incomplete if the second person declarative forms were missing. Elicitation is essential to the completion of paradigm charts.

Often the meaning of a particular morpheme or construction is not clear until the entire range of possibilities that could replace it is identified. The same observation can be applied to syntactic constructions. For example, whether a particular transitive construction is a passive or an ergative depends at least partially on whether there exists a corresponding 'active' construction. Similarly, the precise function of Subject-Verb-Object word order may not be apparent until minimal pairs with Verb-Subject-Object order are obtained. Text data may exhibit other orders, but in examples extracted from texts, there are usually enough other formal differences that the precise contribution of word order to the observed semantic differences is obscured. True minimal pairs are usually obtainable only through elicitation.

The more pragmatic, semantic and subtle parts of language are best approached and analyzed from a function-first perspective, via a large body of naturally occurring text, supplemented by elicitation where necessary. These would include:

a. Intonation
b. Constituent order
c. Inflectional morphology (determining the precise functions, including tense/aspect/mode)
d. Voice (alignment of grammatical relations and semantic roles of verbal arguments)
e. Sentence level particles (evidentials, validationals and pragmatic highlighting particles)
f. Clause combining (including relativization, complementation, adverbial clauses and clause chaining)

g. Lexical semantics (determining the nuances associated with various lexical choices, including derivational morphology and pronouns)
h. Pragmatically marked structures, such as clefts, questions, etc.

7. CONCLUSION. Field linguists have several goals in mind when approaching the task of writing a reference grammar. These include:

a. Communicativity. A grammar should clearly communicate complex facts.
b. Comprehensiveness. A grammar should describe ALL the grammatical features of a language.
c. Usability. A grammar should have a well-defined audience who will find the grammar of genuine use.
d. Accuracy. A grammar should be as technically accurate as possible.
e. Universality. A grammar should relate the language described to known universal principles of human language.
f. Specificity. A grammar should highlight the unique and beautiful features of the language described.

Unfortunately, these goals often seem to conflict with one another, and so grammar writers must sometimes partially compromise one goal in order to fulfill another. Many potential grammar writers are paralyzed by tensions caused by such conflicting goals, as they approach the complex task of writing a linguistic grammar. In this chapter I have described several of these tensions, and have argued for a sense of 'balance' in grammatical description. I hope that the suggestions made in this chapter will encourage writers of descriptive grammars by acknowledging the tensions, and giving grammar writers ways to balance competing goals. Grammatical description is an art as well as a science. Writers of linguistic grammars must remember that they are artists creating an esthetically pleasing and engaging piece of non-fiction literature, as well as scientists producing a precise and informative research report.

References

Becker, Alton L. 1979. Text-building, epistemology, and aesthetics in Javanese shadow theatre. In Alton L. Becker & Aram A. Yengoyan (ed.), *The imagination of reality: essays in Southeast Asian coherence systems*, 211-243. Norwood, N. J.: Ablex.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.) *Speech acts* (Syntax and Semantics vol. 3), 41-58. New York: Academic Press.

Hopper, Paul & Sandra A. Thompson. 1980. *Transitivity in grammar and discourse*. Language 56, 251-99.

Keenan, Edward L. 1985. Relative clauses. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol II: *Complex constructions*, 141-70. Cambridge: Cambridge University Press.

Payne, Thomas E. 1997. *Describing morphosyntax: a guide for field linguists*. Cambridge: Cambridge University Press.

Payne, Thomas E. 2007. A grammar as a communicative act, or what does a grammatical description really describe? In Thomas E. Payne & David J. Weber (eds.), *Perspectives on grammar writing* (Benjamins Current Topics 11), 127-142. Amsterdam and Philadelphia: John Benjamins.

Sperber, Dan & Dierdre Wilson. 1995. *Relevance: Communication and cognition* (2nd Edition). Oxford. Basil Blackwell Ltd.

tpayne@uoregon.edu

APPENDIX: A POSSIBLE OUTLINE FOR A BALANCED GRAMMATICAL DESCRIPTION.

The following is one possible outline for a grammatical description that tries to 'balance' form-driven vs. function-driven approaches. This outline is humbly offered as a source of ideas and guidance to fieldworkers who feel a need for such guidance. As discussed in the chapter itself, every grammar exhibits unique characteristics based on the interests, goals, abilities, and personality of the grammar writer, as well as characteristics of the language and of its sociolinguistic situation. Hence this is not meant to be a 'checklist' a 'field manual' or a 'strait-jacket', but simply a source of ideas for elaborating a grammatical description.

Items followed by an asterisk (*) are considered essential. Other items may or may not appear in the grammar outline, depending on a) the intended use of the grammar, b) the special experience and interests of the author and c) the individual characteristics of the language. Of course any particular grammar may also include more headings than what are found here.

Front matter
- Acknowledgements (*)
- Introduction (Including theoretical assumptions and purpose of the grammar.)

List of abbreviations (*)

Part I: The Cultural, Ecological and Sociolinguistic Context of the Language
- 1.1 The name of the language (*) Including 'endoethnonyms' (the name people use to refer to themselves) 'exoethnonyms' (terms used by outsiders to refer to a particular ethnic group).
- 1.2 Previous research (*)
- 1.3 Demography (*) Number of speakers, location and other linguistic groups in the area.
  - 1.3.1 Map(s)
  - 1.3.2 History/migrations
- 1.4 Ecology
- 1.5 Ethnography (material culture, cosmology) (*)
- 1.6 Genetic and areal affiliations (*)
- 1.7 Literary traditions
- 1.8 Dialects (including classical/written varieties if applicable) (*)

- 1.9 Sociolinguistic situation (*)
  - 1.9.1 Multilingualism and language attitudes
  - 1.9.2 Contexts of use and language choice (*)
  - 1.9.3 Viability (*)
  - 1.9.4 Loan words
- 1.10 The corpus (*)
  - 1.10.1 The nature of the research (affiliation, location, duration) (*)
  - 1.10.2 Consultants and other sources (*)
  - 1.10.3 Presentation of data (*)

Part II: Structural Overview (form driven approach)

- 2.1 Typological Sketch
- 2.2 Phonological inventory and orthography (*)
  - 2.2.1 Consonants (*)
  - 2.2.2 Vowels (*)
  - 2.2.3 Tone / stress (*)
- 2.3 Phonetics (*)
- 2.4 Syllable structure
- 2.5 Word structure
- 2.6 Major phonological and morphophonemic processes (*)
  - 2.6.1 Process 1 (*)
  - 2.6.2 Process 2 (*)
  - 2.6.3 Process 3
  - 2.6.4 Process 4
- 2.7 Relaxed speech rules and contractions
- 2.8 Word Classes (*)
  - 2.8.1 Nouns (*)
    - The structure of the noun word
    - Derivational processes
    - Inflectional processes
    - Count vs. mass nouns
    - Proper names
    - Other grammatically distinct subclasses of nouns
  - 2.8.2 Pronouns and/or anaphoric clitics (*)
    - Personal pronouns (*)
    - Demonstrative pronouns (*)
    - Other (relative pronouns and question words may be more efficiently treated in the sections on relative clauses and questions.)
  - 2.8.3 Verbs (*)
    - Verb structure (a diagram of a verb and its morphology) (*)
    - Derivational (stem-forming) processes
    - Inflectional processes
    - Grammatically distinct verb subclass 1
    - Grammatically distinct verb subclass 2
    - Grammatically distinct verb subclass 3

- 3.2.10 Antipassives
- 3.2.11 Object demotion and/or omission
- 3.2.12 Object incorporation
- 3.2.13 Other valence related constructions

3.3 Nominalization
- 3.3.1 Action nominalization
- 3.3.2 Participant nominalizations
  - Agent nominalizations
  - Patient nominalizations
  - Instrument nominalizations
  - Location nominalization
  - Product nominalizations
  - Manner nominalizations
  - Action or clausal nominalization

3.4 Tense/aspect/modality (*)
- 3.4.1 Tense
- 3.4.2 Aspect
- 3.4.3 Modality
- 3.4.4 Location/direction
- 3.4.5 Evidentiality, validationality and mirativity
- 3.4.6 Miscellaneous

3.5 Pragmatically marked structures (*)
- 3.5.1 Constituent order variation
- 3.5.2 Contrastive/emphatic particles
- 3.5.3 Contrastive/emphatic intonation patterns
- 3.5.4 Cleft constructions
- 3.5.5 Negation (*)
- 3.5.6 Questions (*)
  - Yes/No Questions (*)
  - Question word (information, content) questions (*)
- 3.5.7 Imperatives (*) (including subtypes)

3.6 Clause Combinations (*)
- 3.6.1 Serial verbs
- 3.6.2 Complement clauses (*)
- 3.6.3 Adverbial clauses (*)
- 3.6.4 Clause chaining, medial clauses and switch reference
- 3.6.5 Relative clauses (*)
- 3.6.6 Coordination

3.7 The Language in use
- 3.7.1 Lexical Typology
  - Space, direction and motion
  - Causation
  - Perspective
  - Salience of semantic features

- 3.7.2 Continuity (cohesion) and discontinuity
  - Topic (referential) continuity
  - Thematic continuity
  - Action continuity
- 3.7.3 Episodic prominence
  - Climax/peak
  - Intensification
- 3.7.4 Genres
  - Conversation
  - Narrative
  - Personal experience
  - Historical
  - Folk stories
  - Mythology
  - Hortatory
  - Procedural
  - Expository
  - Descriptive
  - Ritual speech
- 3.7.5 Miscellaneous and conclusions
  - Idiomatic expressions / proverbs
  - Sound symbolism
- 3.7.6 Summary of typological findings

End matter
- Text with interlinear translation
- Glossary
- References cited (*)
- Index (*)

7

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 109-119
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4587

Endangered domains, thematic documentation and grammaticography

Jacob Terrell
University of Hawai'i at Mānoa

When setting out to document a language with the intended goal of describing it (typically through a grammar and dictionary), fieldworkers prefer to collect an array of linguistic data, ranging from elicited words and paradigms to an assortment of texts based on conversations, narratives, procedures and so forth. Capturing a wide variety of speech acts provides a clearer record of the language and its use, and thus offers the potential for a richer description of the language at hand. However, without controlling for content, one may collect linguistic data based on an open-ended amount of topics or themes. The purpose of this chapter is to introduce the notion of endangered linguistic domains and themes in language documentation and description. Even in thriving minority languages, domains such as indigenous music or knowledge of flora and fauna come under pressure from the same forces that eventually lead to language endangerment. Gathering linguistic data based on a particular domain or specialized knowledge can generate a corpus applicable to a wider audience without sacrificing the needs of linguists. Similar to thematic dictionaries in lexicography, this introduces thematic grammars to grammaticography.

1. INTRODUCTION.[1] The purpose of this chapter is to introduce the notion of endangered linguistic domains as they pertain to thematic documentation and grammaticography through corpus building. When setting out to document a language with the intended goal of describing it (typically through a grammar and dictionary), fieldworkers often collect an assortment of texts of various genres (Himmelmann 2006). Capturing a wide variety of speech acts in a corpus provides a clearer record of the language and its use, and thus offers the potential for a richer description of the language. For example, Mosel (2011b) provides a useful overview of what fieldwork guides recommend one collects in terms of data, including recording different genres such as oral histories, narratives, explanatory texts, artistic texts (songs, poems, etc.) and so on. However, without controlling for domain or topic, one may collect linguistic data based on an open-ended number of subjects.

In this chapter I argue that in some cases it is appropriate to control for topic and focus on a single domain or genre in a documentation project. This is a novel concept, and so it leads us to ask how one should go about thematic documentation of an endangered domain, how this approach may have an impact on a grammar, and how the outputs of such a project can be of use to the community, linguists, and others. To address this, Section 2 raises the notion of endangered linguistic domains. The following section introduces the idea of thematic documentation for corpus building and grammaticography. Following this, I will present my work with the Akha shaman of northern Thailand as a case study, discussing

1 I would like to thank the editors of this volume, Toshihide Nakayama and Keren Rice, as well as an anonymous reviewer for their comments on previous versions of this chapter. I am also grateful to Pirma Gavq Lavq for sharing the Ahka shaman chants with me, and to Miqder Saeduq for her constant assistance. Last, the case study described below would not have been possible without support from the Hans Rausing Endangered Languages Project and the World Oral Literature Project.

ISBN: 978-0-9856211-4-8

how the thematic documentation project has shaped the resulting corpus, dictionary, and grammar.

2. ENDANGERED DOMAINS. Even in thriving minority languages, domains such as indigenous music or knowledge of flora and fauna come under pressure from the same forces that eventually lead to overall language endangerment. In fact, one may presume that language endangerment is a systematic process where individual domains first become moribund and then disappear, gradually leading to the endangerment of the language as a whole. This is reflected in the fact that before becoming moribund, the domains of endangered languages are reduced from every type of social application down to the home domain. At the time of writing, there is no definition of an endangered domain or genre, nor any rubric to measure the level of endangerment as is often applied to languages as a whole. However, the concept is beginning to receive some attention.

For example, in a post titled Endangered Genres[2] on the *Endangered Languages and Cultures* blog hosted by Paradisec, Peter Austin writes:

> It is by now well known that around half (or possibly more) of the world's 7,000 languages are endangered and under threat of disappearance during the current century. Perhaps less well known is that many languages that are not (yet) endangered show certain genres, or ways of using the language, that are endangered in that there are few people who can perform them and occasions for their use are diminishing. We could refer to these as 'endangered genres'.

Austin goes on to discuss a literary tradition of the Sasak of Indonesia, where they once recorded manuscripts on lontars, the dried leaves of a type of palm, to be read during performances. These manuscripts are unique in that they are "written in Kawi, a form of middle Javanese, or Sasak, or a mixture of both". Austin further states that this literary practice, or domain, is highly endangered, "as there are probably only 100 people (among a population of 2.5 million) who can read the manuscripts, and performances are discouraged due to cultural associations which conservative Islamic groups on Lombok do not approve of".

Relatedly, Tim Brookes of the Endangered Alphabets[3] project writes:

> Writing has become so dominated by a small number of global cultures that those 6,000-7,000 languages are written in fewer than 100 alphabets. Moreover, at least a third of the world's remaining alphabets are endangered—no longer taught in schools, no longer used for commerce or government, understood only by a few elders, restricted to a few monasteries or used only in ceremonial documents, magic spells, or secret love letters.

Like the previous example of Lontar manuscripts, this is a complex case of endangered domains in that not only does it involve a single domain, that of literacy, but in some cases other domains may be endangered as well, such as chanting, magic, and certain literary practices.

2 http://www.paradisec.org.au/blog/2010/12/endangered-genres
3 www.endangeredalphabets.com

Finally, perhaps no organization has given more attention to the idea of endangered domains than the World Oral Literature Project (WOLP) through the University of Cambridge Museum of Archaeology and Anthropology. The project was initiated in 2009 and its goal is to record, document and archive collections of endangered oral traditions. The WOLP defines oral traditions on their website as follows:[4]

> Most simply, oral literature refers to any form of verbal art which is transmitted orally or delivered by word of mouth. These creative works are increasingly endangered as globalisation and rapid socio-economic change exert complex pressures on smaller communities, often eroding expressive diversity and transforming culture through assimilation to more dominant ways of life. As vehicles for the transmission of unique cultural knowledge, local languages encode oral traditions that become threatened when elders die and livelihoods are disrupted.

All projects funded by the WOLP place an emphasis on the domain of oral literature. These include collections of chants, various literary genres, indigenous religion, epics, ballads, and others. Although still in its infancy, the WOLP is by far at the forefront of thematic documentation. Currently, the organization has funded fifteen projects that are in various stages of completion.

The Lontar manuscripts, the Endangered Alphabets project, and the projects funded by the WOLP all have two things in common. First, the projects control for topic by focusing on a particular domain; they show that language documentation is cultural documentation, a concept that many fieldworkers would agree with. More importantly, they illustrate that the relationship between language documentation and conservation and cultural documentation and conservation is not unilateral. It is common for proponents of language documentation and conservation to argue that when a language is lost, the culture is lost as well. However, the examples above demonstrate that it is equally true that when a particular cultural practice is lost (or replaced), the language associated with that practice is lost too. Thus, there is a bilateral rather than unilateral relationship between the two—a concept not often recognized. For this reason, it is important for fieldworkers who set out to document and describe a language to begin considering thematic documentation of endangered domains.

3. THEMATIC DOCUMENTATION. It is often the case that the younger generations are less likely to be well-versed in detailed cultural practices and indigenous knowledge due to the ease of replacing these with more dominant local or global ones. For example, a community may be transitioning from an indigenous religion to one of the world's major religions, or learning life skills in schools rather than in the forest, or purchasing clothing from a large department store rather than weaving one's own.[5] Whatever the case, it is this detailed cultural or indigenous scientific knowledge that is most likely to be cherished by

4 www.oralliterature.org

5 I should clarify here that I am not suggesting that any group not have access to national education, services, or goods. Instead, once these practices disappear, so does the language and knowledge which is used in these domains. Hence, when the domain is lost without any documentation, so is the language associated with it.

elders and valued by researchers in other fields. Therefore, gathering linguistic data based on a particular domain may generate a corpus applicable to a wider audience.

Of course, a researcher setting out on a thematic documentation should choose a domain or genre related to personal interests. Although potential endangered domains for a thematic documentation project depend on the environment of the speakers, the increasingly globalized world is (unfortunately) providing more topics to choose. As such, these are nearly endless: cyclical and life-cycle ceremonies, knowledge of flora and fauna, indigenous music, religion, textiles, and so on. These are indeed hypothetical examples, but still plausible. In fact, it is not necessary for the domain in a thematic documentation project to be an endangered one. The point of thematic documentation is to control for the topic to create a database of specified knowledge, which in turn should have a wider-reaching audience than just linguists.

With that said, thematic documentation itself is not a new concept—field linguists and anthropologists have long been interested in lexical semantics and domains such as kinship terms or those related to ethnobotany, ethnobiology or folk taxonomy (Haviland 2006). The Dictionary Development Process[6] initiated by Ron Moe at SIL is a good example of thematic documentation in that it provides a template for a lexicography project based on semantic domains. Additionally, theme-based lexicography projects also allow for quick production of thematic dictionaries, each based on a particular domain, creating a focused product that is appealing to not only the community but also researchers in other fields (Mosel 2011a).

As mentioned in section 1, typical products of a documentation project are often a grammar and dictionary, and, similar to thematic dictionaries in lexicography mentioned above, thematic documentation introduces thematic grammars to grammaticography. Of course, this means that examples in a grammar based on a theme-based corpus would be restricted to its domain. However, this does not mean that such a corpus cannot fulfill the needs of linguists. For example, if one's hobby is fishing, it would be possible to gather needed linguistic data for a grammar and dictionary based on the domain of fishing. This could include names of fish, procedural texts on making nets/traps, fishing methods, folklore involving fish, narratives about fishing, and others. A corpus of this type could be of interest to biologists, ichthyologists, anthropologists, and other researchers, depending on what the corpus contains and what is made available in the translations of the texts. Additionally, the researcher would have an excuse to spend much time in the field fishing.

4. THE AKHA SHAMAN, A CASE STUDY. To exemplify the points made above, I draw on my experiences working with the Akha shaman, or Pirma, in Northern Thailand to document the chants of the indigenous religion. Akha is a Tibeto-Burman language, belonging to the Lolo-Burmese branch (Bradley 1979). The language is spoken in five countries: southern China (est. pop. 200,000) Laos (66,100), Myanmar (200,000), Thailand

6 http://www.sil.org/computing/ddp/

(56,600), and Vietnam (1,260) (Ethnologue 2009).[7] With a population well over 500,000, this speech community is much larger than those which are usually the focus of a language documentation project. The language is vibrant and is still being learned by children. However, due to recent cultural changes, the register used in the traditional religion is highly endangered; in fact, it has become moribund in the last two to three generations. According to the Mekong Akha Peace and Sustainability network, an NGO made up of community leaders who work to establish cultural networks across borders, there are only three shaman in China, fewer than fifteen in Thailand, and perhaps four in Myanmar. The number of shaman in Laos and Vietnam is unknown (Wang, pc.).

One of the most important roles of the shaman is to carry out funeral rites. A funeral can be extremely extravagant; it can last up to a week and a number of animals are sacrificed to send off with the deceased, including up to three water buffalos. During the rite, the shaman will sit in front of the coffin and chant for the deceased to take the soul to the border of the afterworld. The chants are epic poems covering topics from creation stories, to conception, death, the relationships between spirits and humans, crops, animals, and much more. There is one volume of chants for each of the three buffalos that are sacrificed at the funeral, and it can take up to a week for a shaman to perform a three-buffalo funeral. The chants are all done by memory, and all the shaman that I interviewed report that it takes up to twenty years to memorize all three volumes. Still, as mentioned earlier, this practice is highly endangered. The shaman are all elders, and I do not know of any apprentice under the age of forty.

It is important to note here that the religious register used by the shaman is not mutually intelligible with contemporary, spoken Akha. The community believes that the religious register is a fossilized version of contemporary Akha. I have played recordings of chants for numerous Akha who will recognize that it is Akha, and that it is the shaman's language, but they always report that they cannot understand it. Given the dire situation of the shaman's language, some community members suggested that I direct my attention towards documenting the religion. Thus, in the fall of 2009 I initiated a thematic documentation project of shaman's chants and indigenous religion based out of Chiang Rai, Thailand.

4.1 DOCUMENTING THE CHANTS.[8] With the help of funding from the Endangered Language Documentation Programme (ELDP) and the World Oral Literature Project (WOLP), I established a team of five native speakers of Akha to assist with the project. Three team members were in their twenties, while two were in their fifties. The team members were trained as native-speaking documenters in basic linguistic description, the applicable tools and technology, and methods of ethnography. For linguistic description, much of the curriculum was based on what is used at the Language Documentation

7 These figures are from Ethnologue, based on Bradley 1997. However, they may not be accurate due to many Akha migrating from Myanmar and into Thailand who do not receive citizenship and are not included in any census. For example, the Akha Foundation NGO in Chiang Rai believes that there are around 100,000 Akha in the 284 villages in the northern Thailand (Kukeuwsakem, pc).

8 In the 1980s, Leo Alting von Geusau and Inga-Lill Hansson worked together to initiate a project to translate cultural texts of Akha. They began with *Oer Zar*, 'Life Cycles', as recited by Pirma Arso Dzoeqbaw and Argaw Dzoeq, and produced a manuscript in 2002. Unfortunately, the project came to a halt after Dr. Geusau's untimely death in the same year.

Training Center (LDTC) run by the Linguistic Society of Hawai'i.[9] The team members were also taught how to use Audacity, ELAN, Toolbox for glossing texts, and Lexique Pro for building the dictionary database. We also practiced using the equipment: shotgun and lavalier microphones, Zoom recorders, and the video recorder. In all, the training session was quite intensive and lasted just over a month.

After the training session, we would take turns traveling to different villages in northern Thailand to record chants and interviews, check collected texts, and take photographs of relevant items related to the chants and the religion. We worked primarily with one shaman, Pirma Gavq Lavq of Huay Pra Sot village. Much of our time was spent annotating, glossing and translating the chants, since we were working with one source language (the religious register), and two target languages (contemporary Akha and English). We found that it would take nearly a month for two team members to process a thirty-minute text, creating a solid first draft of the text—glossed with free translations in Akha and English. After we finished one text, we would travel to two villages to check our work with at least two other shaman. We continued to collect and process data in this manner throughout the duration of the project. In all, we collected and processed the data for almost a year and a half, from August 2009 through December 2010.

5. RESULTING CORPUS.[10] The data in the corpus can be divided into two categories: 1) primary audio recordings of the chants performed by Pirma Gavq Lavq, and 2) secondary audio and video recordings and texts related to the chants.[11] Since the purpose of this project was to record, document, and describe the songs of the Akha shaman, we focused on processing Prima Gavq Lavq's chants at the expense of processing the secondary data.

The corpus contains 10 hours and 9 minutes of Pirma Gavq Lavq chanting Aqnyoq tiq mawr, 'Buffalo one', the first volume of chants performed at a funeral. In total, the chants contain over 6,500 verses of prose.[12] Each verse has been transcribed in ELAN using the standard Akha orthography. We have a free translation from the religious register to contemporary Akha for the majority of the verses (perhaps ninety-five percent), though there are some sections that neither various shaman nor we could decipher. We also have a free translation in English for about seventy percent the verses that could be translated into contemporary Akha. Every verse with an Akha and English translation is also accompanied by a gloss in both contemporary Akha and English.

6. OUTPUTS. In another chapter in this volume, Mosel discusses the importance of the types of texts one collects and methods in building a corpus for writing a grammar. Additionally,

9 http://www.ling.hawaii.edu/~uhdoc/

10 The corpus is currently being prepared for archiving with the Endangered Languages Archive (ELAR) at the School of Oriental and African Studies (SOAS) at the University of London, and with the WOLP. The first three chants of 'Buffalo One' are more public than the others are, and thus may be accessed in the archives without restrictions. The remaining chants may be accessed with permission from Mekhong Akha Peace and Stability (MAPS).

11 The secondary data include chants from other shaman, mourning songs performed at a funeral, interviews with elders on how to carry out funeral rites, written texts involving folklore, and other materials.

12 For a comparison, the English version of the Christian New Testament contains roughly 7,500 verses.

Rehg (this volume) argues for the need to include speakers of the target language in the documentation and description processes, providing speakers with a larger role in both. In this case study we recorded different types of texts related to religion (procedural, narratives, etc.), but translating the chants remained our priority. Furthermore, the Akha assistants were more involved with the documentation of the chants and compiling the dictionary (and are listed as coauthors), while they mainly served as consultants for the grammar.

6.1 DESCRIPTIVE OUTPUTS: DICTIONARY AND GRAMMAR. The most recent draft of the dictionary contains approximately 3,300 entries. It is a trilingual dictionary: Shaman's Akha—Contemporary Akha—English. Lexical items include everyday contemporary Akha words, their clipped forms used in the shaman song, and lexical items that are unique to the shaman's chants. Since this is a thematic documentation project, the majority of the items in the dictionary revolve around the material culture and flora and fauna mentioned in the chants. To some extent this product may be of use to the community, since it is a trilingual dictionary.

I do not believe that an academic grammar will be of great use to the community, other than perhaps adding some prestige to the language and the shaman's role. (This is because, in part, it is written in English.) Instead, the intended audience for the grammar are linguists and others scholars. The grammar is based on the first three chants of 'Buffalo One': *Lavq Khoer Mr*, 'Inner Ceremony' (ch-gl-A.txt), *Gee Jm*, 'the Ancestor Alter' (ch-gl-B.txt), and *Khoer Naevq Dav*, 'Inner Spirits' (ch-gl-C.txt). In total, these texts amount to fifty-four minutes of chanting, some 446 verses. Below are two example verses.

(1) *Khoer Naevq Dav*[13], 'Inner Spirits' (ch-gl-C.004.txt)[14]

CH:	*Nya*	*ˈlaw*	*aq.ˈpoeq*	*ghaq*	*ˈor*	*zer.ˈzaq zer.ˈsanr*
AG:	*Lavq.tav*	*law*	*aq.poeq*	*ghaq*	*or*	*zer.zaq zer.sanr*
EG:	Above	PTC	ancestor	CLF.person	PTC	guardian.of.children

CH:	*tiq*	*ˈeeq*	*ˈor!*
AG:	*tiq*	*eeq*	*or*
EG:	one	gather	PTC

Akha: *Lavqtav paw eq jawr aqpoeq nar a zerzaq zersanr tiq deq mae!*
Eng: Oh!, Above, the ancestors, the guardian spirits of the children gather together!

(2) *Lavq Khoer Mr*, 'Inner Ceremony' (ch-gl-A019.txt)

CH:	*sar*	*ˈnan*	*manq.ˈpaq*	*dziq*	*ˈawr kawr*	*ˈlmr*	*tmr*	*ˈe*
AG:	*yaw sar*	*ar.nan*	*manq.paq*	*dziq*	*awr yaw kawr*	*yaw lmr*	*ngaeq*	*e*
EG:	easy	day	stallion	ride	and quick	warm	speak	GEN

13 Akha is a tonal language, and tones are represented in the orthography as: $-r, high; $-0, mid; $-q, low. Also, $-v indicates creaky voice.

14 CH stands for chant, or what the shaman sings in each verse, AG is the gloss in contemporary Akha, and EG is the English gloss.

CH: *ar.ˡnan ˡleir.*
AG: *ar.nan leir*
EG: day NPST

Akha: *Yaw sar arnan anr manqpaq dziq awr yaw kawr aer ngaeq e arnan lei.*
Eng: A relaxing day, it is a day for riding stallions and making quick, warm conversation.

Both registers use SOV order with modifiers following the head of the noun phrase: i.e., stative verb, degree word, determiner, and then classifier. The language has postpositions rather than prepositions. There is no difference in word order in complex constructions such as passives, causatives or relative clauses. However, the structure of the verses differs from that of contemporary Akha in the interaction between prosody, morphology, and semantics.

For example, in the data above one can see that each verse contains an odd number of syllables. Verses can range in length from five syllables to twenty-one or more. The metric foot in the chants is iambic consisting of two syllables, and the final, odd syllable of a verse is footed. Obviously, not all clauses or sentences in contemporary Akha consist of an odd number of syllables. To maintain the iambic rhythm with an odd number of syllables in the chants insertion and/or clipping is used.

The morphemes glossed as PTC in (1) above are non-lexical vocables[15], or syllables in song that do not have meaning but do have function. Here, their function is to ensure that the verse contains an odd number of syllables. Although the poetic syllables above fall on the stressed beat of the foot, they may also be inserted in unstressed positions. These syllables are always [a], [o], or [u], with either high or low tones.

While insertion and clipping work in tandem to ensure that a verse has an odd number of syllables, clipping is more complex. For example, there is a strong tendency in Akha for verbs, stative verbs and closed word classes to be monosyllabic. Nouns can be disyllabic, many of which have only two syllables. Note in (2) that there are two forms for 'day' in the same verse: *nan* and *arnan*; the latter is the term used in contemporary, spoken Akha (speakers do not recognize *nan* as 'day'). Also, the clipped syllable in 'day' is always *ar-* and never *-nan*. The syllable that can be clipped in a disyllabic noun is specified for each lexical item, as shown below.

TABLE 1. FULL AND CLIPPED NOUNS IN THE CHANTS

A			**B**		
English	**Full Form**	**Clipped Form**	**English**	**Full Form**	**Clipped Form**
water	ir.cuvq	ir	gourd	ir.puq	puq
head	uq.duq	uq	horn	uq.coer	coer
banana	nga.baev	nga	bean, pod	ar.baev	baev
mouth	maeq.boeq	maeq	tongue	maeq.lar	lar
earth	mir.tsaq	mir	country	mir.khanq	khanq

15 This is the term commonly used in music. I prefer the phrase 'poetic syllable' and the adhoc gloss PTC for simplicity.

In column A above the second syllable may be clipped in the chants, while in column B the first syllable of the noun may be clipped. The option to clip either the first or the second syllable depends on what I call the semantic weight of the syllable. For 'water' the syllable with the semantic weight is *ir-* (retained in clipped form), while for 'gourd' it is *-puq*. This is a result of a noun class system in Akha where nouns in a particular semantic domain share a common syllable through extension. Other nouns in the class of water/liquids include *irbanq*, 'cup', *irxoi* 'urine', *irxmr* 'otter', *irnei* 'mud'. Nouns associated with the head often begin with *uq-*, as in *uqduq* 'head', *uqcoer* 'horn', *uqghmq* 'pillow', and *uqlanr* 'turban'. Similar examples can be found for the other nouns above.[16] These types of extensions are frequent in the lexicon, and speakers do not recognize the clipped forms above as lexical entries.

The shaman may use either *ircuvq* or *ir* for 'water', or *irpuq* or *puq* for 'gourd', as long as the syllable with the semantic weight falls on the stressed beat in the iambic foot, *ir* and *puq* respectively. If the syllable with semantic weight is not aligned with the stressed beat of the foot, clipping, insertion or a combination of the two can be used.

Returning to 'day' in (2), the syllable with the semantic weight, *nan*, falls on the stressed beat of the foot; hence *sar ˈnan* for 'easy day', not **sar ˈar nan*. Likewise, this constraint requires the full form *arnan* in *tmr ˈe ar.ˈnan ˈleir* 'a day of speaking' rather than **tmr ˈe nan ˈleir*, since the penultimate syllable in the verse and the final odd syllable must be stressed.

For the purpose of this chapter, this is only a simplified explanation of the interaction between prosody, morphology, and semantics in the chants. The phenomena above, coupled with the unique lexical items, render the chants mutually unintelligible with contemporary, spoken Akha. This adds to the uniqueness of the grammar, since the thematic documentation approach allows for a grammar and dictionary based on a highly endangered genre of a language, one that most native speakers do not grasp. Another impact that this approach has had on the grammar-writing process is that all data revolve around religion—the chants and their translations in contemporary Akha. The distinctive patterns found in the chants are best described in comparison to their translations in contemporary Akha, and despite being focused on a single genre, the size of the corpus still allows for observing grammatical regularities in both registers. In conclusion, though the corpus focuses on an endangered genre, it still offers a wealth of information for a grammar.

6.2 COMMUNITY USES. As a fieldworker, I would like to think that this project will have a positive impact on the community. Unlike the descriptive products mentioned above, the corpus has proved to be of much more use and value. At the very least, the community now has access to audio recordings, transcriptions and translations of the verses in 'Buffalo One'. It is now up to members in the community to choose how they wish to use these materials. To start, in December of 2010 MAPS organized a workshop in Shan State of Myanmar (Burma) to bring the shaman from Thailand to meet with their counterparts in Myanmar. Representatives from China came as well. This was a three-day training workshop on traditional religion and culture. All team members from the thematic documentation project were asked to share the project with the other participants. They discussed using the corpus

16 There are similar rules for compound nouns and other disyllabic words; however, this is beyond the scope of this chapter.

to promote the indigenous religion through learners' materials for a future generation of shaman apprentices, with the intention to produce textbooks with accompanying karaoke VCDs of the shaman song for distribution to villages where there are youth interested in learning the shaman's texts. This is also related to MAPS's first major project; the NGO recently received a grant from the US Embassy in Thailand in order to initiate a project on cultural literacy. This is a two-year long project aimed at producing instructional video and literacy materials explaining how to carry out cyclical festivals and rites. One team member, Lawq Gaw, has taken it upon himself to write the manual. The book is entirely in Akha, 165 pages long (Wang 2011). Another team member has finished editing the accompanying video guide on how to conduct a funeral ceremony in early 2012. Finally, drafts of the first three chants were produced in 2010. These books are meant to be used along with the audio recordings. Each contains a preface with instructions on how to use the book, and each verse in the audio recording is followed by a translation in contemporary Akha and explanatory footnotes where needed.

In summary, although the resulting grammar may not be of use to the community, the corpus that the grammar and dictionary are based on is proving to be useful due to its focus on an endangered genre. I also hope that the corpus described above will be useful to researchers in other fields. Common themes in the chants include creation, life, death, crops, animals, plants, migration routes and so on. For example, the dictionary contains over fifty names of bamboo species, classified according to folk taxonomy (which is reflected in the word-form). This information may be of use to botanists or ethnobotanists. Furthermore, the chants are all sung, which may be of interest to ethnomusicologists. Additionally, as with many religions, the numbers 3, 5, 7, 9, and 12 carry significant importance in the chants—an interesting concept that has received little attention in anthropology and is still not discussed in religious studies.

7. CONCLUSION. It is my hope that this chapter will open a dialogue on thematic documentation for language description with the intent of building theme-based corpora for thematic grammars focusing on endangered domains. As mentioned in section 1, a comprehensive documentation project includes not only corpus, but also a dictionary and grammar based on the corpus. Each field setting is different, therefore making it difficult to provide generalizations for carrying out fieldwork for language documentation and description projects. However, by controlling for topic when setting out for a language documentation project, fieldworkers can better produce corpora that are more useful to the community, and applicable to a wider audience.

References

Austin, Peter K. 2010. *Endangered Genres*. http://www.paradisec.org.au/blog/2010/12/endangered-genres. (05 May, 2011.)

Bradley, David. 1979. *Proto-Loloish*. London: Curzon Press.

Bradley, David. 1997. *What did they eat? Grain crops of the Burmic groups*. Mon-Khmer Studies 27.161-70.

Brooks, Tim. 2011. *Endangered Alphabets*. (26 May, 2011.)

Geusau, Leo A. von (ed.). 2002. *Oer Zar: life cycle of Zaqnyiq peoples, animals, and plants*. Chiang Mai, Thailand: unpublished ms.

Haviland, John B. 2006. Documenting lexical knowledge. In Gippert, Jost, Nikolaus Himmelmann and Ulrike Mosel (eds.), *Essentials of language documentation*, 129-162. Berlin, New York: Mouton de Gruyter.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Gippert, Jost, Nikolaus P. Himmelmann, and Ulrike Mosel (eds), *Essentials of language documentation*, 1-30. Berlin & New York: Mouton de Gruyter.

Kukeuwsakem, Yotesapong. Personal communications. (09 June, 2010.)

Lewis, M. Paul (ed.). 2009. *Ethnologue: Languages of the World, Sixteenth edition*. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com (01 June, 2010.)

Mosel, Ulrike. 2011a. Lexicography in endangered language communities. In Austin, Peter & Sallabank, Julia (eds.), *The handbook of endangered languages*, 337-353. Cambridge: Cambridge University Press.

Mosel, Ulrike. 2011b. Morphosyntactic analysis in the field – a guide to the guides. In Nick Thieberger (ed.). *The Oxford handbook of linguistic fieldwork*, 72-89. Oxford: Oxford University Press.

SIL. 2006. *Dictionary Development Project*. http://www.sil.org/computing/ddp/ (01 January, 2012.)

Terrell, Jake, Argawv Byevtseirguq, Ardoq Ceiganr and Miqder Saeduqguq. 2010. *Pirma dawq – Aqkaq dawq – Pa la dawq Dir xaq naq liq. Archaic Akha – Modern Akhaq – English dictionary*. Unpublished ms.

Wang, Jianhua. personal communication. (28 May, 2011.)

Wang, Jianhua (ed.). 2011. *Tsawrxir pavqkhawq gurtsov: A manual of Akha funeral ceremonies*. Chiang Rai, Thailand: Mekhong Akha Network for Peace and Sustainability (MAPS).

World Oral Literature Project. 2009. *Definitions and undesrstanding oral literature*. http://www.oralliterature.org/about/oralliterature.html. (26 May, 2011.)

jake.terrell@gmail.com

ABBREVIATIONS

AG	Akha gloss	GEN	genitive
CH	chant	NPST	nonpast
CLF	classifer	PTC	poetic syllable or non-lexical vocable
EG	English gloss	$	full syllable, as in (C)V

8

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 121-134
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4588

Walking the line: Balancing description, argumentation and theory in academic grammar writing

Carol Genetti
University of California, Santa Barbara

This chapter explores how to incorporate linguistic typology, argumentation, and theoretical innovation into a reference grammar. It provides recommendations on how to produce a balanced grammar that is firmly grounded in theory, responsible to the unique structures of the language, and comprehensible now and over time. Linguistic typology provides a set of widely recognized linguistic categories used in the classification of grammatical patterns. These can be taken as starting points from which the structures of the language can be compared, contrasted, explored, and explained, profiling the unique shapes of language-particular categories. Argumentation for particular analyses provides clarification and explanation, although excessive argumentation can obscure descriptive facts. Simply asserting facts is appropriate for lower-level linguistic features, simple canonical structures, or uncontroversial elements or their functions. Argumentation is appropriate when structures differ from typologically-expected patterns, when the analysis counters descriptions in the literature, and in cases of multiple interpretations of a structure. Grammar writing immerses researchers in the structure of a language, revealing new vistas of understanding and novel ways of interpreting structure. Theoretically innovative analyses that reflect these insights can be incorporated as long as they are motivated, well-explained, and balanced by a typologically-informed descriptive base.

1. INTRODUCTION. An academic reference grammar is a complex study which can be enriched by incorporating diachronic, ethnographic, and theoretical dimensions. The grammar writer must determine how much he or she can incorporate these dimensions and still achieve the primary goal of presenting the facts of the language in an accessible and interpretable way. For example, inclusion of excessive diachronic discussion can obscure the synchronic details and produce a study that is primarily historical; this is clearly a worthy endeavor, but one with a different function and goals than a descriptive reference grammar. The same is true for theory: a grammar that is devoid of deeper observations lacks richness and depth; on the other hand, grammars with too much theoretical machinery will be dated and obscure to anyone not trained within that theoretical paradigm. It is also true for argumentation: while argumentation provides invaluable depth to the analysis, excessive argumentation can be tedious and frustrating for the reader. This chapter thus addresses a practical question: how does one incorporate typological analysis, structural argumentation, and theoretical innovation into a reference grammar while still presenting the descriptive facts in a clear and accessible manner? Or, phrased differently, how can a grammar writer do justice to the language-specific richness and variety of structural categories without being either straight-jacketed by typological and theoretical convention or overrun by it? It is all about finding balance.

Section 1 of this chapter addresses linguistic typology. Specifically, it discusses how to balance the typological classification of linguistic forms with the imperative to describe

ISBN: 978-0-9856211-4-8

the unique and language-particular shapes of structural categories. Section 2 moves to the question of argumentation and discusses conditions under which incorporating argumentation is appropriate in contrast to conditions that call for the simple assertion and exemplification of descriptive facts. Section 3 then looks to theoretical innovation. Grammar writing allows one to view the complex interactions between grammatical subsystems with exceptional clarity, opening new vistas of understanding and novel ways of interpreting structure. Here the discussion considers how to balance conventional descriptive categories with new insights that push theoretical boundaries.

2. BALANCING TYPOLOGICALLY ESTABLISHED PATTERNS WITH LANGUAGE-SPECIFIC CATEGORIES. Linguistic typology provides us with an inventory of critical structural categories and relationships that guide the fieldworker and constitute the basis for the conception, analysis and presentation of linguistic structure. These categories include, for example, the following: linguistic units (e.g., segments, syllables, words, phrases); word classes (e.g., noun, verb, adjective, numeral); commonly coded meanings (e.g., plural, past tense, perfective, negative, hearsay); grammatical relationships that hold between words and phrases (e.g., subject, head, dependent, embedding); and construction types (e.g., serial verbs, complementation, imperatives, etc.). The grammar writer is thus faced with the question of how much of the grammar should be shaped by our understanding of linguistic typology. Logically, it is possible to provide two extreme answers to this question that represent opposing poles: 'entirely' and 'not at all'. Of course, there is a continuum between these. I will discuss two approaches which can be placed on the continuum approximating these opposite poles: the 'checklist' model of grammar writing, and the 'all is unique' model. Balance is found at an intermediary point.

One approach to grammar writing is to take the identification of pre-established linguistic types as the primary goal of grammatical investigation. This is called the 'checklist' model of grammar writing since one goes down a list of typologically-defined categories and notes the presence or absence of features in the language in question, filling in specific details as appropriate. The obvious problem with this approach is that a language may have features that are not covered in the questionnaire. For example, the *Lingua Descriptive Studies Questionnaire* (Comrie and Smith 1977) has nothing on evidentials.

Even if none of the features in the language were to be absent from the list, there is a more fundamental problem with this approach. Each language has not only a subset of the world's linguistic structures and categories, but also elements that may be unusual or unique, have fuzzy boundaries, be 'bistructural' (Genetti 2007), or be shaped to allow participation in larger language-particular systems in reference to other elements. Identifying types on a checklist can prevent one from seeing both the fine detail and the bigger picture. While a checklist can be a useful starting point for guiding field research, a more balanced, richer, and accurate grammar is produced when one seeks to elucidate the unique shaping of structures and categories that interact with each other in complex ways. Figure 1 contains a visual metaphor for this idea, representing grammar as a multi-dimensional puzzle of differently-shaped, interlocking elements that fit together to create a cohesive whole:

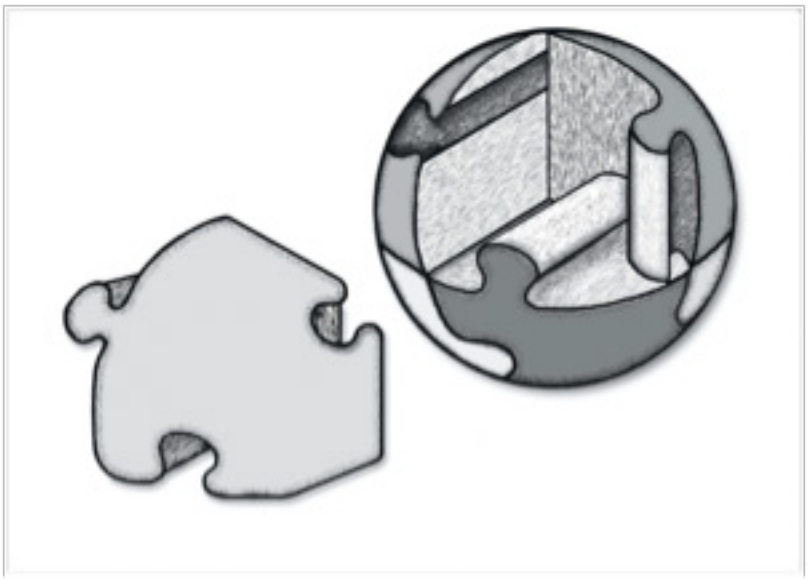


FIGURE 1: Grammar as a set of interlocking, uniquely shaped categories

Only a nuanced approach to representing grammatical categories will provide the data and analysis to expand our set of typological classifications and allow our understanding of linguistic typology to continue to grow.

If going down a checklist to identify whether or not a language has a particular category represents one end of a continuum of incorporating typological theory into descriptive grammars, the other end is then represented by producing a grammar without any reference to cross-linguistically established categories. Such a grammar would be uninterpretable (possibly even impossible to write, e.g., without reference to categories such as noun and verb) as one would need to invent an entirely new, yet comprehensive, set of language-specific terminology. Some grammars go further down this path than others. Consider, for example, the following quote from James A. Matisoff's *Lahu Grammar*:

> The simplest vC's are binary, with a single vV preceding the Vh. We have been using 'β' to symbolize the verbal nucleus of a VP; this is, the obligatory Vh plus any versatiles that may optionally be juxtaposed to the head. We may then generate binary vC's by some such rule as the following: β -> (vV) + Vh. (Matisoff 1972:211)

Mattisoff was writing in the late 1960's, the heyday of Chomskyan generative grammar and the infancy of modern linguistic typology (i.e. Greenbergian and its antecedents). In addition, he was writing about a language that is, in typological terms, vastly different from the Western European languages that were the focus of the early (and much of the later) work in Generative Grammar. Without the tools of contemporary functional-typological linguistics, he had to take a unique, and frequently idiosyncratic, approach. Matisoff's grammar is extraordinary in its richness, depth, and insight; however, it takes commitment

on the part of the reader to learn the terminology sufficiently enough to understand the text. In short, it is not a grammar for the casual browser.

It is possible to find a balance between these poles by taking the linguistic categories identified by typological studies as a starting point for deeper exploration, rather than having the mere identification of the categories be the goal and endpoint of the investigation. For example, one might write:

These are serial verb constructions.

This sentence only identifies a set of examples as belonging to a particular structural type. Argumentation for the analysis could be provided, which would take the grammar to a deeper level, e.g.:

These are serial verbs because they have the following properties that define the serial verb category in the typological literature…

However, one goes even further if one looks beyond the horizon of constructions identified by typologists. There may be considerably more to say about the precise details of the language-particular construction that suggest shades of meaning, relations with other structures, or other conceptualizations. Returning to our example, if one takes the statement above on serial verbs as the starting point of the discussion, it can then be followed by more complex observations, e.g.:

However, some examples suggest that: this is actually clause linkage / some of these verbs have grammaticalized to auxiliaries / some of these verbs function as true prepositions / etc.

This approach allows the emergence of a grammar that is firmly built on linguistic typology—and so is typologically relevant—but that also explicates the nuances and complexities that fully shape linguistic categories. Only with this grammar can one explore those fascinating interfaces between categories in all their messy glory.

As an example, consider the case of Dolakha Newar 'adjectival verbs'. Linguistic typology traditionally provides us with two distinct lexical classes: adjectives and verbs. Dolakha Newar has both of these classes, each with their own characteristic phonological and morphosyntactic features. Some of the properties that distinguish them are given in Table 1:

Adjectives	**Verbs**
Closed class	Open class
Multiple phonotactic shapes	CVC structure of root
No inflection	Inflect for multiple categories
Occur with copula in predicate	Occur independently in predicates
Modify nouns directly	Must be nominalized to modify nouns
Used referentially with clitic	Must be nominalized for referential use
Occur with intensifiers	Do not occur with intensifiers

TABLE 1: Properties of adjectives and verbs in Dolakha Newar

If we were limiting our investigation to a typological checklist, we may decide we had successfully identified and distinguished these categories and stopped at that. However, closer investigation reveals a third category, which I have called 'adjectival verbs'; these are a hybrid category, with the syntactic properties of adjectives, but the phonological and morphological properties of verbs. The seven features presented in Table 1 are distributed for adjectival verbs as shown in Table 2:

Like Adjectives	**Like Verbs**
Closed class	CVC structure of root
Occur w/ copula in predicates	Take verbal inflections
Occur with intensifiers	Must be nominalized for referential use
	Must be nominalized to modify nouns

TABLE 2: Features characteristic of Dolakha Newar adjectival verbs

We can capture the hybridity of this category with the Venn diagram in Figure 2.

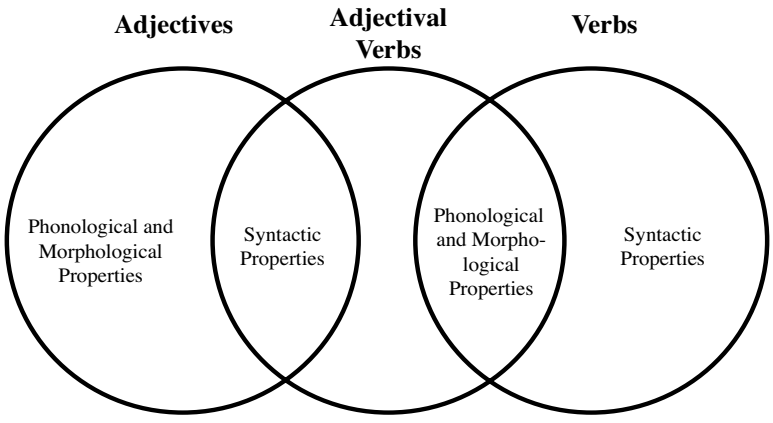


FIGURE 2. Venn diagram illustrating the hybrid nature of adjectival verbs

So while the standard categories of basic linguistic theory provide an important starting point for identifying lexical categories, deeper investigation allows us to identify an additional class with distinct patterns of behavior. This fact, in turn, allows for the enrichment of our typologies and the development of linguistic theory.

In sum, taking typology as the entryway to the investigation provides balance to a grammar, as it (1) allows for the exploration of the language-particular shapes of grammatical categories, (2) incorporates theoretical notions without constraining the description and (3) allows for a continuous feedback loop between linguistic description and the development of linguistic theory.

2. ASSERTION AND EXEMPLIFICATION VERSUS ARGUMENTATION. Another choice facing the grammar writer is just how much argumentation to incorporate in order to justify the presented analyses. The answer ranges logically from providing no argumentation to providing arguments for every point. The former would entail simply asserting descriptive analyses, presumably providing illustrative examples. An example of assertion and exemplification would be:

Complement clauses have structure X.
[Example 1]
[Example 2]
[Example 3]

The latter would be to provide argumentation for each claim, informing the reader why one has analyzed the structure in that way. Continuing our hypothetical example, the argumentation approach might look something like the following:

Complement clauses are clauses that function as noun-phrase arguments of a complement-taking predicate. In this language, complement clauses have the following structure…

There are three facts that confirm that these clauses function as arguments of a complement-taking predicate:
[Argument 1 with examples]
[Argument 2 with examples]
[Argument 3 with examples]

Argumentation clearly enriches the grammar by allowing the reader to understand the grammar-writer's reasoning; on the other hand, there are problems with arguing for every point, just as there are problems with providing no argumentation at all.

There are three problems with providing insufficient argumentation. First, the given analysis is likely to be unconvincing; readers have no means by which to follow the grammar writer's reasoning and might not see the logic of the analysis, even though it may be obvious to the grammar writer. Second, a grammar without argumentation lacks descriptive richness. Argumentation naturally incorporates a deeper level of discussion and the assertion and connection of facts that might not otherwise be made. Finally, the absence

of argumentation results in a grammar that is less interesting to read, as it does not engage the reader in the analytical process.

On the other hand, excessive argumentation can be tedious and runs the risk of obscuring the descriptive facts. In principle, one could present spectrograms that would provide evidence for the phonetic value of every consonant in the language. While this might delight a small number of phoneticians, most people would find this excessive in a grammar and better suited for publication in a phonetics journal. Also, taken to its limit, providing argumentation for every fact of the language would require extensive discussion of non-occurring patterns. For example, if a language has a maximal syllable template of CCV, one might have to argue for this by demonstrating that certain syllable shapes are not attested or accepted by speakers (e.g., that there are no syllables of shapes CCCV, CCVC, CCVV, CCVVC, etc.). An analogous example from syntax would be to list all the possible ordering permutation of noun phrase elements, including those that are ungrammatical to an absurd degree (like **black three bears big the* in English). It makes more sense to simply state what is found, rather than providing long lists of what is not found.

One can find balance between these two extremes by determining which types of phenomena are better served by the two approaches. In my own practice, I found assertion and exemplification to be appropriate for the following types of phenomena (I have illustrated these with examples from *A grammar of Dolakha Newar* (Genetti 2007) where appropriate):

A. LOW-LEVEL DESCRIPTIVE FACTS

- Phonetic values of segments
- Simple phonological processes
- Allomorphic variation (e.g. stem classes)
- How verbs are borrowed
- Phonotactic structures of verbs
- Verb paradigms
- Ordering of elements in the noun phrase

B. DEFINITIONAL STATEMENTS

> Non-finite verb forms differ from finite verb forms in that they do not convey information about tense, person, or number, and in that they do not have separate suffixal paradigms which indicate negation or mood. (Genetti 2007:186)

C. EXPECTED PATTERNS AND STRUCTURES

- Proximal and distal demonstratives
- Numeral systems
- Casemarkers
- Simple clause structure
- Interrogative pronouns

D. USAGE OR DISTRIBUTIONAL PATTERNS

- The uses of the present tense
- The sets of nouns classified by numeral classifiers

- The uses of the various demonstratives
- The distribution of the allative case marker
- Conditions under which noun phrases have post-verbal placement
- Different uses/meanings of two imperative constructions

On the other hand, argumentation enriched the discussion in the following circumstances:

A. CASES WHERE MORE THAN ONE STRUCTURAL ANALYSIS IS POSSIBLE

> The adverbs of location are distinct from locational nouns in that they cannot occur within a noun phrase… (Genetti 2007:230)
>
> These alternative views of the syllable structure have different descriptive goals. One describes the syllable structure as it is likely to be understood by the speakers…The second describes the syllable structure as revealed by patterns of distribution… (Genetti 2007:62)

B. CASES WHERE ILLUSTRATING THE COMPLEXITY OF THE PHENOMENON IS MORE IMPORTANT THAN DEFINITIVELY CHOOSING ONE ANALYSIS

> The primary reason to consider the plural morpheme to be a clitic rather than a suffix is … in the absence of a head noun, it can be bound to a genitive phrase or relative clause … On the other hand, [the morpheme] is not always bound to the final element… [and] it can occur on both elements of a conjoined NP… (Genetti 2007:97-98)

C. CASES WHERE THE LANGUAGE DIFFERS FROM TYPOLOGICALLY-EXPECTED PATTERNS

> In this chapter, I have described two classes of 'adjectivals'…Adjectival verbs still can inflect…The class of simple adjectives, by contrast, has no inflection. This is a major difference in morphological behavior and argues that the two adjectival categories are lexically distinct… (Genetti 2007:212)
>
> In many languages one can grammatically distinguish between classes of objects…direct and indirect objects…[or] primary objects and secondary objects…However, in Dolakha Newar neither of these patterns is in evidence. Instead, all O and R arguments appear to constitute a single grammatical relation of object. (Genetti 2007: 315-317) (The text goes on to present three distinct arguments for this point.)

D. CASES WHERE THE LANGUAGE DIFFERS FROM AREALLY- OR GENETICALLY-EXPECTED PATTERNS

> In the Himalayan area, one typically sees significant split ergativity, however that is not true for Dolakha Newar. To argue for this unexpected pattern, Genetti (2007:109-110) provides examples of consistent ergative marking even in the presence of features which typically condition splits. Hence the following examples are given, each arguing that the designated feature does not affect ergative marking:

(55) ergative case w/ negated verb and non-volitional agent
(56) cognate-object verb *talk a talk* where the object is not differentiated from the action of the verb itself
(57) highly unaffected object
(58) continuous, imperfective aspect
(59) non-active verb, unindividuated object, non-agentive agent, and future imperfective

E. CASES THAT COUNTER EXPLICIT CLAIMS IN THE LITERATURE

> Other restrictions, of the type commonly found on related constructions in other languages (see e.g. Haspelmath and König 1995), are not in evidence in this language…the participial construction does not impose constraints on anaphora, control, the scope of interrogative or imperative mood, or the scope of negation… The argumentation…will be summarized briefly here. (Genetti 2007:446)

> It should be noted that both relative clauses and nominal complements constitute modifiers of nouns within a single unified noun phrase and do not occur as independent noun phrases in appositional relations with the heads (cf. DeLancey 1999; Noonan 1997). This can be seen from… (Genetti 2007:389)

In sum, the best balance can be reached by judiciousness as regards the inclusion of argumentation. Argumentation should especially be included when illustrating multiple analyses is beneficial or in cases that are surprising or go against expected claims or patterns. In other circumstances, simply asserting and describing a given grammatical structure, with plenty of illustrative examples, is likely to be sufficient.

3. THE ROLE OF THEORETICAL INNOVATION IN A REFERENCE GRAMMAR. There is no such thing as an atheoretical grammar, as theory is implicit in all of our terms and concepts, and undergirds how we conceptualize linguistic structures and the relationships among elements. A grammar necessarily reflects the writer's contemporary theoretical orientation, as practically every statement is imbued with assumptions about the nature of language and with current conventional understandings about linguistic structures and categories. As noted above, a grammar that eschewed contemporary theoretical conceptions of language would be very difficult to write and probably impossible to interpret. On the other hand, grammars that are based too heavily on a specific instantiation of an idiosyncratic theoretical paradigm are also problematic due to the fact that narrow linguistic theories can change rapidly and new generations of linguists are not trained in the older theoretical paradigms. This is especially true of formal theories and their representational apparatus. As an example, consider the tagmemic grammars written (primarily) in the 1970s. Statements such as the following necessarily require specific training in tagmemics to interpret:

> The Identificative Adjective Declarative Clause Type has the following identificational-contrastive features:
> it is not a division-subclass of the Submissive passive Complement filler class in the Submissive Passive Clause Type 6;

its minimum nuclear structure is composed of an obligatory, and an optional, nucleus tagmeme.

MINIUMUM NUCLEUS FORMULA

±S ⟨ N-Phr / Pn-Phr / V-Phr / LRelAx-Phr / Adv-Phr ⟩ +Pr[DeclAdj-Phr]

(Liêm 1969: 23)

For this reason, it is best to avoid theory-specific formalisms and to describe the linguistic structures using terms and categories on which there is broad consensus, such as those defined by basic linguistic theory (e.g. Dixon 2010) and used broadly in the literature on linguistic typology.

On the other hand, writing a grammar provides the author with an understanding of the language that is broader and sometimes deeper than can be achieved by analyzing single constructions or subsystems. The very act of explicitly describing the structure in prose necessitates the conscious consideration of every grammatical fact and brings into view structural and semantic nuances, unanticipated connections, and novel ways of conceptualizing the relations between grammatical elements. It is thus a natural venue for proposing unconventional analyses and theoretical innovations. In addition, readers have within the volume deep background information on the relevant grammatical phenomena, so can easily follow the author as he or she lays out the new terrain.

To exemplify this point, I will refer to a portion of the discussion of complex sentences in Dolakha Newar given in Genetti (2007). In Dolakha Newar, the sentence is a central unit of grammatical and discourse structuring. Sentences are bounded by the production of a non-embedded finite verb, and the clauses preceding the main clause link up via a wide variety of constructions. Properties of sentences (both syntactic and prosodic) were the topic of Chapter 21, the final chapter of the grammar, which could refer back to facts that had already been established in the proceeding chapters, from casemarking to grammatical relations to the various clause-combining constructions.

The discourse of typological linguistics and basic linguistic theory include the combinatorial principles of subordination and coordination. However, these terms have a number of implications that do not map well onto the descriptive facts of Newar. Similarly, the structures found in this language are not quite perfect fits to terms like 'clause-chaining' (e.g. Longacre 1972, 1985), 'converbs' (Haspelmath 1995), and 'co-subordination' (Foley and Van Valin 1984:256-263; Van Valin and La Polla 1997:454 inter alia). The theoretical issues surrounding this were fully discussed in an earlier journal article (Genetti 2005) and, to a lesser extent, in Chapters 19 and 20 of the grammar. The embedded structures found in relative clauses and complementation were described previously.

Due to an inexact fit between the facts of Newar and the terms *subordination* and *coordination*, I found it more insightful to conceptualize clause combining as involving two 'design principles' of chaining and embedding. Chaining is the linear ordering of units at the same level of syntax, whereas embedding entirely incorporates one clause into an-

other. The chaining structures in Dolakha Newar include what are typically analyzed as adverbial clauses, converb clauses or clause chains, as well as multiple sentences that are embedded into a direct quotation. These structures are not typically categorized together within linguistic theory. One of the atypical claims of my grammar is that the final clause in a chained structure, which most would consider to be the 'matrix' clause, lacks the syntactic or rhetorical privilege implied by this term. In my view, the final clause has no such privilege in this language, but is simply the last unit in the chain (Genetti 2007: 452-453). For this reason I represent chaining structures as simple linear strings, and diagram them as in Figure 3:[1]

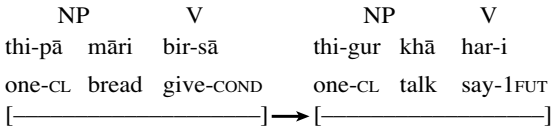


FIGURE 3. A chaining structure diagrammed
Free translation: If you give me a (piece of) bread, I will tell you one thing.

Embedding then comprises the remaining clause-combining strategies: relativization and complementation. Whereas chained clauses were diagrammed on the same horizontal plane, reflecting the symmetrical relation between units that I argued for, embedded clauses were diagrammed on a lower horizontal plane. An arrow was included indicating the noun phrase of the matrix clause which contained the embedded element. An example is given in Figure 4:

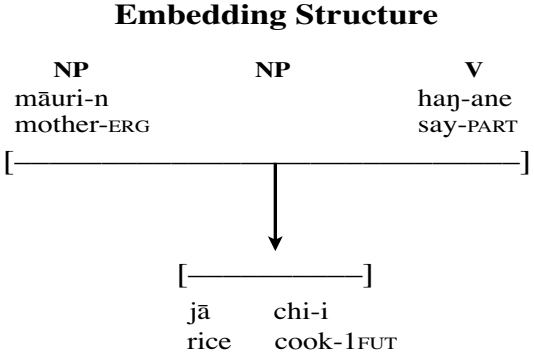


FIGURE 4. Diagram of an embedding structure
Free translation: The mother said: "I will cook rice."

The advantage of this method of diagramming is that it provides one with a simple tool to visually represent the complexity of sentences that arise when the two design principles

1. Full discussion of these structures is beyond the scope of the current chapter. Readers are referred to Genetti (2007; 2011); the latter is a journal article that followed from the new conceptualization of the structure. The abbreviations found in the surrounding examples include: 1 first person; ABL ablative; CL classifier; COND conditional; DAT dative; ERG ergative; FUT future; IND individuation; IMP imperative; LOC locative; NR nominalizer; PART participle (converb).

are recursively applied: any chain can contain embeddings and any embedded element can in turn contain chain. This creates structures such as that in Figure 5:

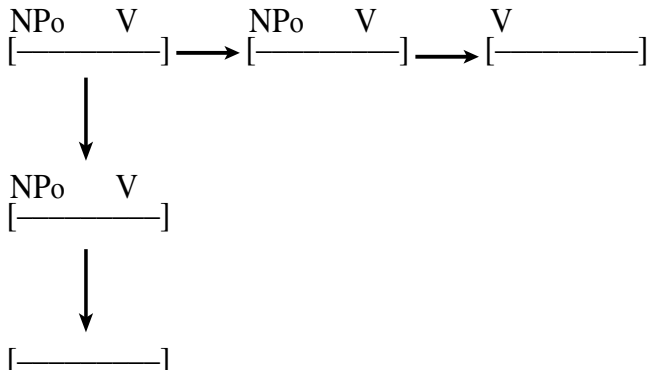


FIGURE 5. Diagram for a relative clause embedded into an object NP, which is within an embedded object complement, which is within the first clause of a participial chain[2]

The innovation in both the conception of the relationships and the diagramming conventions allowed for a richer and more accurate presentation of the facts than would otherwise have been possible.

It is important to note that although this analysis does not follow other theoretical paradigms directly, it clearly draws on them, a point which was made in Chapter 21 of the grammar, which directly cites theoretical work, especially that of Role and Reference Grammar (Foley and Van Valin 1984; Van Valin and LaPolla 1997). Readers are able to understand the new analysis because they share this theoretical base (or can access it, if they do not share it), and because the entire grammar is at their fingertips; I was able to explore the new approach without space limitations or the need to explain all the details from scratch.

In sum, it is best to avoid presenting the analysis using a narrow theoretical model that is likely to go out of date; it is preferable to use terminology from basic linguistic theory that has come into widespread international use, then explain any terminological deviations with reference to this work. Writing a grammar provides the author with the depth of insight and the opportunity to understand grammatical categories and constructions in new ways and to present the language to others in these terms. However it is important to provide

2. The sentence this diagrams is the following (Genetti 2007: 500-501; Genetti 2011):

(15) [[[*chẽ=ku=ri=na* *chanta bi-e*]$_{REL}$ *māsāku māsāku cijbij*]$_{NP-O}$
house=LOC=IND=ABL 2SDAT give-NR untasty untasty things
sumake na-e]$_{NP-O}$ *sukā-en* *janta* *hā-en* *bi-u*
silently eat-NR pretend-PART 1SDAT bring-PART give-IMP
'Silently pretend to eat the not-tasty not-tasty food that they give you from in the house and bring and give it to me.'

argumentation that supports innovative analyses, and to explain, justify, and be explicit as to precisely how the analysis is innovative. However, it is also critical that the innovations be clear and easily understood by the reader; analyses that are excessively idiosyncratic will cause many people to stop reading. Balance can be found by building on the bedrock of conventional structural description and argumentation and moving upwards from there.

4. CONCLUSIONS. Academic reference grammars are squarely theoretical works and it is helpful to think through precisely how theoretical terminology, conceptions, argumentation, and innovations are best incorporated. To write a grammar that will be interpretable for centuries to come requires incorporation of the set of theoretical terms and conventions that have emerged internationally through the practice of grammar writing (i.e. basic linguistic theory). At the same time, however, the grammarian's insights into the uniqueness of a particular language are invaluable for our collective understanding of linguistic diversity, the cognitive sciences, and anthropological linguistics. Here I have attempted to present some guidelines for finding balance in the incorporation of theory and argumentation into a reference grammar. I hope that these will be useful to others who look to find their own balance in grammar writing, and that this approach can be fruitfully applied to historical, ethnographic, or other domains in the production of these complex works.

References

Comrie, Bernard and Norval Smith. 1977. *Linguistic descriptive studies: questionnaire*. *Lingua* 42: 1–72, http://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaire/linguaQ.php

DeLancey, Scott. 1999. Relativization in Tibetan. In Yogendra P. Yadava &Warren W. Glover (eds.), *Topics in Nepalese linguistics*, 231–249. Kamaladi, Kathmandu: Royal Nepal Academy.

Dixon, Robert M. W. 2010. *Basic linguistic theory*, Volumes I and II. Oxford: Oxford University Press.

Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Genetti, Carol. 2011. The tapestry of Dolakha Newar: Chaining, embedding, and the complexity of sentences. *Linguistic Typology* 15. 5-24.

Genetti, Carol. 2007. *A grammar of Dolakha Newar* (Mouton Grammar Library 40). Berlin: Mouton de Gruyter.

Genetti, Carol. 2005. The participial construction of Dolakha Newar: Syntactic implications of an Asian converb. *Studies in Language* 29.1, 35–87.

Haspelmath, Martin. 1995. The converb as a cross-linguistically valid category. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms*, 1–56. Berlin & New York: Mouton de Gruyter.

Liêm Nguyễn Đăng. 1969. *Vietnamese grammar: A combined tagmemic and transformational approach; A contrastive analysis of English and Vietnamese*, Vol. 2 (Pacific Linguistics Series C No. 4). Canberra: Pacific Linguistics.

Longacre, Robert E. 1972. *Hierarchy and universality of discourse constituents in New Guinea languages*. Washington, D. C.: Georgetown University Press.
Longacre, Robert E. 1985. Sentences as combinations of clauses. In Timothy Shopen (ed.), *Language typology and syntactic description*: Volume 2, 236–286. Cambridge University Press.
Matisoff, James A. 1972. *The grammar of Lahu* (UC Publications in Linguistics 75). Berkeley: UC Press.
Noonan, Michael. 1997. Versatile nominalizations. In Joan Bybee, John Haiman, & Sandra A. Thompson (eds.), *Essays on language function and language type*, 373–394 (Typological Studies in Language). Amsterdam & Philadelphia: John Benjamins.
Van Valin, Robert D. & Randy LaPolla. 1997. S*yntax: Structure, meaning, and function*. Cambridge: Cambridge University Press.

cgenetti@linguistics.ucsb.edu

9

Language Documentation & Conservation Special Publication No. 8 (July 2014):
The Art and Practice of Grammar Writing,
ed. by Toshihide Nakayama and Keren Rice, pp. 135-157
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4589

Corpus linguistic and documentary approaches in writing a grammar of a previously undescribed language

Ulrike Mosel
Universität Kiel

Drawing on her experiences with writing a grammar in the course of the Teop language documentation project, the author explores how corpus linguistic methods can be employed for the analysis and description of a previously undescribed language. After giving a short introduction into the creation of a digital corpus and complex corpus search methods, the chapter focuses on the importance of creating a diversified corpus. It demonstrates that different text varieties such as spoken and written legends, procedural texts and descriptions of objects show different preferences for certain ways of expression and thus represent valuable resources for various grammatical phenomena. Accordingly, a grammar which is based on texts should account for this variation by incorporating a detailed description of the corpus, giving references and metadata for each example and providing information on the kind of contexts particular grammatical features are usually associated with.

1. INTRODUCTION. Most publications on linguistic field methods emphasize that a collection of recorded, transcribed and analyzed texts is the most important source for the grammatical description of a previously undescribed language (see Bright 2007:16, Chelliah 2001, Crowley 2007:121, Dixon 2010:321 among many others). But only two older field manuals (Samarin 1967:55-68, Rivierre 1992:56-63) and the recently published *Handbook of descriptive linguistic fieldwork* (Chelliah & De Reuse 2011:422-44) give some information on what constitutes a good corpus for grammaticographers and how the texts that are typically collected during fieldwork can be classified. The crucial questions, however—how an annotated corpus of texts is created and what kind of grammatical information can be gained from different text varieties—have been neglected in descriptive and typological linguistics.

Therefore, I would like to open the discussion on this topic by making a few suggestions of how the writers of grammars of previously undescribed languages can build up a diversified text corpus, and illustrate this corpus linguistic approach by examples from my own research on Teop. Teop is an Oceanic Meso-Melanesian language of the North-West Solomonic linkage (Lynch et al. 2002:101-102), spoken by approximately 6000 people in the Autonomous Region of Bougainville, Papua New Guinea. Our project was one of the first language documentation projects funded by the Dokumentation Bedrohter Sprachen program of the Volkswagen Foundation (Mosel et al. 2007), but besides this documentation I continuously worked on a Teop Reference Grammar. At the same time I learned to use the language documentation tool ELAN (see §2) and became interested in modern corpus linguistics, which completely changed my way of writing a grammar compared to the methods we employed when writing the *Samoan Reference Grammar* (Mosel & Hovdhaugen 1992).

ISBN: 978-0-9856211-4-8

In the following, I will briefly explain some corpus linguistic methods of grammatical analysis and grammar writing in §2, then in §3 discuss how a corpus can be compiled that meets both the wishes of the speech community and the interests of the grammaticographer, and in the following sections focus on three kinds of grammatical variation:

1. the grammatical variation in spontaneous oral texts and the edited versions of these texts (§4);
2. the preference for certain grammatical constructions in particular text varieties (§5);
3. the pervasive use of certain constructions in texts on special themes (§6).

My experiences suggest that the four phases of the grammar writing process — text recording in the field, corpus compilation and annotation, data analysis and description — are so closely interrelated that they should be integrated into a holistic methodology.

2. CORPUS LINGUISTIC METHODS IN GRAMMATICAL ANALYSIS AND GRAMMAR WRITING. The use of text collections as the basis of grammatical analysis makes the writing of grammars of previously undescribed languages a kind of corpus linguistic enterprise, although it is impossible to meet the demands of quantitative corpus linguistics and investigate grammatical variation on the basis of a corpus of millions of words as it is nowadays done for the compilation of grammars of European languages, e.g. Biber et al. (1999). But what seems worth doing is to gather a corpus that comprises texts of various kinds, analyze and describe grammatical categories and constructions, identify linguistic variation across text varieties and interpret the preferences for certain linguistic features in relation to the contexts where they occur. As for the terms text and text variety, I follow Biber and Conrad (2009). While the term text refers to 'natural language used for communication, whether it is realized in speech or writing', text varieties are defined by their situational characteristics, which include the channel, relations among participants, production circumstances, communicative purposes and the topic (Biber & Conrad 2009:5, 40).

Linguistically significant variation is especially noticeable in comparable corpora where two text varieties only differ with respect to one or two variables as, for instance, the transcription of a spontaneously narrated legend and the edited version of this transcription (see §3), or a narrative about the butchering of a chicken and a procedural description of how people butcher chickens (see §5.3).

A corpus gathered in the course of fieldwork is certainly not representative for the language as such, but only for a few selected text varieties. As will be further elaborated on in §3, fieldwork corpora differ from conventional corpora in that the selection of texts is not primarily guided by linguistic or demographic criteria. Rather, especially in the beginning of the research project, the sampling is determined by the external conditions of the fieldwork site, and consequently, classifies as "haphazard, convenience, or accidental sampling" (Kalton 1983:90, quoted in Meyer 2002:43).

For writing a grammar the most useful kind of text collection has the form of a digitalized annotated corpus that links audio or video recordings to transcriptions and translations, provides for each text metadata, is accessible via the internet (Austin 2006),

and allows the user to search with a query language like Regular Expressions (see below §2.2).

2.1. ANNOTATION. The most sophisticated tool for compiling a corpus of a previously undescribed language is ELAN which besides or in combination with Toolbox is widely used in language documentation projects.

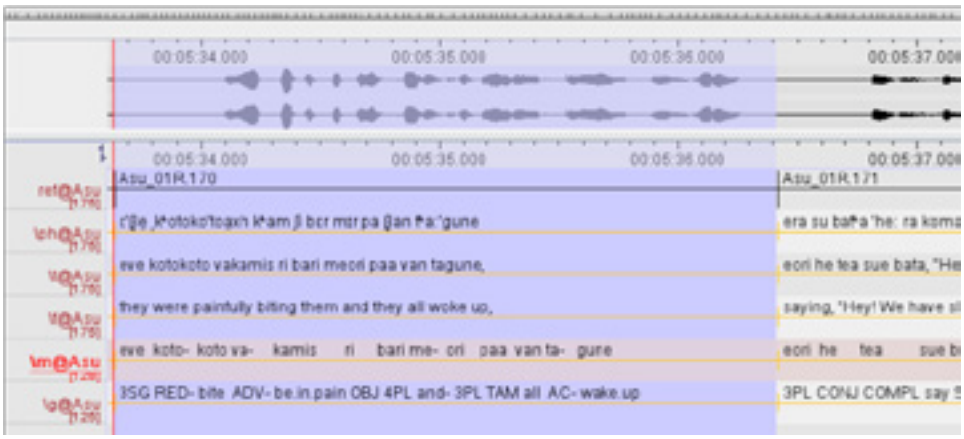


FIGURE 1. Annotation of a Teop audio file in ELAN

ELAN allows a text with various kinds of annotation on separate tiers to be presented, as illustrated in Fig. 1, which shows a narrow phonetic and an orthographic transcription, a free translation, morphological segmentation, and interlinear morpheme glossing. An annotation like the transcription of clauses can be time-aligned to the sound file, whereas other annotations as, for instance, the translation of the clauses are aligned to the corresponding transcription. The annotations can be exported as Toolbox, Praat or text files. For an excellent overview of annotation systems used in documentary linguistics see Schultze-Berndt (2006).

In grammars, individual examples and text samples, which are usually presented in an appendix, are provided in a three-tiered format, a transcription with morphological segmentation, interlinear morpheme glossing, and a free translation in order to show how the meaning of a construction that is rendered in the free translation relates to the constituent parts of the construction. For the grammatical analysis of texts, however, it may be useful to have additional tiers that provide information on the constituent structure of phrases and clauses and thus facilitate the exploration of syntactic structures and the interface of discourse and syntax. Such a system called GRAID (Grammatical Relations and Animacy in Discourse) has been recently developed by Haig & Schnell (2011) and tested for five genetically and typologically divergent languages (Haig, Schnell & Wegener 2011).

2.2. CORPUS-BASED GRAMMATICAL ANALYSIS AND DESCRIPTION. What makes ELAN most suitable for the grammatical analysis of a text corpus is that it facilitates complex searches with the query language Regular Expressions on multiple tiers.

With Regular Expressions you can search not only for all occurrences of a particular word, but also for discontinuous sequences of particular words or bound morphemes, for two or more alternative expressions at the same time, for a particular expression with the exclusion of other expressions, and even for reduplicated word forms.

A typical example for a complex construction in Teop that can easily be searched for with Regular Expressions is the negation of predicates which is expressed by the discontinuous morpheme *saka*/*sa* ... *haa*-. The first part has the variants *saka* and *sa*, the second part may stand by itself or have a suffix.

The search finds 338 tokens in the current Teop corpus of 258,866 words and presents them in a concordance. As Fig. 2 shows, the construction *saka*/*sa* ... *haa*- accommodates nouns, verbs and adjectives. Further searches show that with 229 tokens the first element *sa* is much more frequent than *saka*, and that contrary to our expectations both forms are equally distributed in spontaneous spoken and edited written text varieties.

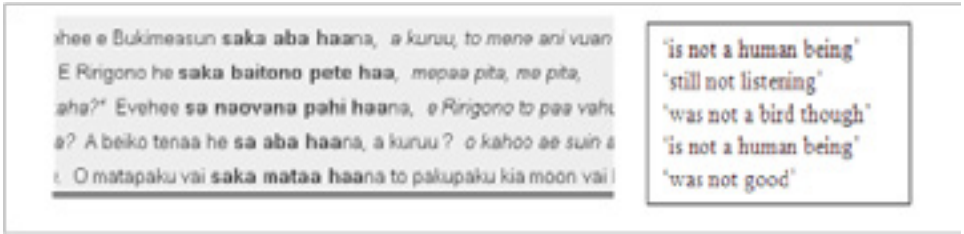


FIGURE 2. Concordance for the negation of predicates

Other typical examples of searches for discontinuous elements include the search for patterns of word-formation by prefixes and suffixes as in English *un*...*able* or the collocation of particular verb forms with temporal adverbs, which is useful for the investigation of aspect and Aktionsart (Van Valin 2005:32-42).

The so-called multilayer search enables you to search on more than one tier. For example you can search for one sense of a polysemous or homonymous lexical item by searching simultaneously on the transcription and the translation or glossing tier. By using the wild card * for the translation tier, the concordance shows the searched Teop items in context on the left-hand side and the corresponding translation on the right-hand side (Fig. 3).

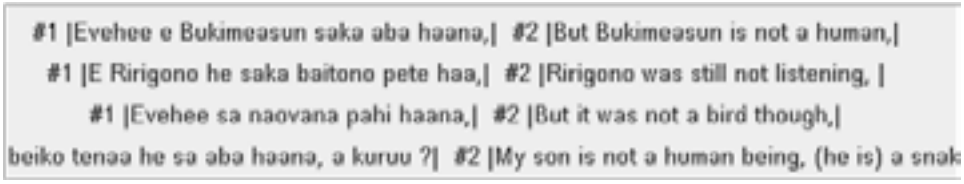


FIGURE 3. Multiple tier searches

Working with a digital corpus that is linked to sound files and facilitates complex searches has a number of advantages over traditional text analysis and thus ultimately leads to more reliable grammars. When browsing texts in search of examples for a particular construction, your attention may be biased; you only note down what seems interesting at the moment, because you cannot make notes for all examples. In contrast, the corpus search gives you all tokens in a concordance and enables you to make explicit statements about the frequency of constructions and their distribution in various text varieties. Since the concordance is linked to the corpus, you can jump from each token to the text file with one mouse click and immediately check the wider context of the token, listen to the sound file and check its annotation.

For the writing of a grammar the corpus-based approach means that linguistic phenomena are described with reference to their context in natural language use and that their frequency can be stated. So instead of giving the vague information that a linguistic phenomenon is rare or more frequent than another one, the grammarian can give exact figures of the number of occurrences in the corpus or a subcorpus, including the search method and the date of access in case the corpus in growing. In Teop, for example, nouns like *aba* 'person, human being' may function not only as the head of a NP, but also as the head of a verb complex (VC) as in

(1) *E* *Magaru* *kou* *na* *aba* *vakis* *nana*
ART Earthquake PART TAM person still IPFV:3SG
te- *a* *taem* *vai*.
PREP ART time DEM
'Earthquake was still a human being at that time.' (Val_02R.31)

But only concrete figures of the functions of prototypical nouns will show how rare the use of nouns as VC heads is (see Table 1).The figures as such are not explanatory, they only show patterns of language use that need to be interpreted. In the case of the distribution of prototypical nouns in the position of VC and NP heads, a probable explanation is that these nouns denote entities that in most contexts are conceptualized as time-stable (cf. Givón 2001:51), whereas the use of a word as the head of a VC implies a change over time.

		VC head	**NP head**
aba	'person'	8	370
beiko	'child'	1	504
moon	'woman'	9	767
naono	'tree'	-	232

TABLE 1. The distribution of prototypical nouns in the Teop Language Corpus (31.12.2011)

2.3. THE ROLE OF METADATA IN GRAMMARS. Metadata are data about data. In the context of language documentation metadata can be classified into

- collection-level metadata giving information on the circumstances of data collection such as the scope and duration of the project and the equipment used for recording;
- item-level metadata giving information on the name of the language, the recording date, the collector and the speaker, the kind of media, the content and the size of the recording;
- biographical information about each participant of the recording session (for details see Conathan 2011:246-248).

In the grammar the information on the circumstances of the data collection and biographical data of the speakers and the people who did the recordings will be given in the introductory chapter, whereas an appendix may contain a list of the names of the primary data and their specific item-level metadata. In addition, each example taken from the corpus should get a label which indicates its source and some information on the text variety in the form of abbreviations. Nordhoff (2009), for example, uses labels that indicate the town and the date of the recording and the text variety, e.g. 'nar' for narrative, 'cvs' for conversation, 'sng' for song etc.

To date many corpus-based grammars of previously undescribed languages do not give any detailed information on the content and structure of the corpus, let alone references or metadata for the examples. The readers of these grammars are not informed whether a particular example has been elicited or comes from a legend, a procedural text, ritual or a certain genre, who the speaker was, and when and under which circumstances the recording was done. Since all languages show variation and grammars never capture the full range of variation, the grammar user should be informed about the text varieties that served as the basis of the grammatical analysis and description.

2.4. ACCESSIBILITY OF THE CORPUS. From a scientific point of view, it should be a matter of course to grant access to the corpus that served as the source for the grammatical analysis because otherwise the grammatical description could not be scrutinized by other researchers (Himmelmann 1998:165). Since grammarians may be misguided by their hypotheses and overlook examples that would not fit their hypotheses, the results of their grammatical analysis remain preliminary as long as they are not replicable by other researchers (for a discussion of replicability in corpus linguistics see McEnery & Hardie 2012:14-16). Consequently, the grammar should either contain a DVD with the corpus or inform the readers how they can access the corpus via the internet (cf. Thieberger 2006, Nordhoff 2009).

3. BUILDING UP A GRAMMAR WRITER'S CORPUS. Linguists who intend to write a grammar of a previously undescribed language will use a variety of field methods to collect data (Bowern 2008, Chelliah & De Reuse 2011, Mosel 2006, Mosel 2012), and sooner or later start collecting texts. What kind of texts are recorded and in what format they are published depends in the first place on the speech community's interests and values

(Woodbury 2011:180-182). The request for recording "a variety of informal communicative events ... to give an authentic impression of the language" (Seifart 2008:69) cannot always be responded to. Haig, Nau, Schnell & Wegener (2011:4) observe that in most projects of the Dokumentation Bedrohter Sprachen (DoBeS) program, "it still tends to be more traditional monologues than everyday conversational interactions that find themselves as fully-annotated records in the archive".

The speech community's right to set the agenda can mean that linguists who intend to collect texts as their data basis for a grammatical description might have to change their plans and adapt their project to the decisions of the speech community. If, for example, the speech community is only interested in documenting their traditional oral literature, then obviously the grammar that is based on these texts only represents this text variety, and from the point of view of grammaticography there is nothing wrong with this. It goes, however, without saying that the more diverse text varieties the corpus contains, the greater are the chances that the grammar can comprehensively represent the language (Foley 2003:95).

The speech community may also, as it happened in the Teop language documentation project, insist on editing the texts before publication, which clearly contradicts the aim of language documentation to record "the linguistic practices and traditions of a speech community" (Himmelmann 1998:166). As will be shown below, the editing of texts will provide the grammaticographer with a new quality of data because it shows what native speakers are actually doing when transforming speech to writing and thus not only contributes to the analysis of the particular language in question, but also to research on the differences between spoken and written language. The drawback of editorial work, however, is that it is time consuming and requires a well organized workflow as, for example, the work of the Dauenhauers on the documentation of Alaskean Tlingit oral culture shows (Dauenhauer & Dauenhauer 1996). A brief description of my own experiences is given in Mosel (2006:82f).

In order to minimize the danger that editorial work in our Teop project was influenced by the native speakers' knowledge of English text varieties, they were advised to keep the original speakers' way of expression, their phraseology and discourse structure, and thus avoid the dangers of westernizing traditional oral literature. Each edited text was independently checked by at least two other native speakers. Both the edited texts and the original recordings are archived in the DoBeS Archive, but the original recordings with their transcriptions and translations are only accessible under the condition that the users register themselves. A comparison of the spoken and edited text varieties shows that in spite of my advice the editors made quite a number of changes (see §4). The edited version has been printed without translations and is now used in schools (Magum et al. 2007) and is also available in the DoBeS archive.

After they had done transcriptions and editorial work during several fieldwork seasons, some Teop research assistants started writing example sentences for the grammar and the dictionary, stories, and descriptions of animals, plants, artifacts, and everyday activities. These are definitely not traditional, but innovative text varieties. But this does not mean that they are less authentic than, for example, spoken legends or conversations, as long as the linguist does not teach the native speakers what in his or her view a good story is. Furthermore, when speech communities want their language to become a written language and the means of instruction in primary schools, it certainly belongs to the responsibilities

of linguists to help them create it by keeping the uniqueness of their language, but also avoiding a rigid purism that would put off younger speakers. Linguistically these new text varieties are interesting because they allow us to observe the process of putting a previously unwritten language into writing.

In all publications on the Teop language, the references for examples indicate whether the example is taken from the recordings of spontaneous speech (abbr. R), from an edited version (abbr. E) or from a written text that has not been derived from a transcription (abbr. W). In the grammar I try to present wherever possible examples from spoken and edited or written text varieties for each grammatical phenomenon.

In the reminder of this chapter, I show how useful even a relatively small but diversified corpus can be and what special kinds of grammatical constructions are offered by different text varieties.

4. VARIATION IN THE GRAMMAR OF ORAL LEGENDS AND THEIR EDITED VERSIONS. When we analyzed the two subcorpora of spoken and edited Teop legends, which comprise 31,909 and 31,294 words, respectively, we could identify four types of syntactic changes in the edited versions: elaboration, linkage of paratactic clauses, compression of paratactic clauses, and decompression of complex constructions (Mosel 2008).

All constructions found in the edited versions are also found in the oral versions, but the two registers differ in the frequency of certain constructions:

- In the edited versions, the replacement of paratactic constructions by compressed constructions is more frequent than the reverse kind of replacement.
- Elaboration often results in complex structures (e.g. adjectival attributes, serial verb constructions, relative clauses, clausal adjuncts).
- The edited versions make more use of explicit clause linkage.

Table 2 gives a summary of the observed changes in edited narratives, which on the whole result in more complex structures.

Strategy	**Syntactic change**
Elaboration	addition of linguistic units (words, phrases, clauses)
Linkage of paratactic clauses	1. linkage by cross-clausal dependency without embedding (chained Tail-Head-Linkage, adjoined adverbial clauses) 2. integration by embedding (relative and adverbial clause constructions) 3. interlacing by raising in complement constructions
Compression of paratactic clauses	1. serial verb constructions 2. nominalizations 3. ditransitive constructions
Decompression	resolution of complex constructions into paratactic constructions

TABLE 2. Syntactic changes in edited narratives

The following citations from an oral legend (2) and its edited counterpart (3) illustrate the replacement of two coordinate clauses by a so-called Tail-Head construction and a few other changes.

(2) *Me-* *paa* *vahuhu* *bona* *taonim* *a* *si* *iana*.
and- TAM give.birth.to ART five ART DIM fish
'And gave birth to five little fish.'

Me- *a* *taonim* *a* *si* *iana* *bona* *vue*
and ART five ART DIM fish DEM particular
'And these five little fish'

na *vaatii* *roho* *e* *te-* *a* *boon ...*
TAM put first 3SG PREP- ART mangroves ...
'she put in the mangroves.' (Ata_01R.01)

(3) *Me-* *paa* *vahuhu* *bona* *taonim* *a* *si* *iana*.
and TAM give.birth.to ART five ART DIM fish
'And gave birth to five little fish.'

Vaahuhu *vakavara* *vai* *ri* *bari*
give.birth finish then 3PL.OBJM 4PL
'Having given birth to them,'

me *paa* *varavihi* *ri* *bari* *koma-* *n-*
and TAM hide 3PL.OBJM 4PL inside- 3SG.POSS-
a *boon*
ART mangroves
'hid them inside the mangroves.' (Ata_01CE1.01)

In the spoken version (2) the second clause is joined to the first one by the conjunction *me* 'and', the repetition of 'five little fish' and the anaphoric demonstrative *bona* , whereas in the edited version the verb of the first clause (the tail) *vahuhu* 'give birth' is repeated at the beginning (the head) of the second clause. While the clause linkage in the spoken version is similar to an English type of clause linkage, the one of the edited version is not. Furthermore, the editor exchanged the verb *vaatii* 'put' for the semantically more specific verb *varavihi* 'hide', inserted an object marker and the pronoun *bari* 'them' and replaced the multipurpose preposition *te* PREP by the more elaborate locative construction *koma na boon* 'inside the mangroves'.

Since the Tail-Head construction is typical for Oceanic and Papuan languages, but is not found in English, we were interested in how the editors treated these constructions and counted all Tail-Head constructions in which the Head is modified by *vakavara* 'finish' as in (3). The result was that with 51 tokens the Tail-Head construction is much more frequently used in the edited legends than in the original spoken versions which only show 28 occurrences. Thus we have the impression that the edited version represents a more conservative style of story telling than the original spoken version.

5. GRAMMATICAL VARIATION ACROSS TEXT VARIETIES. The Teop Language Corpus comprises several subcorpora which on the basis of their content and the circumstances of their production can be classified as shown in Table 3. Not unexpectedly, these text varieties do not only differ in their vocabulary, but also in their preferences for certain syntactic constructions, as the remainder of this section will illustrate with examples from legends, dictionary definitions and procedural texts.

Genres	**Themes**	**Production**
legends	fights with giants and witches, bad treatment of children by their stepmothers, controversies between two brothers, origin of natural phenomena and artifacts	spoken and edited; some only written
personal narratives	autobiographies, survival during the Second World War, travel	spoken and edited; two only written
encyclopedic descriptions	plants, animals (mammals, birds, reptiles, fishes, crabs, shells), house and canoe building, fishing, butchering, cooking, cultural practices	descriptions of things only written; procedural texts spoken, edited and written
interviews	young native speakers interviewing elders about customs and the Second World War	spoken and edited
example sentences	not specified	only written

TABLE 3. Text varieties in the Teop Language Corpus

5.1. LEGENDS. Since legends are situated in imaginary worlds where animals can talk or magic allows transformations of things into living beings or living beings into things, they may offer interesting data on the functional flexibility of lexemes (see §2.2) and noun classification. In Teop, for example, animal names belong to the unmarked class of common nouns, but move to the class of personal proper names when referring to one of the protagonists of a legend. This phenomenon, which shows that under certain conditions the classification of nouns is variable, would not be attested and, consequently, not described in the grammar if our corpus did not contain legends.

For the description of argument structure and discourse pragmatics, the beginnings of legends provide easily retrievable data on how new participants are introduced into the discourse. Furthermore, legends may contain direct speech with colloquial expressions of surprise and anger, which are interesting for the description of phraseology and the grammar of interjections, but are difficult to record otherwise (Seifart 2008:73).

5.2 THE GRAMMAR OF DICTIONARY ENTRIES. Since it is impossible to produce a dictionary within a short-term language documentation project, we decided to compile a series of small thematically specialized dictionaries on plants, fishes, house building, cooking, etc. These mini-dictionaries (MD) contain short encyclopedic articles in Teop with an English translation (Mosel 2011, Mahaka et al. 2010). In addition, the dictionaries of the material culture are supplemented by procedural texts which describe selected traditional techniques like thatching the roof of a house, making fishing nets, butchering a pig, etc. Both the definitions and the procedural texts are valuable sources for gathering grammatical data, because they contain some constructions at a much higher rate than narrative texts.

Since the purpose of a dictionary entry is to define the meaning of a word, the entries show a variety of topic constructions that are not encountered in narratives in this density. The definitions of nouns frequently start with a non-verbal clause consisting of a topical subject NP followed by a classifying predicative NP that is modified by an adjectival phrase (AP) or a relative clause:

(4)

SUBJ.NP		PRED.NP		QUALIFICATIVE ATTRIBUTIVE AP	
A	*bokua*	*a*	*iana*	*a*	*beera ...*
ART	bokua	ART	fish	ART	big, ...

'The bokua is a big fish.' (Vaa_09W.068)

(5)

SUBJ.NP		PRED.NP		POSSESSIVE ATTRIBUTIVE AP		
A	*havanao*	*a*	*iana*	*a*	*kapa*	*kikis*.
ART	havanao	ART	fish	ART	skin	strong

'The havanao is a fish with a strong skin.' (lit. '(having) a strong skin') (Sii_11W.039)

(6) SUBJ.NP PRED.NP RELATIVE CLAUSE

O	*poka*	*o*	*hum*	*to*	*vavaobetera-*	*ra-*
shelf	ART	place	REL	put	ART	1PL.INCL.IPFV-
ara		*bona*		*maa*	*taba*.	
1PL.INCL		ART		PL	thing	

'The shelf is a place where we put things.' (MD House, poka)

While the definitions of nouns supply excellent examples for topicalization, non-verbal clauses, adjectival phrases, relative clauses, and the expression of habitual activities (4-6), the definitions of verbs are a good source for nominalizations and complement clauses in predicative function:

(7)

A	*siri*	*atovo*		*ei*	*be-*	*ara*	*gono*
ART	tear	sago.palm.leaf		DEM	when-	1PL.INCL	get
kahi	*o*	*paka*	*bono*	*sikiri*	*na-*	*e*.	
from	ART	leaf	ART	midrib	3SG.POSS	3SG	

'The tearing of the sago palm leaf, this (is) when we remove the midrib from the leaf.' (MD House, siri atovo)

To conclude, although a grammar writer's task is not collecting data for a dictionary, it seems worthwhile asking native speakers to formulate some definitions of animal and plant species, artifacts, and special activities.

5.3. PROCEDURAL TEXTS VS. NARRATIVES. Similar to dictionary entries, procedural texts are not an indigenous, conventionalized genre in Pacific cultures, as people prefer to demonstrate how this or that is done instead of describing it (Mosel 2006:73f). Consequently, the speakers have not yet developed conventionalized ways of describing procedures and seem to be free in their choice of pronouns to refer to generic agents. Some prefer the second person singular, others the first person inclusive plural or the third person plural pronoun. One speaker consistently uses the first person exclusive plural (cf. 13), which the editors of her texts always replace by the first person inclusive pronoun. This variation in the use of pronouns for generic agents is remarkable and needs to be mentioned in the grammar chapter on pronouns.

Another remarkable feature of the procedural texts is that all speakers and writers use the same kind of clause linkage construction when explicitly referring to the regular fixed order of actions. While in Teop narratives the sequence of events is simply expressed by paratactic and coordinate clauses, or the so-called Tail-Head construction (see §3), the procedural texts show constructions with adverbial clauses. Our first example (8) comes from a legend in which a giant scrapes the bark of *kave* vines for making a fishing net. In the Tail-Head construction the narrator repeats the head of the VC *kahu* 'scrape', but modifies it by *vakavara* 'finished' expressing that this action was finished, before he did the next one, i.e. *taatagi* 'prepare'.

(8) *me-* *ori* *paa* *dee* *voosu* *maa,* *me-* *ori* *paa*
and- 3PL TAM carry home DIR and- 3PL TAM
ma *kahu,*
come scrape
'and they carried (the kave vines) home, and they scraped them'

me- *ori* *kahu* *va-* *kavara* *bona* *kano-* *kanono* *te-*
and- 3PL scrape ADV- finished ART RED- rope PREP-
ori,
3PL
'and they finished scraping their ropes,'

a- *maa* *kara* *kave* *te-* *ori,* *me-* *ori* *paa*
ART- PL string kave PREP- 3PL and- 3PL TAM
taatagi *bari,*
prepare 4PL.OBJ
'their kave strings, and they prepared them.' (Sii_06R.56-60)

The second example (9) comes from a written description of how Teop people made nets for catching turtles in former times. Here the fixed sequence of two actions is expressed by a *be-re* 'when-then' construction, which is very frequent in procedural texts.

(9) *Be-* *ve* *obete* *nana* *te-* *o* *kasuana,*
when- 3SG lie 3SG.IPFV PREP- ART ground
'When it is lying on the ground,'

eara *re-* *paa* *kahu* *a* *kapa* *nae*
1PL.INCL then- TAM scrape ART bark 3SG.POSS

bono *kehaa*
ART shell
'then we scrape its bark off with a shell'

to *dao* *ra-* *ara* *bono* *sui.*
REL call 1PL.INCL.IPFV- 1PL.INC ART sui
'that we call sui.'

Be- *ara* *kahu* *vaka-* *va-* *kavara* *e,*
when- 1PL.INC scrape RED- ADV- finished 3SG
'When we have finished scraping it,'

eara *re* *paa* *vaaroava* *e* *bono* *buaku*
1PL.INCL then TAM dry.in.sun 3SG ART two

ge	*o*	*kukan*	*o*	*bon*.
or	ART	three	ART	day

'then we put it into the sun for two or three days.' (Eno_08W.4-6)

Other variants of this construction in procedural texts include:

(10)

be-	AGENT	X	*va-*	*kavara,*	AGENT	*re-*
when	AGENT	X	ADV-	finished	AGENT	then-

paa	Y
TAM	Y

'when AGENT has finished doing X, then AGENT does Y'

(11)

be	*kavara,*	AGENT	*re-*	*paa*	X
when	finished	AGENT	then-	TAM	X

'when it is finished, then AGENT does X'

(12)

be-	AGENT	*tau*	X,	*AGENT*	*re-*	*paa*	Y
when	AGENT	about to	X,	AGENT	then-	TAM	Y

'when AGENT is about to X, then AGENT does Y'

(13)

be-	AGENT	*mei*	*tea*	X,	AGENT	*toro*	Y
when	AGENT	not.yet	COMP	X	AGENT	must	Y

'before AGENT X, AGENT must do Y'
(lit. 'when AGENT has not yet X, AGENT must Y')

In order to get further evidence for the difference in clause linkage in narratives and procedural texts, I bought a rooster from a neighbor and asked him to butcher it while I was taking a series of photographs. Luckily his four years old twins were helping him butcher the rooster, while his wife was watching, so that three months later I could ask her to look at the photographs and narrate the story of how her husband and her children butchered a rooster during my last visit. In addition, I asked another woman to have a look at the photographs and describe how Teop people butcher a rooster.

While in the procedural text nine clauses out of a total 40 clauses are adverbial clauses introduced by *be* 'when' (14), the narrative text, which consists of 53 clauses, has none of these constructions, but uses paratactic clauses instead (15):

(14) Procedural text

Be	*kavara,*
when	finished

'When it is finished,'

be-	*nam*	*pee-*	*pee*	*va-*	*ruta-*	*rutaa*	*va-*
when-	1PL.EXCL	RED-	cut	ADV-	RED-	small	ADV-

kavara	*eve,*
finished	3SG

'when we have finished cutting it into small pieces,'

o-	*re*	*paa*	*vahio*	*bari*	*te-*	*o*	*suraa*.
3PL-	then-	TAM	put	4PL	PREP-	ART	fire.

'they put it onto the fire.' (Hel_13R.33-34)

(15) Narrative text

Eove	*he*	*kaku*	*va-*	*kavara*	*bene*	*toa*
3SG	but	butchered	ADV-	finished	ART	chicken

'But he finished butchering the rooster,'

me-	*ori*	*paa*	*vaa-*	*tei*	*bari*	*te-*	*a*
and-	3PL	TAM	CAUS-	be	4SG/PL	PREP-	ART

sosopene.
saucepan
'and they put it into the saucepan.' (Pau_01R.51-52)

As the preceding examples illustrate, the distinction between Tail-Head constructions and the adverbial clauses is most clearly shown by comparing narrative and procedural texts of the same or closely related contents as, for instance, net making (8,9) or butchering a chicken (14,15). Consequently, it seems practical to include this kind of comparison in a grammar.

6. DIFFERENT THEMES - DIFFERENT GRAMMATICAL PHENOMENA. People talk about different themes in different ways. For the collection of grammatical data this means that some themes will provide more and better data for certain grammatical phenomena than others. Thus inanimate topics are certainly better represented in descriptions of how certain artifacts are manufactured than in autobiographies, whereas ditransitive constructions with agents, recipients and themes are most likely to be found in texts about trading and ceremonial exchanges of food and valuables.

6.1. TROPICAL FISHES ARE COLORFUL. The question of whether in Oceanic languages lexemes denoting properties form a word class in its own right, i.e. adjectives, or are better classified as a subclass of verbs is probably as old as Oceanic linguistics itself, but a thorough corpus-based study of property words in any of these languages is still missing. A distributional analysis of dimensional and evaluative adjectives such as *beera* 'big' and *mataa* 'good' in Teop shows that they often occur as the head of VCs, but that they are distinct from intransitive verbs as they never occur as the head of NPs, which all intransitive verbs do, e.g. *a pita* 'the walking', *a mate* 'the dying, death', but not **a beera* 'the being big' or **a mataa* 'the being good'. Secondly, these adjectives differ from intransitive verbs in that they must take the prefix *va-* when modifying a verb, e.g. *vabeera* 'to a great extent', *vamataa* 'well, properly'.

lexeme	**VC head**	**AP head**
beera 'big'	96	295
mataa 'good'	83	133

TABLE 4. Distribution of *beera* 'good' and *mataa* 'good'.

For color words we did not have comparable data until we started compiling a small fish dictionary in which the fish names are defined by descriptions in Teop with English translations. Most descriptions contain color words and clearly show that in Teop color words behave exactly like dimensional and evaluative adjectives. They enter into the comparative construction (cf. 16 and 17) and are transformed into an adverb by the prefix *va-* when modifying a verb, e.g. *tara vamataa* 'look good'.

lexeme		**VC head**	**AP head**
gogooravi	'red'	12	23
kakaavo	'white'	7	25
paru	'black'	5	34

TABLE 5. Distribution of three colour words in the fish dictionary.

(16)

	NP		AP		VC		
evehee	*a*	*toobono*	*a*	*beera,*	*[na*	*beera*	*oha*
but	ART	toobono	ART	big	TAM	big	pass
		NP					
nana]		*bona*		*pasupua*			
3SG.IPFV		ART		pasupua			

'(The toobono looks like the genuine pasupua,)
but the toobono is big, is bigger than the pasupua.' (MD Fishes, toobono)

(17)

NP		AP	predicate		
A	*aranavi*	*[a*	*gogooravi*	*vasihum] ...*	
ART	aranavi	ART	red	a.bit	

'The aranavi is a bit red...'

NP		VC				NP
A	*sinarona*	*[na*	*gogooravi*	*oha*	*nana]*	*bona*
ART	sinarona	TAM	red	pass	3SG.IPFV	ART
aranavi.						
aranavi						

'The sinarona is redder than the aranavi.' (MD Fishes, aranavi)

Similar to *tara vamataa* 'look good', we find derived colour adverbs modifying *tara* 'look':

(18) *Be-* *ori* *hovo* *ruene* *o-* *re* *paa* *tara* *va-*
when- 3PL enter river 3PL- then TAM look ADV-
paru.
black
'(While they are still staying in the ocean, they look white.)
When they enter the rivers, they look black.' (MD Fishes, ovunaa)

As far as we can judge from our limited set of data in Tables 3 and 4, adjectives occur more often as the head of an AP than as the head of a VC, but the difference between these figures is not as marked as those of the distribution of nouns as NP and VC heads. A possible explanation for these findings may be that these adjectives denote less time-stable concepts than nouns.

The preceding examples illustrate that frequency analyses can be helpful in formulating hypotheses about the interaction of lexis and grammar. Munro (2007:72) stresses the importance of dictionary work for grammatical analysis, "Making dictionaries helps in grammatical analysis, and in fact in the absence of dictionary work a grammatical description is very likely to miss important things".

6.2 WHAT TREES ARE GOOD FOR. The Teop language is a verb second language. This means that the verb complex always occurs in the second position of the clause, while the first position is held by the topic of the clause, which can be the subject, a primary object, a secondary object, or an adjunct. If the topic can be recovered from the preceding context, the topic position can be left empty. With ditransitive verbs, Teop shows the following clause patterns:

TOPIC	**VC**	**Argument**	**Argument**
SUBJ (subject)	VC	OBJ1 (primary object)	OBJ2 (secondary object)
OBJ1 (primary object)	VC	SUBJ (subject)	OBJ2 (secondary object)
OBJ2 (secondary object)	VC	SUBJ (subject)	OBJ1 (primary object)

TABLE 6. Clause patterns.

Teop does not have a passive construction. If the agent of an action is not identifiable, the third person plural pronoun functions as a non-topical subject.

The 2007 version of Teop Language Corpus gives the impression that constructions with the subject in the first position represent the dominant word order. For the ditransitive verb *hee*, for example, we find the following frequencies of clause patterns (Mosel 2007, 2010):

clause patterns	**frequency**
SUBJ VC OBJ1 OBJ2	25
OBJ1 VC SUBJ OBJ2	6
OBJ2 VC SUBJ OBJ1	4

TABLE 7. Clause patterns of *hee* 'give' (Sept. 2007).

With *hee* 'give', the primary object (OBJ1) refers to the recipient and the secondary object (OBJ2) to the theme. Other ditransitive verbs like *nahu* 'cook' govern a primary object referring to the patient and an optional secondary object referring to the instrument:

(19) SUBJ:agent VC OBJ1:patient OBJ2:instrument
... a-re *ma* *nahu* *a* *guu* *vai* *bona* *tahii*.
1PL.INCL-then come cook ART pig this ART saltwater
'(You must fetch some saltwater) so that we can cook this pig with saltwater.' (Mat. 1.68R)

When analyzing clauses of this kind, I had the impression again that the dominant, unmarked order was SUBJ VC OBJ1 OBJ2. But when the Teop research assistants collected descriptions of trees and what the parts of trees are used for, I realized that it would only make sense to speak of a dominant word order with respect to a particular text variety. If as in the tree descriptions the topic of discourse is a patient or instrument, the noun phrases denoting these roles function as objects, but occupy the first position of the clause, as the following dictionary entry for *asita* 'putty nut tree' nicely illustrates. The entry starts with the sentence:

(20) OBJ2 VC SUBJ OBJ1
O *asita* *[na* *asi-* *asita* *ri-* *]* *ori* *bono*
ART putty.nut TAM RED- plaster 3PL.IPFV 3PL ART
sinivi.
canoe
'The putty-nut tree, they use it for plastering the canoe.' (i.e. the nuts of the tree) (MD Plants, asita)

In the second clause of the entry (21), the topic position is empty. The topic is still *asita* in the function of a secondary object, but as it is easily recoverable from the context, it does not need to be mentioned.

(21) VC SUBJ OBJ1
[Na *asita* *ri-* *]* *ori* *[bona* *maa* *panapana]*
[TAM plaster 3PL.IPFV] 3PL ART PL knotholes
'They plaster the knotholes (of the canoe with it).' (MD Trees, asita)

This sentence is then followed by two other sentences of the same structure, while the last sentence shows a construction in which the valency of a ditransitive verb—here *porete* 'treat s.o. with s.th. (some kind of traditional ditransitive verb—is reduced by the particle *ni,* resulting in a transitive construction meaning 'use s.th. as traditional medicine' (Mosel 2010:493).

(22) OBJ VC

Asita	*me*	*[na*	*pore-*	*porete*	*ni*	*ri]-*
plaster	also	TAM	RED-	make.medicine	APPL	3PL.IPFV
SUBJ						
ori.						
3PL						

'Asita is also used for making medicine.' (MD Plants, asita)

8. CONCLUDING REMARKS. The present chapter suggests that the grammaticography of previous undescribed languages can profit from an approach that combines language documentation with corpus linguistic methods. In contrast to traditional grammar writing, the corpus linguistic approach accounts for language internal variation in relation to text varieties. As Conrad (2010:228) puts it, "corpus analyses lead us to describing grammar not just in structural terms, but in probabilistic terms—describing the typical social and discourse circumstances associated with the use of particular grammatical features". The modern technology of corpus linguistics allows us to systematically search for particular lexical items and their collocations as well as for constructional patterns and the lexical items they accommodate, to view all findings in a concordance and to analyze the grammatical structures in their natural context.

This chapter emphasizes the need for a diversified corpus and shows what kind of data is provided by different text varieties. In particular we examined spontaneously spoken and edited versions of legends, procedural texts and dictionary definitions, and discovered that due to their different contents and discourse structure these text varieties provide useful data for various grammatical phenomena:

1. The comparison of oral and edited legends shows what kind of constructions native speakers regard as synonymous, in particular variation in narrative clause linkage.
2. Comparable narrative and procedural texts about the very same topic show how the contrast between specific and habitual sequential actions is expressed.
3. Monolingual dictionary definitions of nouns provide data of how the classification of living beings and things is expressed, which in the case of Teop involves non-verbal predicates, various kinds of adjectival attributes and relative clauses. The definitions of verbs typically contain nominalizations in subject position and complement clauses as predicates.
4. In the descriptions of trees and their parts we find numerous examples for constructions with inanimate topics and the expression of the semantic role of instrument.

The macro-structure of a corpus-based grammar may follow the traditional ascending model starting with a chapter on phonology and concluding with a chapter on complex sentences (Mosel 2006b), but its content would probably differ in the following aspects:

1) The introductory chapter would provide explicit information on
   a) fieldwork methods (cf. §3),
   b) the sociolinguistic profile of the speech community,
   c) the sociolinguistic background of those native speakers who were recorded or otherwise involved in the project (cf. §2.3), and
   d) the genres (§5), the topics (§6) and the size of the texts as well as the technology of recordings and the annotation methods (cf. §2.1)
2) In addition, the appendices of the grammar may supply detailed information on the individual texts and speakers in the form of tables (cf. §2.3).
3) Within the chapters the description of grammatical phenomena would account for variation in linguistic form and function and, wherever it seems reasonable and significant, make statements about preferred structures in terms of frequencies. This may, for example, include
   a) the syntactic distribution of words or word classes (cf. Table 1, 4, 5),
   b) the frequency of clause patterns (cf. Table 7), or
   c) the occurrence of particular constructions in certain text types (cf. §5.3).
4) The examples would get labels that inform the reader on their origin and facilitate their identification in the corpus, which ideally is easily accessible.

In the near future digital linguistics will develop electronic formats of grammars and new tools assisting in grammatical analysis (Evans & Dench 2006:28-30, Nordhoff (ed.) 2012), but the arguments for a corpus based grammaticography as outlined in this chapter will certainly not lose their validity.

References

Austin, Peter. 2006. 'Data and language documentation,' in Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.) *Essentials of language documentation*. Berlin, New York: Mouton de Gruyter, 87-112.

Biber, Douglas & Stig Johansson & Geoffrey Leech & Susan Conrad & Edward Finegan. 1999. *Longman Grammar of spoken and written English*. Harlow, Essex.

Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge:CUP.

Bouquiaux, Luc and Jacqueline Thomas (eds.). 1992. *Studying and describing unwritten languages*. Dallas: SIL

Bowern, Claire. 2008. *Linguistic fieldwork. A practical guide*. New York: Palgrave Macmillan.

Bright, William. 2007. 'Contextualizing a grammar', in Thomas E. Payne & David J. Weber (eds.). *Perspectives on grammar writing*. Amsterdam: John Bejamins Publishing Company, 11-17.

Chelliah, Shobhana L. 2001. 'The role of text collection and elicitation in linguistic fieldwork', in Paul Newman and Martha Ratliff (eds.). *Linguistic Fieldwork*. Cambridge: Cambrige University Press, 152-164.

Chelliah, Shobhana L. & Willem de Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht etc.: Springer.

Conathan, Lisa. 2011. Archiving and language documentation. In Peter Austin & Julia Sallabank (eds.). *The handbook of endangered languages*. Cambridge: Cambridge University Press, 235-254.

Conrad, Susan. 2010. What can a corpus tell us about grammar? In Anne O'Keeffe & Michael McCarthy (eds.). *The Routledge handbook of corpus linguistics*. Abingdon: Routledge.

Crowley, Terry. 2007. *Field linguistics. A beginner's guide*. Oxford: Oxford University Press.

Dauenhauer, Nora & Richard Dauenhauer. 1999. The paradox of talking on the page: Some aspects of the Tlinglit and Haida experience. In Laura J. Murray & Karen Rice (eds.). *Talking on the page. Editing aboriginal oral texts*. Toronto, Buffalo, London: University of Toronto Press, 3-41.

Dixon, R.M.W. 2010. *Basic linguistic theory*. Vol. 1. Methodology. Oxford: Oxford University Press.

Evans, Nicholas & Alan Dench. 2006. Introduction: Catching language. In Felix Ameka, Alan Dench & Nicholas Evans (eds.). *Catching language. The standing challenge of grammar writing*. Berlin, New York: Mouton de Gruyter, 1-39.

Foley, William A. 2003. Genre, register and language documentation in literate and preliterate communities. In Austin, Peter. *Language Documentation and Description*. vol. 1, pp. 85-98.

Givón, Talmy. 2001. *Syntax*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Good, Jeff. 2002. *A gentle introduction to metadata*. Open Language Language Archives Community. www.language-archives.org/documents/gentle-intro.html (31.12.2011)

Haig, Geoffrey & Nicole Nau & Stefan Schnell & Claudia Wegener. 2011. Documenting endangered languages before, during, and after the DoBeS programme. In Geoffrey Haig & Nicole Nau & Stefan Schnell & Claudia Wegener /eds.), *Documenting endangered languages. Achievements and perspectives*,1-14. Berlin: De Gruyter Mouton.

Haig, Geoffrey & Stefan Schnell. 2011. *Annotations using GRAID (Grammatical Relations and Animacy in Discourse). Introduction and guidelines for annotators*. Version 6.0. http://vc.uni-bamberg.de/moodle/course/view.php?id=9488 (31.12.2011)

Haig, Geoffrey & Stefan Schnell & Claudia Wegener. 2011. Comparing corpora from endangered languages projects: Explorations in language typology based on original texts. In Geoffrey Haig & Nicole Nau & Stefan Schnell & Claudia Wegener /eds.), *Documenting endangered languages. Achievements and perspectives*,1-14. Berlin: De Gruyter Mouton, 55-86.

Himmelmann, Nikolaus. 1998. Documentary and descriptive linguistics. In *Linguistics* 36, 161-195.

Kalton, Graham. 1983. *Introduction to survey sampling*. Beverly Hills, CA: Sage.

Lynch, John & Malcolm Ross & Terry Crowley. 2002 *The Oceanic languages*. Richmond: Curzon.

Magum, Enoch Horai & Joyce Maion & Jubilie Kamai, & Ondria Tavagaga, & Ulrike

Mosel & Yvonne Thiesen, (eds) 2007. *Amaa vahutate vaa Teapu. Teop Legends*. Kiel: CAU, Seminar für Allgemeine und Vergleichende Sprachwissenschaft, and www.mpi.nl/DOBES/projects/teop (31.12.2011)

Mahaka, Mark & Enoch Horai Magum & Joyce Maion & Naphtaly Maion & Ruth Siimaa Rigamu & Ruth Saovana Spriggs, & Jeremiah Vaabero & Ulrike Mosel & Marcia Schwartz & Yvonne Thiesen. 2010- *A inu. The Teop-English dictionary of house building*. Kiel: CAU, Seminar für Allgemeine und Vergleichende Sprachwissenschaft, and www.mpi.nl/DOBES/projects/teop (31.12.2011)

McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.

Meyer, Charles. 2002. *English corpus linguistics. An introduction*. Cambridge: Cambridge University Press.

Mosel, Ulrike. 2006a. Fieldwork and community language work. In Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.) *Essentials of language documentation*. Berlin, New York: Mouton de Gruyter, 67-85.

Mosel, Ulrike. 2006b. Grammaticography: The art and craft of writing grammars. In Felix Ameka, Alan Dench & Nicholas Evans (eds.). *Catching language. The standing challenge of grammar writing*. Berlin, New York: Mouton de Gruyter, 41-68..

Mosel, Ulrike. 2007. A corpus based approach to valency in a language documentation project. Pre-ALT workshop, Paris, 24.9.2007. http://www.uni-muenster.de/imperia/md/content/allgemeine_sprachwissenschaft/forschen/lingtypolangdoc/handout_mosel_corpus_based.pdf (31.12.2001

Mosel, Ulrike. 2008. 'Putting oral narratives into writing – experiences from a language documentation project in Bouganville, Papua New Guinea'. Paper presented at the Simposio Internacional Contacto de lenguas y documentación, August 2008. Buenos Aires, CAIYT, http://www.linguistik.uni-kiel.de/mosel_publikationen.htm#download

Mosel, Ulrike. 2010. Ditransitive constructions and their alternatives in Teop. In Malchukov, Andrej & Martin Haspelmath & Bernhard Comrie (eds.) *Studies in ditransitive constructions: a comparative handbook*. Berlin, New York: De Gruyter Mouton, 486-509.

Mosel, Ulrike. 2011. Lexicography in endangered language communities. In Peter Austin & Julia Sallabank (eds.). *The handbook of endangered languages*. Cambridge: Cambridge University Press.

Mosel, Ulrike. 2012. Morphosyntactic analysis in the field – a guide to the guides. In Nick Tieberger (ed.) *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press.

Mosel, Ulrike & Even Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Oslo University Press.

Mosel, Ulrike & Enoch Horai Magum, & Jubilie Kamai & Joyce Maion, & Naphtali Maion & Siimaa Ruth Rigamu, & Ruth Saovana Spriggs & Yvonne Thiesen. 2007. *The Teop language Corpus*. www.mpi.nl/DOBES/projects/teop (31.12.2011)

Mosel, Ulrike & Yvonne Thiesen. 2007. *Teop sketch grammar.* www.mpi.nl/DOBES/projects/teop (31.12.2011)

Munro, Pamela. 2007. Form parts of speech to grammar. In: Thomas E. Payne & David Weber (eds.) *Perspectives on grammar writing*. Amsterdam & Philadelphia: Benjamins, pp. 71-111.

Nordhoff, Sebastian. 2009. *A grammar of Upcountry Sri Lanka Malay*.Utrecht: LOT Publications, http://www.lotpublications.nl/publish/articles/003745/bookpart.pdf

Nordhoff, Sebastian (ed.). 2012. *Electronic grammaticography*. Language Documentation & Conservation Special Publication No. 4. Honolulu: University of Hawai'i Press.

Rivierre, Jean Claude. 1992. 'Text Collection', in Luc Bouquiaux and Jacqueline Thomas (eds.). 1992. *Studying and describing unwritten languages*. Dallas: SIL, 56-63.

Samarin, William J. Field linguistics. *A guide to linguistic fieldwork*. New York etc. Holt, Rinehart and Winston.

Seifart, Frank. 2008. On the representativeness of language documentations. In Peter K. Austin (ed.). *Language documentation and description*. Vol. 5. London: School of Oriental and African Studies, 60-76.

Schultze-Berndt, Eva. 2006. Linguistic annotation. In Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.) *Essentials of language documentation*. Berlin, New York: Mouton de Gruyter, 213-251.

Thieberger, Nicholas. 2006. *A grammar of South Efate*. Honolulu: University of Hawai'i Press.

Van Valin, Robert D. 2005. *Exploring the syntax-semantic interface*. Cambridge: Cambridge University Press.

Woodbury, Anthony C. 2011. Language documentation. In Peter Austin & Julia Sallabank (eds.). *The handbook of endangered languages*. Cambridge: Cambridge University Press, 159-186.

umosel@gmx.de