1. Aims
To consider the following general questions 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 the details of 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?

2. Writing a grammar: What is the nature of the task and how should one prepare for it?

2.1 The main steps
Anyone who would produce a grammar must go through at least the following steps or stages: (1) plan the project, (2) gather the data, (3) analyse the data, and (4) produce a written account of the analysis. The focus in this talk will be mainly on stage (4), the final writing up or presentation of the grammar. However, the steps are not simply sequential but cyclical to some extent.

2.2 Planning the project
Many things have to be planned at the outset of the thesis project, beginning with the choice of language. One needs to plan the topic, scope and timetable of the research, the data-gathering operations, 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?
A decision has to be taken at the outset about what kind of grammar to do. 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. The primary readership of a PhD grammar will be professional linguists.

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. Are there circumstances in which it is acceptable to do a grammar of such a language for a dissertation?

A compromise is to do a ‘topics in the grammar of X’ thesis, 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 discuss this matter early on. 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 because he or she objects to the model of grammar you are using. For another thing, the kinds of data to be collected will, to some extent, to be dependent on
the questions that the theoretical framework focuses on.

Specific theories of grammar are notorious for having a short shelf life and even in their heyday are accessible only to specialists. Accordingly, most grammarians of my acquaintance take the view that where possible one should try as far as possible avoid terminology that is local to a particular model of grammar and use instead what R.M.W. Dixon calls “basic linguistic theory” as a source of terms and notational conventions. However BLT is not a fixed thing. The tools of grammatical analysis continue to evolve and one should be ready to take on board new and useful concepts.

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? As a general rule, I think one needs at least nine or ten months in order both to obtain sufficient data for a first grammar of little-documented language and to gain a pretty good level of competence in the language.

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 longer, up to six months, aimed at yielding enough material to draft a substantial part of the thesis. One can then return home, refresh the body and mind, write up first drafts of a large section of the thesis, 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. The situation is different if the field site is easily accessible.

3. Some notes on data-gathering

3.1 The importance of lexical data
Keeping a lexical file is an important part of grammar-writing. 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.” 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.

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 some 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 words, morphological forms and other grammatical combinations will not occur in a corpus of a million or even 10 million words.

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?

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 word forms and glosses, paradigms 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.

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. How large and how comprehensive should a PhD reference grammar be?

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 2009). These are comparable to the largest grammars of English, e.g. by Quirk et al. (1100 pages) and Huddleston and Pullum (1800 pages). My home library contains a few other very 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 700. None of these works were PhD theses. All were done by established 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 generally advise students that 300-350 pages, excluding appendices, should be ample. I compared 17 published grammars that are versions of PhD theses submitted either to Australian or Dutch universities and the details are given below. 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.

<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>No. pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowden</td>
<td>Taba: Description of a South Halmahera language (2001)</td>
<td>408 (451)</td>
</tr>
<tr>
<td>Ezard</td>
<td>An Austronesian language of the Milne Bay area, Papua New Guinea (1997)</td>
<td>297 (320)</td>
</tr>
<tr>
<td>Farr</td>
<td>The interface between syntax and discourse in Korafe, a Papuan language of Papua New Guinea</td>
<td>417 (459)</td>
</tr>
<tr>
<td>Hyslop</td>
<td>The Lolovoli dialect of the NE Ambae language (2001)</td>
<td>438 (476)</td>
</tr>
<tr>
<td>Jones</td>
<td>Towards a lexicogrammar of Mekeo (an Austronesian language of western central Papua) (1998)</td>
<td>553 (601)</td>
</tr>
<tr>
<td>Obata</td>
<td>A grammar of Bilua, a Papuan language of the Solomon Islands (2003)</td>
<td>281 (329)</td>
</tr>
<tr>
<td>Quick</td>
<td>A grammar of the Pendau language of central Sulawesi, Indonesia (2007)</td>
<td>601 (716)</td>
</tr>
<tr>
<td>Teng</td>
<td>A reference grammar of Puyuma, an Austronesian language of Taiwan (2008)</td>
<td>279 (309)</td>
</tr>
<tr>
<td>van Klinken</td>
<td>A grammar of the Fehan dialect of Tetun (1999)</td>
<td>322 (355)</td>
</tr>
<tr>
<td>Eades</td>
<td>A grammar of Gayo, a language of Aceh Sumatra (2005)</td>
<td>316 (350)</td>
</tr>
</tbody>
</table>
Fifteen of the grammars fall between 240 and 440 pages, with just a couple of outliers in the 600 range. The median is 336 pages. Appendices with texts usually amount to 30-50 pages but in one case (van Staden’s Tidore grammar) they run to 200 pages. My impression is that PhD grammars have grown a bit larger in recent decades. If so, no doubt the main factor driving it is competition: survival of the fittest in the job market, Darwinian natural selection.4

4.2 What should be in a grammar? Scope and organisation

What should go in a reference grammar and how should it be organised?

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 may vary considerably in scope and degree of detail. I think grammars in general, including PhD grammars, today cover a wider range of phenomena than those done, say, in the 1950s, 60s and 70s, and that the modern grammars are more discursive and readable than those produced at the height of the structural and transformational-generative grammar eras. In those times grammarians were much 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 more expansive and more readable accounts of grammatical constructions. I think there is room in a grammatical description 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 advise 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.

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

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

The content and organisation 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)

**Taba, S. Halmahera** (Bowden 2001)

**Fehun Tetun, East Timor** (van Klinken 1999)

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)
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

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 quotes a general principle from the Cambridge Grammar Series:

```plaintext
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?

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. 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. In my sample an exception is Robert Bugenhagen’s grammar of Mangap-Mbula, which has an appendix on how to say things that treats 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).

## 5. The finer details of presentation

### 5.1 Some questions

Let us now consider some of the nuts and bolts of presentation, the finer details.

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? For each construction 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
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.

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 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 words 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): it is best to describe morphology mainly with a form-to-function orientation and syntax mainly with a function-to-form orientation. This may be a useful general rule of thumb but the issues here are complex and I hope will be get some attention at the conference. 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 PhD include an index giving page references for key topics and terms.

**6. How not to complete a thesis: some bad strategies in data gathering and writing up**

A proportion of PhD students never finish their grammars. 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.
2. Perfectionism.
3. The all chapters at once strategy.
A combination of two or more of 1-3 is generally lethal. 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)
There are times where the advisor 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.

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.

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.

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-capturing and analysis. They are enthusiastic and hard-working, enjoy fieldwork and are well-organised. They have the good judgment and flexibility to recognise and accept good advice and to question that which is dubious.

REFERENCES


