



OXFORD
LINGUISTICS

THE ART OF GRAMMAR

A PRACTICAL GUIDE

ALEXANDRA Y. AIKHENVALD

The Art of Grammar

The Art of Grammar

A Practical Guide

ALEXANDRA Y. AIKHENVALD

Language and Culture Research Centre,
James Cook University

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Alexandra Y. Aikhenvald 2015

The moral rights of the author have been asserted

First Edition published in 2015

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence, or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2014935566

ISBN 978-0-19-968321-5 (hbk.)

ISBN 978-0-19-968322-2 (pbk.)

Printed and bound by
CPI Group (UK) Ltd, Croydon CRO 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

*For Bob,
indefatigable grammar-writer,
grammar-reader,
and the inspiration behind this work*

Contents

Preamble: what this book is about xiv

Acknowledgements xvii

Conventions xix

List of tables, figures, and boxes xx

Abbreviations xxii

- 1 Introduction: to write a grammar 1
 - 1.1 The language and its ‘genius’ 1
 - 1.2 What linguistic diversity is about 2
 - 1.3 The limits of grammar 3
 - 1.4 Serving many masters: a glimpse into the multiplicity of grammars 6
 - 1.4.1 The essence of a reference grammar 6
 - 1.4.2 Grammars of further kinds 7
 - 1.4.3 The place of a linguist 8
 - 1.5 Documenting a language: an open-ended task 9
 - 1.6 The boundaries of a language and the individual voice 11
 - 1.7 Constructing a grammar 12
 - 1.7.1 The building blocks 12
 - 1.7.2 To be precise 14
 - 1.7.3 Labels, categories, and meanings 15
 - 1.7.4 How to structure a grammar: some hints 15
 - 1.8 How this book is organized 18
- Notes and sources 19
- Appendix: Excursus—Linguistic fieldwork 20
- 2 A language and its setting 30
 - 2.1 To launch a grammar: the language and its profile 30
 - 2.2 The language and its speakers 32
 - 2.3 The language, its relatives, and its neighbours 33
 - 2.3.1 Chance similarities and shared structural features 33
 - 2.3.2 Common origins and genetic connections 34
 - 2.3.3 Geographical proximity and language contact 35
 - 2.3.4 The language in its genetic and areal context: what to include 36
 - 2.4 The language in its social setting 37
 - 2.5 How the people used to live: non-linguistic background and social organization 39

2.6	Previous studies of the language	44
2.7	Materials and speakers: basis for the grammar	44
	Notes and sources	45
3	Basics	46
3.1	Sound systems	46
3.2	The unit 'word'	48
3.3	The building blocks of a grammatical word	48
3.4	Inflection and derivation	50
3.5	Delineating word classes	52
3.6	Categories and strategies	52
3.7	Phrases and clauses	53
3.8	Clauses and speech acts	54
3.9	Clause types and clause structures	54
3.10	Order of words and order of constituents	57
3.11	Sentence and discourse	57
	Notes and sources	58
4	Sounds and their functions	59
4.1	Segmental phonology	60
4.1.1	Consonants	60
4.1.2	Vowels	66
4.1.3	Syllable structure	68
4.1.4	Phonotactics	70
4.1.5	Unusual sounds in unusual forms	71
4.2	Beyond segments: stress and tone	72
4.3	Phonological word	73
4.4	Grammatical word versus phonological word	74
4.5	Beyond the 'word'	77
4.6	How to decide on an orthography	79
	Notes and sources	79
5	Word classes	81
5.1	A statement of word classes	81
5.2	Essential features of word classes	81
5.2.1	Morphological categories	81
5.2.2	Syntactic functions: word classes and functional slots	82
5.2.3	Derivation between word classes	84
5.2.4	One form, several functions	85
5.3	Concomitant features of word classes	86
5.4	Grammatically defined subclasses of nouns	87

5.5	Grammatically defined subclasses of verbs	90
5.5.1	Transitivity classes	90
5.5.2	Further grammatical subclasses of verbs	91
5.5.3	Secondary concept verbs	94
5.5.4	Copula verbs and copulas	94
5.6	Grammatically defined subclasses of adjectives	95
5.7	The elusive class of 'adverbs'	97
5.8	Closed word classes and their properties	97
5.9	Summing up	99
	Notes and sources	100
6	Nouns	103
6.1	Reference classification: genders and classifiers	103
6.1.1	Genders	104
6.1.2	Noun classifiers	106
6.1.3	Numeral classifiers	107
6.1.4	Classifiers in possessive constructions	108
6.1.5	Verbal classifiers	109
6.1.6	What reference classification devices are good for	110
6.2	Number	111
6.2.1	The meanings of number	111
6.2.2	How number is expressed	114
6.2.3	Number in its further guises	115
6.3	Possession classes of nouns	115
6.3.1	What can be possessed	115
6.3.2	How to express possession	116
6.3.3	The meanings of possessive constructions	117
6.4	Case	118
6.5	Definiteness and specificity	121
6.6	Making new nouns: derivation and compounding	122
6.7	Unravelling a noun phrase	124
6.8	Summing up	126
	Notes and sources	127
7	Verbs	129
7.1	What may go onto a verb	129
7.2	Non-spatial setting and speech acts	132
7.2.1	Moods, speech acts, and sentence types	132
7.2.2	Tense	134
7.2.3	Aspect, and the structure of activity	136
7.2.4	Modality, and attitude to knowledge	138

7.2.5	Evidentiality	139
7.2.6	Reality status	140
7.3	Valency-changing derivations	141
7.3.1	Reducing valency	141
7.3.2	Increasing valency	143
7.4	Noun incorporation	146
7.5	Making new verbs	147
7.6	Multi-verb constructions	149
7.6.1	Serial verb constructions	149
7.6.2	Constructions with auxiliary verbs and support verb constructions	151
7.6.3	Multi-verb predicates in their further guises	152
7.7	Verbal word and verb phrase	153
7.8	Summing up	154
	Notes and sources	155
8	Adjectives and adverbs	157
8.1	Adjectives: syntactic functions and semantic types	157
8.2	How adjectives are special	160
8.2.1	When adjectives are 'like' nouns	160
8.2.2	When adjectives are 'like' verbs	161
8.2.3	Adjectives in their own right	162
8.2.4	Why adjectives?	163
8.3	Making new adjectives	164
8.4	Beyond adjectives: property concepts through other means	166
8.5	Adverbs	166
8.5.1	Syntactic functions, grammatical categories, and semantic types	167
8.5.2	Making new adverbs	168
8.6	Summing up	169
	Notes and sources	169
9	Closed classes	171
9.1	Personal pronouns	171
9.1.1	Person, number, and gender in free pronouns	172
9.1.2	Interpersonal relations in free pronouns	178
9.1.3	Special features of free personal pronouns	180
9.1.4	Possessive pronouns	182
9.1.5	Pronouns of argument identity	183
9.1.6	Bound pronouns	185
9.2	Demonstratives	186

9.2.1	The meanings of demonstratives	187
9.2.2	What demonstratives can do	190
9.3	Articles and noun markers	192
9.4	Interrogatives and the like	193
9.4.1	Interrogative words	193
9.4.2	Indefinites	195
9.4.3	Negative and relative words	195
9.4.4	Determiners of further kinds	196
9.5	Quantifiers and number words	197
9.6	Closed classes of further kinds	198
9.7	Summing up: a statement of closed classes	201
	Notes and sources	202
10	Who does what to whom: grammatical relations	205
10.1	Core grammatical relations and transitivity	205
10.2	How to mark grammatical relations	206
10.3	Nominative-accusative systems	207
10.4	Absolutive-ergative systems	208
10.5	Split ergative systems	210
10.5.1	The meaning of a noun phrase and the Nominal hierarchy	210
10.5.2	The meaning of a verb	212
10.5.3	Tense and aspect	214
10.5.4	Split patterns with more than one factor at play	214
10.6	'Non-canonically' marked arguments	215
10.7	Case marking, definiteness, and topicality	216
10.8	Arguments of copula clauses and of verbless clauses	218
10.9	Capturing argument structure	220
10.10	Concluding remarks	222
	Notes and sources	222
11	Clause and sentence types	225
11.1	Internal structure of clauses and clause types	225
11.1.1	Copula clauses	226
11.1.2	Verbless clauses	227
11.1.3	Possession within a clause	228
11.1.4	Comparative constructions	230
11.2	Clauses and their syntactic functions	232
11.3	Speech acts and sentence types	234
11.3.1	Imperatives and commands	234
11.3.2	Interrogatives	236
11.3.3	Exclamations	238

11.3.4	Versatile sentence types	239
11.4	Negation as a clausal category	241
11.5	Concluding remarks	242
	Notes and sources	243
12	Complex sentences and clause linking	245
12.1	Coordinated clauses	245
12.2	Dependent clauses	246
12.2.1	Relative clauses	246
12.2.2	Complement clauses	250
12.2.3	Adverbial clauses and clause chaining	253
12.3	Switch reference	256
12.4	Ellipsis and pivots	257
12.5	Reporting what someone else has said	260
12.6	Putting a sentence together	261
12.7	Concluding remarks	263
	Notes and sources	264
13	Language in context	267
13.1	Information structure	267
13.1.1	To mark a topic	267
13.1.2	Topic, focus, and contrast	270
13.1.3	Topics, participants, and pivots	272
13.1.4	'Asides' or parentheticals	274
13.2	Putting sentences together	275
13.2.1	Linking sentences: recapitulation, repetition, and 'bridging'	275
13.2.2	Ellipsis	277
13.3	To use the language	278
13.3.1	Speech formulae, greetings, and farewells	278
13.3.2	Speech genres and story-telling	280
13.4	Issues in semantics and features of lexicon	282
13.5	Envoi	283
	Notes and sources	284
14	Why is a language the way it is?	286
14.1	Shared origins	286
14.2	Language contact	287
14.2.1	What to borrow and how	288
14.2.2	Linguistic areas	291
14.3	Physical environment	293
14.4	Social structures, lifestyle, and beliefs	293

14.4.1	Genders and social stereotypes	294
14.4.2	Culture, lifestyle, and values in reference classification	295
14.4.3	Possession	296
14.4.4	Personal pronouns, questions, and commands	297
14.4.5	‘To be precise’: information source and attitudes to information	298
14.4.6	Attitudes, beliefs, and practices through grammar	301
14.5	The question ‘why?’: explanation and prediction	303
	Notes and sources	304
15	How to create a grammar and how to read one	306
15.1	Putting a grammar together	306
15.1.1	What goes into a grammar	306
15.1.2	Synchrony meets diachrony	309
15.1.3	Analysis, presentation, and argumentation	310
15.1.4	Making the most of language examples	312
15.1.5	On linguistic terminology	313
15.2	To read a grammar	314
15.3	Envoi: the makings of a reference grammar	316
	<i>Glossary of terms</i>	317
	<i>References</i>	332
	<i>Index of languages, language families, and linguistic areas</i>	357
	<i>Index of authors</i>	366
	<i>Index of subjects</i>	371

Preamble: what this book is about

Several thousand distinct languages are currently spoken across the globe, many of them by small tribal communities. Each has its own grammatical system. Comprehensive reference grammars are the basis of our understanding of linguistic diversity, and of cultural diversity as embedded in human languages. Grammars offer a unique window into the structure and cognitive underpinnings of languages, and the ways they reflect the changing world. A reference grammar brings together a coherent treatment of a language as a system, within its cultural context. Ideally, it also touches upon the history of the language.

Comprehensive grammatical analysis of the world's languages is the backbone of linguistics. Grammars are the foundation for any meaningful generalizations about what human languages are like. Linguistic typology—the science of generalizations and predictions about human languages and the underlying cognitive patterns—is based on what we learn from grammars.

Writing a reference grammar is a considerable task. It should outline the distinctive features of the language without being too esoteric. The discussion within the grammar needs to be placed within the established parameters of linguistic typology. And at the same time, a comprehensive grammar will shed light on new, perhaps previously unknown, categories and meanings.

In many ways, a reference grammar serves different masters—experts in the languages of the area, general linguists and linguistic typologists, anthropologists, and other kinds of interested readers. It may also spark the interest of the speakers themselves (who may or may not be co-authors), in uncovering the distinctive genius of their languages. It is up to the grammar writer to satisfy a number of possible readerships.

A reference grammar creates the basis for each of the other, more focused kinds of grammars (e.g. pedagogical or historical grammars), ideally creating a definitive analytic document for a language. It combines scholarly description, documentation, and analysis of language facts. It may even venture an explanation—why the language is the way it is.

Description of linguistic facts goes hand in hand with the analytic perspective one takes on them. The most comprehensive grammars are cast in the typologically informed framework recently given the name of basic linguistic theory (see Dixon 2010a, 2010b, 2012).

Maintaining a subtle balance between being detailed and being comprehensive, between being language specific and yet of interest to a more general audience is an intellectual art. Hence the title of this book, which purports to explain some of the tricks of the art.

My aim in this book is to introduce the principles of grammar writing. The structure of this book in itself follows the ways in which most grammars tend to be organized. One starts with the social and cultural setting of the language, then goes on to its phonetic and phonological make-up, and from there on to the basic building units of a grammar, and then discourse, and lexicon. However, it should not be taken as a prescriptive ‘recipe’—it is essential that the grammarian should structure their analysis in the way best for each particular language.

So as to show what a grammar ought to cover, I focus on a discussion of a variety of grammatical topics—and some parameters of their variation—across languages. As a result, the discussion here partly overlaps with topics covered by books on linguistic typology and basic linguistic theory. This is understandable, as we have to cover essentially the same ground—the structure of human languages and their recurrent features.

Note that Dixon’s (2010a, 2010b, 2012) three-volume magnum opus on basic linguistic theory differs from this book in three major ways:

- First, the present book addresses a fair number of topics not covered by Dixon in any detail. These include noun classes and classifiers, derivation and compounding, definiteness and specificity, serial verbs and other multi-verb constructions, imperatives and commands, exclamations and versatile sentence types, switch reference, information structure, and issues in linking sentences such as recapitulation and repetition.
- Secondly, the topics Dixon discusses are addressed in more depth and with more examples than will be appropriate here.
- And thirdly, the present book contains hints as to what to include in a grammar, and how to organize it.

My aim is to offer guidelines for writing grammars of spoken languages. These may be judiciously used for sign languages—with the proviso that sign languages are linguistic systems in their own right and not adaptations of the spoken ones.

A further prefatory note is in order. I use the word ‘sex’ in its traditional meaning, for the difference between women and men, rather than ‘gender’, which is a grammatical label. This allows me to say ‘female sex is marked by feminine gender’, which is a clear statement, whereas ‘female gender is marked by feminine gender’ sounds odd and confusing.

Who this book is for

The word ‘linguist’ has at least two senses. Some people see a linguist as a polyglot, who knows many languages. There is another sense of ‘linguist’: an analyst who studies—in a scientific manner—how languages are structured, where they come from, and how they work. This book is primarily for linguists in this second, scholarly, sense.

Its broader audience includes anthropologists, educationalists, and other scholars interested in how languages work and why they are so diverse. It presupposes knowledge of basic concepts of linguistics, such as phoneme and morpheme. The readers are expected to have undertaken a couple of undergraduate courses in linguistics. Many technical terms are explained in the Glossary.

Plea

This book is not the last word on the art of creating, composing, and writing a grammar. I welcome reactions, counterexamples, new facts, new ideas, to further develop, refine, and perhaps redefine the hypotheses and generalizations put forward here. Please send them to me at the LCRC, CASE, James Cook University, Cairns, North Queensland, 4870, Australia, or, in a quicker manner, to Alexandra.Aikhenvald@jcu.edu.au.

Acknowledgements

This book is a product of about thirty years of writing and reading grammars, supervising students, and commenting on grammars written by Post-doctoral Fellows and colleagues, and also conducting annual workshops on grammar-writing for PhD students and Research Fellows, jointly with R. M. W. Dixon. A set of handouts prepared for these workshops were the foundation for this book.

My own experience in writing grammars is at the heart of the book. I have written grammars of three Arawak languages (Warekena of Xié, Baré, and Tariana), one language from Papua New Guinea (Manambu), and two grammars of Hebrew. My typological work, and work on language contact patterns (especially in Amazonia), has helped me enrich my own grammars, and get a feel for what may make a grammar more enticing and more comprehensive. Over the years, I received feedback from many people, of different continents and backgrounds, and am indebted to them all.

My gratitude goes to native speakers of Amazonian languages who taught me their languages. First, my Tariana family: José, Jovino, Olívia, Rafael, Leo, Maria, Diká, Emílio, Juvenal, the late Gracialiano, Ismael, and Cândido Brito, together with other members of the Brito family (speakers of Tariana of Santa Rosa), Marino, Domingo, Ismael, Jorge, and Batista Muniz, and other members of the Muniz family (Tariana of Periquitos). And also Arthur, Amilton, and Floriano Baltazar, and the late Humberto Baltazar and Pedro Ângelo Tomas (Warekena); Ilda Cardoso da Silva, Afonso, Albino, and João Fontes, Celestino da Silva, Cecília and Laureano da Silva, and the late Marcília Rodrigues (Baniwa); the late Tiago Cardoso (Desano, Piratapuya), the late Candelário da Silva (Baré), and Marilda-Mamori and Carlito Paumarí, and Alfredo Fontes (Tucano). I am also grateful to my adopted family in the Sepik area of Papua New Guinea who revealed to me the beauty of their native Manambu (especially Pauline Agnes Yuaneng Luma Laki, James Sesusu Laki, David Takendu, and Jacklyn Yuamali Benji Ala).

I am grateful to all my former students and friends from various parts of the world—Simoni Valadares, Marcos Baltar, Cristiane Cunha de Oliveira, Claudete Lucyk, Rute Maria Coelho, Silvana and Valteir Martins, Michael Dunn, Kristina Sands, Rina Marrita, Yunseok Lee, Kazuko Obata, Antoine Guillaume, Carola Emkow, Simon Overall, Mark Post, Chia-jung Pan, Sihong Zhang, Dineke Schokkin, Hannah Sarvasy, Mikko Salminen, Juliane Böttger, Grant Aiton, Kasia Wojtylak, Nick Piper, Valérie Guérin, Elena Mihás, and many more. I have learnt a great deal from them.

My work on Amazonian languages would have been impossible without the assistance, support, and friendship of Lenita and Elias Coelho, and Silvana and Valteir Martins. I continue being inspired by Lenita, Elias, Silvana, and Valteir, and also by

the pioneering work of other Brazilian linguists in the Makú-speaking area—Cácio and Elisângela Silva and Marcelo and Claudinha Carvalho.

The members and visitors of our Language and Culture Research Centre within James Cook University—Lourens de Vries, Carol Genetti, Knut Olawsky, Ken Sumbuk, Catherine Travis, Gwen Hyslop, Azeb Amha, Nerida Jarkey, Anvita Abbi—have provided an exciting intellectual ambiance. Mike Wood and Rosita Henry have helped create a wonderful collegial atmosphere, and gave me insight into cultural features which may be included in a grammar. I am indebted to our Post-doctoral Fellows for maintaining high standards of intellectual endeavour—Tianqiao (Mike) Lu, Elena Mihas, Angeliki Alvanoudi, Simon Overall, Valérie Guérin, and Diana Forker.

Invaluable comments on just about every page came from R. M. W. Dixon, without whose incisive criticism, and constant encouragement and support, this book would not have appeared. Hannah Sarvasy, Mikko Salminen, Angeliki Alvanoudi, Kasia Wojtylak, and Nick Piper read through the draft of the book, making sharp comments and correcting errors. Valérie Guérin read through a number of chapters, pinpointing contradictions and inconsistencies. It is hard to find words to express my gratitude to you all!

Brigitta Flick carefully read through various drafts of this book and corrected it with her usual skill, perspicacity, dedication, and good humour. Without efficient and cheerful support from Amanda Parsonage who took many an administrative burden from me, I would never have managed to complete this book.

This volume owes a good deal to John Davey and Julia Steer, the Linguistics Editors of Oxford University Press. Their indomitable support makes an author, and their book, feel wanted.

The stimulating and cheerful atmosphere, encouragement, and warm support of senior management, colleagues, and friends at James Cook University have helped make it all possible.

Conventions

The AFFILIATION and LOCATION of each language is given in brackets at its first mention. The affiliation includes the subgroup if known and generally established, and then the family—for instance, Nanti is introduced as Nanti (Campa, Arawak). I only include a subgroup if it has been firmly established, as is the case with the Campa subgroup of Arawak languages in South America. I cannot expect the reader to remember the affiliation of each and every language. So, this same information is repeated now and again. The index of languages also contains this information.

Examples are numbered separately for each chapter. For instance, examples in Chapter 1 will be numbered 1.1, 1.2, and so on. All language examples are supplied with an interlinear morpheme gloss, and then translated into English. The symbol ‘+’ is used to indicate fused morphemes, e.g. a Tariana form *nhupa* is glossed as ‘1sg+grab’, its underlying form being *nu-* ‘first person singular prefix’ plus *-hipa* ‘to grab’. Portmanteau morphemes are glossed with ‘.’, for instance, Tariana *-mahka* (RECENT.PAST. NONVISUAL). All grammatical morphemes are glossed in small caps while lexical morphemes are in lower case. Pronominal prefixes are shown as 1sg, 3pl, in lower case.

Cross-references are of two kinds:

- Those preceded by \$ refer to chapter and section number: for instance, \$11.1 refers to section 1 of Chapter 11;
- Those beginning with a number refer to examples in the grammar: for instance, 11.1 refers to example 1 in Chapter 11.

References to quoted and cited material are *not* generally given in the chapters themselves. They are included, together with additional notes, in a section ‘Notes and sources’ at the end of each chapter.

List of tables, figures, and boxes

Tables

- 2.1 No chance coincidence: 'One' through 'four' in English and German 34
- 3.1 A sample paradigm in Latin: the verb *amāre* 'to love': indicative mood 49
- 3.2 Typical properties of inflection and derivation 51
- 4.1 A sample consonant matrix 61
- 4.2 Passive articulators and their labels 61
- 4.3 Active articulators, passive articulators they combine with, and their labels 62
- 4.4 Natural classes of consonants by manner of articulation 64
- 4.5 Primary cardinal vowels 67
- 4.6 Contrasting clitics and affixes 75
- 5.1 Word classes and functional slots in Tariana 84
- 6.1 How numeral classifiers in Burmese highlight different meanings of one noun 110
- 6.2 Number systems: a snapshot 112
- 6.3 Number in Warekena 113
- 6.4 A selection of oblique cases in Estonian: *kivi* 'stone' 119
- 9.1 Personal pronouns in Amele 173
- 9.2 Personal pronouns in Tamambo 174
- 9.3 Personal pronouns in Longgu 175
- 9.4 Personal pronouns in Tok Pisin 175
- 9.5 Personal pronouns in Hdi: a minimal/augmented system 176
- 9.6 Free personal pronouns in Hausa 177
- 9.7 Personal pronouns in Gala 177
- 9.8 Personal pronouns in Spanish 178
- 9.9 A selection of personal pronouns in Korean 179
- 9.10 A selection of personal pronouns in Lao 180
- 9.11 Two sets of possessive pronouns in English 182
- 9.12 Demonstratives in Palikur 188
- 9.13 Demonstratives in Hua 189
- 9.14 Properties of nominal demonstratives in English 191
- 9.15 Imperative particles in Lakhota 199
- 9.16 A selection of closed word classes in Manambu: a comparison 201
- 12.1 Semantic types of clause linking 262

Figures

- 6.1 Structure of an NP in Longgu 125
- 10.1 Nominal hierarchy and split ergativity 211

Boxes

- 1.1 Essential principles for successful fieldwork in linguistics 26
- 4.1 One segment or more? 66
- 5.1 Word classes and functional slots: a checklist 85
- 7.1 How to recognize a passive 141
- 7.2 How to recognize an antipassive 142
- 7.3 What is a causative? 143
- 7.4 What is an applicative? 145

Abbreviations

1, 2, 3 – first, second, third person

A – transitive subject

ABS – absolutive

ACC – accusative

ADJOINED.CL – adjoined clause

ADV – adverb

ALL – allative

AOR – aorist

ART – article

ART.DEF – definite article

ART.INDEF – indefinite article

AUG – augmentative

AUX – auxiliary

CA – common argument

CAUS – causative

CC – copula complement

CL – classifier

CL:ANIM – classifier for animates

CL.INAN – classifier for inanimate objects

CoCL – complement clause

CoCL:A – complement clause in A function

CoCL:O – complement clause in O function

CoCL:S – complement clause in S function

COMP – comparative

COMPL – complementizer

COMPL.CL – complement clause

COMPL.DS – complement clause with
different subject

COP – copula

COP.INTER – interrogative copula

CORR – correlative marker

CS – copula subject

DAT – dative

DECL – declarative

DEF – definite

DEF.ACC – accusative definite

DEF.ART – definite article

DEF.NOM – definite nominalizer

DEM – demonstrative

DEM.DIST – distal demonstrative

DEP.CL – dependent clause

DEP.CL(DS) – dependent clause different
subject

DIR – directional

DS – different subject

du – dual

DUR – durative

E – extended argument

ERG – ergative

EVID – evidential

f – feminine

fem – feminine

FEM – feminine

FINAL.CL – final clause

FOC – focus

FOC.A/S – focused A/S

FUT – future

GEN – genitive

H – high tone

IMPERF – imperfective

IMPERS – impersonal

IMPERV – imperfective

IMPV – imperative

INAN – inanimate

INDEF – indefinite	PRESENT.INDEF – present tense of indefinite conjugation in Hungarian
INDIC – indicative	PROHIB – prohibitive
INF – infinitive	REACT.TOP – reactivated topic
INSTR – instrumental	REAL – realis
IPA – International Phonetic Alphabet	RECIP – reciprocal
L – low tone	REC.P – recent past
LK – linker	REC.PAST – recent past
LOC – locative	REC.P.VIS – recent past visual
m – masculine	RED – reduplication
MAIN.CL – main clause	REFL – reflexive
masc – masculine	REL – relative clause
MASC – masculine	REL – relativizer
NCL – noun class	REM – remote
NEG – negation	REM.P.REP – remote past reported
nf – nonfeminine	REM.P.VIS – remote past visual
NOM.MARKER – nominal marker	S – intransitive subject
NOMZ – nominalizer	Sa – subject of an active intransitive verb
non.fem – nonfeminine	SAP – speech act participant
nsg – nonsingular	SEQ – sequencing
NP – noun phrase	sg – singular
NUM.CL – numeral classifier	SG – singular
O – direct object of transitive verb	sg.nf – singular nonfeminine
obj – object	SIM – simultaneous
OBJ – object	So – stative of a stative intransitive verb
OBL – oblique	SS – same subject
PASS – passive	subj – subject
PERF – perfect	SUBJ – subject
PERT – perturbative	TH – theme
pl – plural	TOP – topic
PL – plural	tr – transitive
POSS – possessive	TRANS – transitivizer
POSS/ATTRIB – possessive/attributive	VCC – Verbless Clause Complement
PRED – predicate	VCS – Verbless Clause Subject
PRES – present	VOC – vocative
PRESENT.DEF – present tense of definite conjugation in Hungarian	

Introduction: to write a grammar

Linguistics is the science of language, similar to how mathematics is the science of numbers. A reference grammar is a scientific enterprise. It brings together a coherent treatment of each language as a system where everything fits together, within the cultural, and historical, context of the language. The aim of this book is to offer a guide for creating a reference grammar based on empirical facts and combining description, interpretation, and analysis.

Linguistics can be considered a branch of natural science. The great linguist N. S. Trubetzkoy came to the conclusion in 1909 (at the age of 18):

that linguistics was the only branch of ‘human lore’ with a scientific approach and that all the other branches of this lore (ethnography, history of religion, history of culture, and so forth) would leave their prescientific, ‘alchemic’ stage only when they followed the example of linguistics. (Lieberman 1991: 304)

1.1 The language and its ‘genius’

Several thousand different languages are currently spoken across the world. The exact figure is hard to pinpoint. Estimates vary depending on how one counts—whether or not one includes more or less mutually intelligible varieties. Generous counts offer a figure of 6,000, and more restrictive ones suggest about 4,000 or fewer. Only a few of these are spoken by more than a million people. And indeed, English, Mandarin Chinese, Spanish, varieties of Arabic, Portuguese, and French are threatening to take over the world. In 1992, Michael Krauss estimated that at least half of languages spoken then would become extinct during the twenty-first century.

Less than a third of those languages which are still spoken have been extensively described and understood. Analysing and documenting languages on the path towards extinction is a race against time.

Every language is a repository of beliefs, heritage, history, and traditional laws. The loss of a language is bound up with the loss of indigenous knowledge about the environment, its biological diversity and traditional means of sustainable use. Language loss deals a blow to an ethnic group’s identity and self-esteem. Speakers of Tariana, an

endangered language from north-western Brazil, often complain to me that they have to speak a ‘borrowed’ language now that their own is being lost. They feel impoverished.

For scientific linguists, loss of a language means loss of a unique system, perhaps a missing link, which might have turned out to be crucial for our understanding of the workings of the human mind, peoples’ histories and the ways we interact.

A comprehensive reference grammar—accompanied by a detailed dictionary and a collection of natural stories in the language, as well as audio and perhaps video recordings—will ensure that the language is not irretrievably lost. Ideally, a reference grammar should reflect language use, and its history, and also the ways in which it relates to other languages within its area, and within the world. It will form the basis for further knowledge of the language’s history, and for generalizations and predictions about language in general.

A grammar will capture the unique genius of the language—the way meanings are expressed, and how categories are realized. This is how we can understand how languages differ, and why, and what cognitive and other mechanisms they may reflect.

1.2 What linguistic diversity is about

In recent times, a great deal of attention has been directed to biological diversity—what gives rise to it, where it is distributed, its implications, and prognostications for the future. Linguistic diversity is an equally important topic. Why is there a high concentration of different languages in certain areas, but a relative paucity in others? Is linguistic diversity something to be valued, or deplored, and what can we learn from it? How is it being affected by rampant globalization, and what does the future hold?

When we talk about ‘diversity’ we refer to at least three possible things.

1. DIVERSITY IN NUMBERS. As we survey the world, the number of languages spoken in a certain-sized area differs a great deal. Papua New Guinea (PNG, 463 km²) and Paraguay (407 km²) are of comparable size but there are over 850 distinct languages spoken in PNG and (before the European invasion) Paraguay had only about 20. This difference can partly be explained by geography. PNG has many swamps and high mountains which impede communication, these being absent from Paraguay.

2. DIVERSITY IN ORIGINS. Most languages—with the exception of a few isolates, or linguistic orphans—are known to belong to a genetic family. Multiplicity of families creates diversity in language origins. This is higher in some areas, and lower in others. The island of New Guinea is home to over 60 unrelated families. About 370 Bantu languages are spoken over a considerable-sized area in East and Central Africa. These languages belong to one family, and, despite their distinctness, they share many similarities in their grammars.

Establishing linguistic families is based on fairly strict principles of comparative linguistics—perhaps the most scientific branch of the discipline. And surface

similarities between languages should not be taken as indicators of their relationship. More on this later.

Most language communities—with the exception of the few confined to an isolated island or a remote mountain region—are in contact with other communities. The communities interact—through trade, intermarriage, shared festivals and rituals, and also military conflicts. Their languages also interact. They may come to sound similar. Some vocabulary may be borrowed. Some structural features of languages may converge: for instance, neighbouring languages may develop similar systems of noun classes. And yet, the languages remain distinct.

3. DIVERSITY IN STRUCTURES. In some parts of the world, the languages spoken within a given area demonstrate remarkable diversity of structure. In other regions, there is a considerable degree of uniformity for the goodly number of structural and lexical features. Benjamin Lee Whorf (1956) coined the term ‘Standard Average European’ as a way of drawing attention to the fact that more than a dozen languages spoken in Europe show many similarities in their grammatical properties, lexical extensions, and idioms. These languages have no more than two or three genders, most are straightforwardly nominative-accusatives; many have similar systems of auxiliary verbs. In contrast, languages in a region of the Amazon basin of about the same size as Europe, exhibit a mind-boggling diversity of grammatical structures, lexical subtleties, and culturally-determined aphorisms. The number of gender-like noun classification systems varies from one language to the next. Hardly any language is straightforwardly nominative-accusative. And many categories—such as nominal tense, evidentiality, frustrative modality—are alien to a European centred scholar.

Structural diversity cannot be fully understood without comprehensive grammatical study of each language. This is what a reference grammar is primarily useful for.

No language exists in a vacuum—it is used by people within their own distinct social structure. Linguistic diversity goes hand in hand with social diversity. Diversity in the physical environment and many extra-linguistic features may serve to explain diversity of other sorts. A comprehensive reference grammar will provide a coherent foundation for understanding the language’s special character as a system where ‘everything fits together’. This is a basis on which the understanding of ‘how’ and ‘why’ will then be built.

1.3 The limits of grammar

No two languages are entirely the same, nor are they entirely different. It is as if there were a universal inventory of possible grammatical and lexical categories and meanings, and each language makes a different set of choices from this inventory. Any meaning can be expressed in any language. But it does not have to be.

As Franz Boas (a founding father of modern linguistics) put it, languages differ not in what one **can** say but in what kind of information **must** be stated: ‘grammar . . . determines those aspects of each experience that *must* be expressed’ (Boas 1938: 132). French has an obligatory two-term gender system, German has three, while Armenian and Hungarian make no gender distinctions at all in their grammar. In French and in German every noun requires agreement in gender on adjectives and articles. In Armenian and Hungarian the same form will be used with any noun.

However, in every language one can distinguish male from female, as a natural gender—or sex—category. This can be achieved through special words for males and females. The number of distinctions in a lexicon can be unlimited. A grammar forces us to make limited choices. It may focus on gender as it is expressed through an obligatory mechanism. Words for men, women, and other males and females will then be relegated to the lexicon.

That is, every language has a number of obligatory grammatical categories from which a choice must be made in order to construct an acceptable sentence. To translate into French *The child fell down* one must know the gender of the child, since this has to be specified. To translate this sentence into Tariana (from Amazonia) one must specify the source of information on which the statement is based—whether you saw it, or heard it, or inferred it, or were told about it by someone else (this is a grammatical system of evidentiality).

Many languages have a closed set of grammaticalized expressions of location in time: these can involve present, past, and future; or nonpast, recent past, and remote past, etc. Amele, a Gum language from Papua New Guinea, distinguishes four verbal forms which can be defined as: (i) today’s past, (ii) yesterday’s past, (iii) remote past (what happened before yesterday), and (iv) habitual past—something that often occurred in past time.

Each sentence in Amele has to be specified for one of these parameters. Their application is not rigid: what occurred in the hours of darkness the previous night can be referred to either by today’s past tense or yesterday’s past tense depending on whether the speaker considers them related to other events on the previous day or on the day of the utterance. English and many other languages have special words for ‘today’, ‘yesterday’, ‘the day before yesterday’, and so on which one may use or not depending on the speaker’s whim. So does Amele. A closed grammatical system will coexist with sets of *lexical items* which refer to location in time and a potentially unlimited number of ‘composite lexical expressions’ for measuring time intervals.

In other words, a closed grammatical system offers limited options. This is in contrast to the lexicon, where the choices are potentially open. So, for grammatical tense, ‘even the maximal system would have at most tens of categories, rather than the several orders of magnitude made possible in the lexicon’ (Comrie 1985: 9).

A grammar of a language is not a random set of facts. It is rather like a mechanism which organizes the language, or a motor that sets it in motion. At a very early

stage of learning a language, a child acquires the principles of grammar and applies them. This is the basis of a speaker's generative ability of creating an infinite number of texts and sentences, based on the rules of grammar. How these principles, and their grammatical meanings, differ from one language to the next is what we aim at discovering.

What is expressed through a grammatical distinction in one language may have to be phrased lexically in another. This is something we have just seen with yesterday's past in Amele. A grammatical form can be shown to develop from a lexical item. This is known as 'grammaticalization'. In Ewe, a major language of Ghana, the word *nO* means 'mother'. The same form can be used as a suffix to names of animals as a feminine marker, for example *nyi-nɔ* (cattle-feminine) 'cow'. Facts like this one are useful to include in a grammar: they show the ways in which the lexicon and grammatical forms interact.

Lexicon and grammar are intertwined. Different classes of lexicon may have different grammatical properties. Grammatically defined subclasses of nouns often include body parts, kinship terms, and place names. Grammatically defined subclasses of verbs may include verbs of perception and cognition, verbs of stance and posture, and verbs of giving. Colour adjectives may behave differently from adjectives referring to size. (We turn to these in Chapters 5 and 8.) Meanings within the lexicon may, at least partly, shape the meanings reflected in grammar. Inasmuch as this is the case, information about lexical categories are relevant as a background for a full grammar.

In summary: the lexicon and grammar of a language are two complementary parts, each in its own right. The lexicon of each language reflects the world in which it is spoken in numerous minute distinctions, realized as many individual items. It is open to new words and notions. The grammar—much more restricted and much more mechanistic—is a closed system. In some ways, it may be seen to reflect some real-life distinctions and depend on them. A reference grammar may address at least some of these.

TO REMEMBER: Every language has a grammar, and no grammar is primitive. However, some grammatical descriptions are. A few missionaries of the colonial era claimed that non-European languages (Chinese, South American—you name it) have 'no grammar'. What they meant was that the grammatical mechanisms of these languages were beyond them to discover: they do indeed differ from what a European might expect. What they also meant is that those languages had no 'prescriptive' grammar. And this is something we turn to in the next section.

Ideally, a reference grammar will be accompanied by a comprehensive lexicon, and a collection of stories of different genres—showing how people actually talk. If appropriate, there may be also a collection of electronically accessible audio and video resources. What is the place of a reference grammar within grammars and language materials of other sorts?

1.4 Serving many masters: a glimpse into the multiplicity of grammars

1.4.1 *The essence of a reference grammar*

A reference grammar is a comprehensive result of language analysis. A grammarian's first task is to study the complete system of a language at some point in time—that is, focus on its synchrony. As Antoine Meillet (1926: 16) put it, ‘une langue constitue un système complexe de moyens d’expression, système où tout se tient’ (‘a language makes up a complex system of means of expression, a system in which everything holds together’). Scientific linguists who produce comprehensive grammars of languages naturally follow this tenet. Those who look at isolated bits of language, for some particular issue, go against this fundamental principle of systematic analysis.

The study of language can be approached in at least two ways (more details are given by Dixon 2010a). One involves the postulation of a ‘formal theory’ or a framework which puts forward certain deductive hypotheses about language structure and examines selected language data for confirmation of these ideas. There are, typically, many competing ‘formal theories’, each making claims about different aspects of language (these have some similarities to competing theories of economics, or of literature). Many ‘formal theories’ are associated with Chomsky and various generations of his students. These may make some contribution to understanding some aspects of language organization, especially through the prism of European languages—and mostly English—but they never offer a full picture.

A reference grammar (also called ‘analytical’ grammar and ‘descriptive’ grammar) will outline the distinctive features of the language—and transmit its ‘linguistic genius’—with just the right amount of detail, including reference to the language’s history if possible. Ideally, the discussion within a grammar is expected to be placed within the established parameters of linguistic typology—reflecting how languages work, and expanding our understanding of the categories and principles of their organization. At the same time, a comprehensive grammar will disclose new, previously unknown, categories and meanings.

Description of linguistic facts goes hand in hand with the analytic perspective one takes on them. The lasting comprehensive grammars are cast in a typologically informed framework based on cross-linguistic inductive analysis of numerous languages. This framework has been recently given the name of basic linguistic theory (see Dixon 2010a, 2010b, 2012). In this framework, every analytic decision has to be proved, but is not constrained by the requirements of an ad hoc formal model. It is oriented towards expanding our view of structural diversity. This is the perspective taken here.

A reference grammar should last beyond the life-span of its author. This is the main reason why it should not be cast in any of the time-line formalisms which come and go with startling frequency. Two grammars of Bolivian Indian languages—the two isolates Itonama and Movima—were cast in mathematical-type tagmemic

framework-of-the-day. They are a puzzle to a modern reader. One gets an idea of the order in which morphemes go, but not a hint on their semantics or any of the intricacies of these languages—which one learns about from later work, cast within a much less restrictive framework. Migliazza's (1972) grammar of Yanomami, a South American language, is cast in the Chomskyan 'generative' framework of the day. His concern was to fit the language into a transformationalist framework, rather than to see what distinctions were expressed in the language itself. As a consequence, there is no mention of classifiers or evidentials (grammaticalized expression of information source). One has to study other, much less restricted, grammars of Yanomami to learn about these—Borgman (1990), Gomez (1990), and Ramirez (1994).

Being able to read and understand a typologically oriented grammar implies being acquainted with basic notions of linguistics, and principles of analysis. A reference grammar is aimed at a scholarly audience, and is based on linguistic analytical methodology which need to be mastered.

1.4.2 *Grammars of further kinds*

Reference grammars are the foundations for other grammars, whose objectives and audiences are more specific.

A purely HISTORICAL grammar focuses on a diachronic approach and the history of a language within the context of its proven linguistic relatives. It is not sensible to run before one can walk. What Leonard Bloomfield (1933: 18) called 'the natural relation between descriptive and historical studies' reflects a common-sense assumption—'the need of descriptive data as a prerequisite for comparative work'. A historical study can 'only be as accurate and only as complete as these data permit it to be' (Bloomfield 1933: 19). In other words, a historical grammar is derivative of a comprehensive statement of facts and their coherent analysis. A reference grammar should ideally be informed by existing historical and comparative studies: these may shed light on otherwise seemingly random and only partly predictable variation.

A PRESCRIPTIVE grammar offers discussion of norms developed through the language's history, often in an artificial way as a result of language reforms. This type of grammar is especially appropriate for a language which has a literary norm, such as many Indo-European and Semitic languages. A prescriptive grammar may contain evaluation of different variants, in terms of how they reflect the class, educational background, and social standing of the speakers.

A PEDAGOGICAL—or a teaching—grammar aims at teaching the language to a wider audience of native speakers, or second language learners, or—not infrequently—those members of a community who had lost the command of a language and wish to learn it back. A pedagogical grammar needs to be organized so as to best suit their aims. It cannot be considered 'primitive', or inferior to any other grammar. This is a different type of enterprise. Importantly, a linguist taking part in this will need to have some training and experience in how to teach languages.

A pedagogical grammar is organized in such a way as to facilitate learning, and may be accompanied by a series of exercises. It may also be suited to a particular language situation (that is, depending on whether the language is actively spoken in a community or not), and to a type of language programme to which it may be tailored. In Mithun's (2007) words, speakers dedicated to preserving their languages 'are acutely interested in the words of the language but also its history, and its special structures'.

A collection of texts, or a dictionary, may be accompanied by a GRAMMAR SKETCH. In contrast to a reference grammar, a grammar sketch—or a sketch grammar—offers the basic grammatical facts of a language, without going into justification and analytical decisions. A sketch grammar may just say: the language has three genders. A reference grammar has to justify why. A sketch grammar is a subsidiary adjunct, while a reference grammar is a complete document in its own right. A sketch grammar can be accompanied by a glossary of morphemes. This serves as an aid to the dictionary, whose purpose is to cater for native speakers, and learners, of the language.

The reference grammar may serve as the basis for a historical grammar. As the definitive analytic document for a language, the reference grammar can be adapted to other purposes—such as teaching, or reclaiming, a language. Using a reference grammar to learn or teach a language, or to issue 'normative' statements, will not be putting it to good use.

Further by-products of a reference grammar may include literacy materials, readers, dictionaries, vocabularies, videos, multimedia resources, and statements of a linguistic norm. Linguists may help, or be instrumental, in translating religious and other materials. Many grammars have been produced by members of the Summer Institute of Linguistics. Further outcome of their activity is translating the Bible—which sometimes gives the people the feeling for a status of their language in the eyes of others. I took part in translating a set of Sunday prayers into Tariana—on the request of the speakers. These additional activities are extraneous to grammar writing.

1.4.3 *The place of a linguist*

A grammarian—the linguist—does not have to speak the language natively, nor to be born a native speaker of the language. A symbiotic relationship is usually established between the linguist and the community of speakers for whose language the reference grammar is being created. We turn to the ensuing 'priceless partnership' in the Excursus on linguistic fieldwork. The linguist is sometimes bound to take on a social role—in Nora England's (1992) words,

every time we write an article about a language we do several things: we make an analysis of some body of linguistic data, we discuss that analysis in the light of current pertinent theory, we select examples of speech to illustrate our points, and we bring that language into at least momentary prominence. . . . Language prominence resulting from linguistic research has many non-linguistic consequences. (England 1992: 31)

By-products of a reference grammar which we mentioned in §1.4.2 are among the non-purely-linguistic obligations for the linguist.

The linguist is primarily a scholar, and their obligations are scholarly honesty and systematicity. Nora England (1992: 34–5) sums it up: ‘publishing descriptions and analyses of the language we work on [which are] of the highest possible quality, and making those publications available to speakers of the language’.

But as we get involved in the social fabric of the language, we cannot help assuming social obligations—involving, in Nora England’s words,

1. Recognizing the political and social context for our research and, where necessary, taking the part of the language we study and its speakers;
2. Recognizing the rights of speakers of politically subordinate languages over those languages, and paying attention to public presentation of facts about their languages;
3. Contributing to the training of linguists who are speakers of subordinate languages, at every level from the empirical to the theoretical. (England 1992: 34)

An analytic reference grammar is the responsibility of the linguist. It has their authorship and is ultimately their intellectual property. The cultural heritage belongs to the people linguists work with.

1.5 Documenting a language: an open-ended task

Language documentation in its proper sense involves ‘documenting a language as it is used by speakers in various settings from everyday conversation to formal oratory’ (Mithun 2007). A reference grammar is its most vital component, together with a comprehensive dictionary, a collection of texts of various genres, and also—if viable—practical orthographies, teaching materials, and multimedia.

Ideally, a reference grammar may include not just the spoken part, and extend to the role of ‘body language’ as a further modality. Some speech-cum-gesture combinations are a case in point. Saying *yay* (or *yea*) *high* in American English is accompanied by a gesture showing how high an object, or a person, is. Tariana has a set of ‘gestural’ deictics: a form goes together with a gesture indicating the object’s size and shape. Indicating the position of the sun, and direction, is part and parcel of story-telling in Amazonian cultures. A comprehensive multimodal analysis of any language is a task for the future.

TO REMEMBER: different audiences value different types of output. Most speakers value dictionaries above grammars. But this is not to say that reference grammars are appreciated by no one but linguistic scholars. On the contrary. Jovino Brito, the President of the Association of the Tariana and a highly competent speaker of the language, said to me, contentedly, after having received a copy of the Tariana grammar and other materials: ‘Good, my older sister. Now we have a real language, with a grammar book, a dictionary, a manual, and a book of stories’.

As Mithun (2007) put it, proper language documentation involves

documenting the language as it is used for speakers in various settings from everyday conversation to formal oratory. . . . Particularly in the case of endangered languages, what is documented now will be utilised for purposes well beyond those we can imagine at the present time. (Mithun 2007: 44, 55)

For languages which are still spoken, a linguist is instrumental in providing an ‘open-ended’ documentation—the more we document now, the fewer unanswered questions will come up in the future.

Modern technology, especially audio-recording equipment, allows us to ‘record spontaneous, unscripted speech in real time’ (Mithun 2007: 55). Thanks to technological advances, our data and our analyses are more precise. We can now offer an analysis of intonation patterns. Video recordings help register embodied language—gestures and hand and face movements, including lip-pointing, which may well form an integral part of the grammatical structure.

One should not, however, mistake the means for the purpose. Technology should be seen as an accessory, and not an end in itself. If overused, and overtrusted, technological marvels may become a mixed blessing. In a humid tropical environment—such as north-west Amazonia and the Sepik area of New Guinea, with no electricity supply—a computer may become an encumbrance. And the same for video-recorders. We should also recall that speakers of previously undocumented languages, in remote locations, may feel intimidated by flashy gadgets. The constant presence of a video recorder or a computer may do nothing but alienate the fieldworker from the community where they are trying to establish themselves. Even a tape-recorder can be an unwelcome intruder.

There is no doubt that putting conversations, texts, and other information on the world-wide web, and producing videos and web-based archives, is close to the heart of many of our computer-loving linguistic colleagues. But materials without proper grammatical and lexical analysis are of little use to either linguistic posterity or to speakers themselves. For one thing, we can only hope that computerized databases will survive for more than a few years; books have and will survive for centuries. Putting web-based data together may be easier and quicker than painstakingly writing a grammar and producing a competent dictionary. But web-based archives need constant updating, and book pages do not. And if a language is spoken in a remote community, say, in Amazonia, Africa, or New Guinea, with no constant electricity supply, what use is a website? As a colleague from Europe remarked to me, the current focus on computer-based ‘documentation’ is akin to racism, or ‘neo-colonialism’, deepening the gulf between the ‘haves’ and the ‘have-nots’. Leo Yabwi, headmaster of a local school in the Manambu-speaking village of Avatip in the Sepik area of New Guinea, concurred: ‘We need books. What use are computers to our kids? They need to learn their heritage, and not computer games.’

He had a point here: we must not forget that the more the speakers of a language are exposed to new technologies, the more danger there is that it might get ousted by dominant, 'global', communication devices. Rephrasing Dixon (2007a: 144), 'self-admiration in the looking glass' of computer technology can wait; 'linguistic description must be undertaken now'. Documenting a language in all its cultural manifestations does not involve making a fetish out of appliances.

1.6 The boundaries of a language and the individual voice

What is a language, and what is a dialect? Varieties of speech which are similar to each other and mutually intelligible are called dialects of a single language. A language then can be viewed as an ensemble of dialects. Dialects can be differentiated geographically, for example urban, rural, or regional, or socially, for example standard and vernacular. And when dialects come to differ from each other to such an extent that they stop being mutually intelligible, they will have attained the status of different languages. Brazilian and European Portuguese have a somewhat different pronunciation, grammar, and vocabulary. However, speakers can easily understand each other—we can thus speak about European and Brazilian dialects of Portuguese. Dutch and English are different languages: speakers of Dutch cannot understand English without learning it. The issue can be complicated by the fact that mutual intelligibility may be a matter of degree, and may vary with individual people. Scots English may be unintelligible to speakers of some varieties of American English (to an extent that the Scottish-made film 'Rob Roy' was subtitled in the USA!). Some speakers of Portuguese cannot understand Spanish. Others can—given a certain amount of time, goodwill, and effort.

Mutually unintelligible languages can be referred to as 'dialects' for historical and political reasons. Italian dialects ('dialetti') are not mutually intelligible—a speaker from Rome will not immediately understand one from Sardinia without additional learning. Many of the Chinese 'dialects' are in fact very different linguistic systems. In Arabic linguistics, the term 'dialect' refers to many modern varieties, most of which are not mutually intelligible. This follows an ages-old tradition. The interested reader could learn more about this in Owens' (2006) *Linguistic History of Arabic*.

Norwegian and Swedish provide an opposite example. The standard languages are mutually intelligible. But they are not referred to as 'dialects', as they ought to be on linguistic grounds. These are national languages of two different countries, with different literary traditions. That is, cultural identity and sociolinguistic considerations determine the use of the term 'language' rather than 'dialect'. Even mutual intelligibility may depend not only on linguistic factors, but also on attitudinal and sociolinguistic ones.

As Max Weinreich (1945) put it, 'a language is a dialect with an army and a navy'. Serbian and Croatian are dialects of what used to be called the Serbo-Croatian language. But once Serbia and Croatia became separate states, there was a policy to make words in these two languages as different as possible to try to show that they are totally

distinct languages. Orthodox Serbians use Cyrillic script, and predominantly Roman Catholic Croatians use Latin letters: these differences obviously help press the point about them being distinct.

A combination of national and cultural identity, and political and sociolinguistic considerations, often makes it difficult to decide where one language stops and another one begins. In Hock's (1991: 381) words, 'there is no clear line of demarcation between "different dialect" and "different language": 'it is gradient, not discrete'. One often finds a range of varieties—dialects in the linguistic sense—spread across a territory. Varieties next to each other will be mutually intelligible. Those at the extremes, far from each other, will not be. This is what we refer to as a 'dialect continuum'.

A grammar, ideally, focuses on just one geographical—or dialectal—variety of a language. However, other dialects may shed light on apparent exceptions or irregularities. A more archaic dialect may provide an explanation for something otherwise bizarre.

How representative do we expect the grammar to be, and whom does it represent? Ideally, a grammar ought to reflect a linguistic community—that is, 'the regular patterns that characterize the natural exchanges in the speech community' (Poplack 1993: 263).

But the notion of community goes only so far. A careful fieldworker may be able to identify groups of speakers along which variation is observed. Among the Urarina, younger people use a somewhat different grammar.

Variation can serve as a 'boundary marker': there may be different forms and rules depending on social class, education, and gender. In a ground-breaking study of physical features of 'women's' speech among the Tohono O'odham (a Uto-Aztecan group from Mexico), Hill and Zepeda (1999) show how women (not men) use a pulmonic ingressive airstream in order to construct a special atmosphere of conversational intimacy, taking advantage of size differences between male and female vocal tracts: such sound production is easier to achieve with the smaller female larynx and pharynx. Physical attributes—including high-pitched voice—typical of women come to be associated with 'female talk'. Just how much of such variation is included depends on the grammar writer and the data, and the size and diversification of the community itself.

If we are dealing with a language used by just a handful of last speakers, the degree of individual variation may be high. This is especially the case in languages which are falling out of use—like Embo Gaelic, Saaroa, or Tariana. Then, in Johnstone's (2000: 411) words, speakers may 'have different grammars'. Describing the special grammar of each individual—however few of them there are—is a strenuous task. This is one of the many problems one faces when working on an obsolescent—or a moribund—language.

1.7 Constructing a grammar

1.7.1 *The building blocks*

Linguistics is a scientific enterprise where every statement requires substantiation and proof. Creating a grammar involves building up an empirical set of data, and then

proceeding towards their organization and explanation by making inductive generalizations, and uncovering the principles behind them. The steps in building a reference grammar are:

I. Collection of data and their description. This is done in terms of an overarching typological theory of language structure (each description providing feedback to refine the theory); this is describing ‘how’ languages are the way they are.

A grammar can be based on numerous sources. If the language is still in use, the major set of data should come from original fieldwork.

In the introduction to his famous anthropological account of the Akwẽ-Shavante, David Maybury-Lewis (1968) comments:

Most anthropological reports nowadays specify how long the author spent in the field, but they do not always indicate how much of that time was actually spent in daily contact with the people studied and how much elsewhere—for example in a near-by city. Nor do they always mention other pertinent details of such contacts. We are not always told how the field-worker was received by the people [they] studied and how [they] went about collecting [their] information. It is often difficult to discover whether [they] shared living quarters with the people, or occupied a separate dwelling in the same community, or one at some distance from the community, or whether [they] commuted from another community altogether. . . . I suggest that it is time we abandoned the mystique which surrounds field-work and made it conventional to describe in some detail the circumstances of data collecting, so that they may be as subject to scrutiny as the data themselves. (Maybury-Lewis 1968: xix–xx)

Similar comments are relevant for linguistic accounts of work on little-known languages: as much detail as possible concerning fieldwork, and the narratives, conversations and other materials used for the grammars, should be included.

Older documentation and earlier sources—including texts (recorded or simply written down)—may also be used, with as much caution and appreciation as necessary. It is *de rigueur* to always distinguish one’s own results and other sources. There are examples of people who muddle up sources—and the results are poor. Written sources in the language can also be used—but with care. Haspelmath’s (1993) grammar of Lezgian heavily relies on translations into Lezgian from Russian, the dominant language (partly due to the limited amount of fieldwork the author chose to undertake). Such materials often reflect the original language and show calques. This is why caution is due. The same applies to web-based examples from Google and such. If the language is no longer in use, one has little choice. But care, caution, and precision are recommended.

II. Explanation. This involves addressing ‘why’ things are as they are. Why does Swahili have eight genders, German three, French two, and Hungarian none at all? Why does Hua, spoken in the highlands of Papua New Guinea, have several demonstratives, indicating both ‘close’ and ‘distant’, as well as ‘level’, ‘up’, and ‘down’, while English has just *this* and *that*? This may offer a pathway for generalizations about the language type, and the ways in which categories are expressed.

But remember: As Leonard Bloomfield (1933: 20) put it,

The only useful generalizations about language are inductive generalizations. Features which we think ought to be universal may be absent from the very next language that becomes accessible. . . . The fact that some features are, at any rate, widespread, is worthy of notice and calls for an explanation; when we have adequate data about many languages, we shall have to return to the problem of general grammar and to explain these similarities and divergences, but this study, when it comes, will not be speculative but inductive. (Bloomfield 1933: 20)

III. Prediction. Synchronic prediction involves saying ‘if a language has X, it is highly likely also to have Y’. For example, if verbs have pronominal affixes marking subject and object, then the class of adjectives is likely to be grammatically more similar to verbs than to nouns. Diachronic prediction foresees the way a language is likely to change over time. For example, in English irregular verbs are continually being regularized, as in *strove* being replaced by *strived*. This can be expected to continue, with—in the fullness of time—*sang* giving way to *singed*, and so on.

To fully understand a language, a linguist has to be more than just a scholar of languages. Kate Burridge (2007) makes this point very clear, with regard to the archaic Pennsylvania German spoken by an isolated group of Mennonites in Canada:

But it wasn’t long before I realised how inadequate my grammatical descriptions would be if I ignored the social and cultural information. Even my forays into Middle Dutch syntax had taught me that much. In the case of the conservative Mennonites, every aspect of their life style is saturated with symbols that express a commitment to qualities like frugality, equality and humility and, in particular, the subordination of the individual to God’s will. These symbols of subordination are evident in the shape of the lexicon, but are also deeply embedded in the grammatical structuring of the language. (Burridge 2007: 39)

1.7.2 To be precise

Description of linguistic facts goes hand in hand with the analytic perspective one takes on them. The typologically informed framework recently given the name of basic linguistic theory has been, for centuries, the tacitly accepted framework for those grammars that have outlasted the fads.

A number of principles underlie successful grammar writing:

- A. Clarity of presentation.
- B. User-friendliness.
- C. Explicitness.
- D. Presenting alternative solutions where possible, and assessing each.
- E. The quest for explanation: why?

This is where a synchronic grammar may turn to a diachronic perspective.

1.7.3 Labels, categories, and meanings

What term to use? Basic labels for categories and forms usually reflect functions and a cross-linguistic definition. The term ‘serial verb’ refers to a sequence of verbs which form one predicate and bear no marking of subordination or coordination of any sort. It is therefore not advisable to extend this term to cover auxiliary verb constructions as in English. However, some linguistic traditions go against general linguistic categories.

A classifier is a morpheme which is chosen, in a particular morphosyntactic context, to categorize the noun in terms of its intrinsic properties. But in Athapaskan linguistics, classifier means ‘voice’ marker. And ‘gender’ is used in lieu of ‘classifier’. A grammar writer must strike a balance between terminologies of different sorts. Some choose to have a glossary of terms at the end. Or they invent a new, ad hoc term, reflecting a novel or hard to capture notion.

Caution needs to be exercised. As Colin Masica (1991: 774), a major scholar in Indo-Aryan linguistics, puts it,

at one period in the history of structural linguistics . . . traditional grammatical terminology was avoided in favour of *ad hoc terms* arising supposedly only from the exigencies of the language being described, often taken from the form in question itself: “the *te*-form”. Whatever the merits of this position in, for example, language teaching, it makes cross-linguistic comparison of different forms with similar functions difficult. For that general labels are needed, and traditional terminology is probably the best source for them, although it needs refinement and standardization. (Masica 1991: 774)

A grain of salt should be added here. There are not enough terms, and new ones keep being introduced. A careful grammarian may initially describe a new phenomenon, for which a label may come years later. Christaller (1875: 144) was the first linguist to have identified the existence of a sequence of verbs which form one predicate in Akan. Without referring to the term ‘serial verbs’, he identified the phenomenon of verb serialization central to the grammar of many West African, and other, languages. The term ‘serial verb’ didn’t come into general circulation until the late 1950s to early 1960s.

In his magisterial grammar of Takelma, Sapir (1922) did not use the term ‘mirativity’ (introduced into general linguistic currency by DeLancey 1997). But what he describes as ‘surprise overtones’ of an inferential evidential (1922: 158, 200) are what is now known as mirative extensions of an evidential (in itself, an obligatory marker of information source).

Whichever term you use, be consistent, clear and explain your choice well—this is the rule of thumb in assigning a proper label for a category.

1.7.4 How to structure a grammar: some hints

And now, to a frequently asked question: how does one go about organizing a grammar? Sadly, some linguists, and especially arm-chair typologists, tend to scour a

grammar just for a particular topic (we return to this in Chapter 15). This is why it is important to have a clearly stated organization, and also an index. Others read a grammar from the beginning to end, following the story line as it unfolds—like one reads a novel, or a biography. A reference grammar should be written in such a way that it can be read through like this, with the structure of the language unfolding as the text flows.

A grammar starts with an introduction containing basic facts about the language and its social setting, the family it belongs to, and the cultural background. This is followed by a statement of phonology and phonetics. Then comes morphology, then syntax, then sometimes also discourse properties and some notes on lexical semantics. Other types of organization are also possible. Syntax may be placed before morphology; it is however important that the relevant facts about inflectional morphology be summarized first. Otherwise the discussion of syntax may become unintelligible. A detailed discussion of phonological processes may appear later in the grammar; however, the phonemes of the language have to be introduced at the beginning, as building blocks for the understanding of what follows.

A linguist may choose to put the most exciting part of the language more upfront. Dixon's (1972) classic grammar of Dyirbal begins with a short discussion of word classes and basic inflections in the language, and then turns to the most complicated (and most interesting) part—the syntax. Most of the morphology, phonology and semantics are discussed later. The grammarian decides how to organize the presentation—it just needs to be comprehensive.

The basic rule of thumb is as follows: if the analysis of a category A refers to some facts concerning category B, then the chapter, or section, dealing with B needs to be placed before the section dealing with A.

An important point in grammar writing is choosing appropriate examples to illustrate each phenomenon. To make the language, and the culture, come alive through its grammar, examples need to be based on participant-observation, and accompanied by the context in which they occurred. If they come from traditional narratives, placing them in the context of the narrative makes the grammar an interesting read.

Examples are essential—they support what you are saying, and are the foundation for your argument. In order for the examples to be clear and user-friendly, they need to be glossed. Here is an example of conventional glossing, with lower case for lexemes and pronominal number (sg, du, pl) and SMALL CAPS for grammatical elements.

- (1.1) yaw s-i-ht ti o-d
 3sg.masc:NOM 3sgA-3sgO-look.for ART:ACC universal-ACC
 ‘He looks for a universal’