

BASIC LINGUISTIC THEORY

R. M. W. DIXON

VOLUME 1

METHODOLOGY

Basic Linguistic Theory 1

Basic Linguistic Theory

R. M. W. Dixon

The three volumes of *Basic Linguistic Theory* provide a new and fundamental characterization of the nature of human languages and a comprehensive guide to their description and analysis. The first volume addresses the methodology for recording, analysing, and comparing languages. Volume 3 (which will be published in 2011) examine and explain every underlying principle of grammatical organization and consider how and why grammars vary.

Volume 1

Methodology

Volume 2

Grammatical Topics

Volume 3

Further Grammatical Topics (*in preparation*)

A complete list of R. M. W. Dixon's books may be found on pp. 380–1

Basic Linguistic Theory

Volume 1 Methodology

R. M. W. DIXON

The Cairns Institute
James Cook University

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford ox2 6DP

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 in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© R. M. W. Dixon 2010

The moral rights of the author have been asserted
Database right Oxford University Press (maker)

First published 2010

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, 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 book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
CPI Antony Rowe, Chippenham, Wiltshire

ISBN 978-0-19-957105-5 (Hbk)
978-0-19-957106-2 (Pbk)

1 3 5 7 9 10 8 6 4 2

Outline contents for whole work

Volume 1 Methodology

- 1 Basics 1
- 2 Principles to follow 57
- 3 Grammar overview 92
- 4 Analysis, argumentation, and explanation 182
- 5 Terminology 214
- 6 Doing typology 242
- 7 Phonology 264
- 8 Lexicon 289
- 9 Field linguistics 309

Volume 2 Grammatical Topics

- 10 Grammatical word and phonological word 1
- 11 Distinguishing noun and verb 37
- 12 The adjective class 62
- 13 Transitivity 115
- 14 Copula clauses and verbless clauses 159
- 15 Pronouns and demonstratives 189
- 16 Possession 262
- 17 Relative clause constructions 313
- 18 Complement clauses and complementation strategies 370

Volume 3 Further Grammatical Topics

This is planned to include the following chapters

- Non-spatial setting
- Noun categorization devices
- Number systems
- Negation
- Reflexives and reciprocals
- Passives, antipassives, and general intransitivizing derivations
- Causatives
- Applicatives
- Serial verb constructions
- Pivots, and switch-reference
- Comparative constructions
- Noun incorporation
- Interrogatives
- Imperatives
- Culture and language, some correlations

Contents

List of tables and figures	x
How to read this book	xi
Preface	xii
Abbreviations and conventions	xiv
1. Basics	1
1.1. What this book is about	1
1.2. Linguistics as a branch of natural science	1
1.3. Prototypical patterns and variations on them	4
1.4. Grammatical labels	9
1.5. Limitations of a language	12
1.6. Why is each language the way it is?	14
1.7. Meaning and its organization in a language	22
1.8. A grammar as an integrated system	24
1.9. Grammar and meaning	27
1.10. The basics of grammar	35
1.11. Grammar and lexicon	47
Sources and notes	54
2. Principles to Follow	57
2.1. Writing a grammar	57
2.2. Always go back to the primary sources	64
2.3. Unfashionable issues	66
2.4. Avoid the fashion fads	70
2.5. Avoid sloppy terminology	75
2.6. How to learn linguistics	80
Sources and notes	90
3. Grammar Overview	92
3.1. Basic units: word and clause	92
3.2. The clause	93
3.3. The two major word classes, Noun and Verb	102
3.4. Phrases	106
3.5. Clause structure and word classes	110
3.6. Adjectives	112
3.7. Pronouns, demonstratives, and interrogatives	114
3.8. Syntactic specification of space and time	118

-
- 3.9. Marking of core and peripheral arguments 122
 - 3.10. Complement clauses 128
 - 3.11. The sentence 132
 - 3.12. Negation 137
 - 3.13. Morphology 138
 - 3.14. Derivations 149
 - 3.15. Non-spatial setting 152
 - 3.16. Noun classes and genders, and classifiers 155
 - 3.17. Number systems 158
 - 3.18. Definiteness 160
 - 3.19. Dependencies between grammatical systems 162
 - 3.20. Changing valency 165
 - 3.21. Topic and pivot 171
 - 3.22. Argument identity: reflexives and reciprocals 175
 - 3.23. Comparative constructions 177
 - Sources and notes 179
 - 4. Analysis, Argumentation, and Explanation 182
 - 4.1. What is linguistic analysis? 182
 - 4.2. Analyses should be clear and plausible 184
 - 4.3. Argumentation needed to justify an analysis 186
 - 4.4. Choosing between alternative analyses 191
 - 4.5. Different analyses for varying purposes 194
 - 4.6. Different analyses of similar data 196
 - 4.7. Intertwining of levels 199
 - 4.8. Encountering the unexpected 201
 - 4.9. Explanation 205
 - Sources and notes 213
 - 5. Terminology 214
 - 5.1. Grammar and lexicon 214
 - 5.2. Morphology 217
 - 5.3. Inflection and derivation 218
 - 5.4. Clitic, affix, and adposition 221
 - 5.5. Morphological types 226
 - 5.6. Syntax 228
 - 5.7. Markedness 235
 - Sources and notes 240
 - 6. Doing Typology 242
 - 6.1. Requirement for consistent analysis 243
 - 6.2. What should be compared 247
 - 6.3. Phonological typology 249
 - 6.4. Grammatical typology 251

- 6.5. Lexical typology 255
- 6.6. The question of sampling 257
 - Sources and notes 261
- 7. Phonology 264
 - 7.1. Phonemes 264
 - 7.2. Consonants 266
 - 7.3. Vowels 273
 - 7.4. Phonotactics and syllables 275
 - 7.5. Prosodies 279
 - 7.6. Stress 280
 - 7.7. Balancing parsimony and clarity 283
 - 7.8. Orthography 286
 - Sources and notes 287
- 8. Lexicon 289
 - 8.1. Organizing the lexicon 291
 - 8.2. Constructing a lexicon 296
 - 8.3. Structure of the lexicon 300
 - 8.3.1. Nouns 302
 - 8.3.2. Adjectives 304
 - 8.3.3. Verbs 305
 - Sources and notes 307
- 9. Field Linguistics 309
 - 9.1. The fundamentals 309
 - 9.1.1. What is linguistic fieldwork? 309
 - 9.1.2. Why do it? 309
 - 9.1.3. Poor reasons for doing fieldwork 310
 - 9.2. Getting started 311
 - 9.2.1. The ethics of fieldwork 311
 - 9.2.2. How to choose a language 311
 - 9.2.3. Fieldwork locations 314
 - 9.3. Working in the field 315
 - 9.3.1. How to do it 315
 - 9.3.2. What to get 318
 - 9.3.3. What to do 320
 - 9.3.4. Other fieldwork situations 325
 - 9.4. Making sure you have it right 326
 - 9.5. What not to do 326
 - Appendix 1 Describing the fieldwork situation 328
 - Appendix 2 Planning a fieldwork PhD 329
 - Appendix 3 Field methods courses 329
 - Note 330

Glossary 331

References 342

Combined indices for Volume 1 and Volume 2

Author Index 356

Language Index 364

Subject Index 372

List of tables and figures

Tables

- 1.1. Forms of 2nd person pronouns in languages with different number systems 10
- 1.2. Locational suffixes in Dyirbal 16
- 1.3. Examples of noun class membership in Dyirbal 28
- 1.4. Sample of verbs taking the four main types of complement clause in English 31
- 1.5. Complement clauses co-occurring with a selection of THINKING and LIKING verbs 33
- 3.1. Sample semantic types of the Verb class in English, and their roles 104
- 3.2. Marking of core arguments in Cashinawa 124
- 3.3. Linkage types, illustrated with sample markers in English 134
- 3.4. Samples of derivational processes in English 149
- 4.1. Main syntactic possibilities for word classes in Fijian 191
- 4.2. Possibilities for *-er* and *more* 202
- 6.1. Syntactic properties for languages 1, 2, and 3 243
- 6.2. Standard errors for different population sizes 259
- 7.1. A sample consonant matrix 267
- 7.2. Cardinal vowels 274
- 7.3. Consonant system at an earlier stage of Nyawaygi 285

Figures

- 1.1. Reference of tenses in multiple-past tense systems 23
- 3.1. Types of clause combining 94
- 3.2. Nominative–accusative system 122
- 3.3. Absolutive–ergative system 123
- 3.4. The nominal hierarchy 123
- 3.5. Split-S system 125
- 3.6. Fluid-S system 125

How to read this book

This book is, of course, designed to be read from first to last page. But other strategies are possible.

Chapter 1 outlines the approach followed throughout, and should be consulted first. The lengthy Chapter 3 provides an overview of grammatical structures and systems found across the languages of the world. Many of these topics are dealt with in more detail in the chapters of Volume 2 and of the projected Volume 3. Ideally, Chapter 3 should be studied at an early stage, although experienced linguists may choose to skim it.

The remaining chapters of Volume 1, and those of Volume 2, could be read in any order. However, recurrent themes are developed across chapters and maximal benefit will be obtained by reading the chapters in the order in which they were written.

Preface

For more than four decades I have been doing linguistics in the true sense of the word—undertaking immersion fieldwork, writing grammars, compiling lexicons. I've studied, in fair detail, more than two hundred published grammars, and consulted several hundred more. I have worked—by inductive generalization—on a number of topics in typological theory, and have read everything I could lay my hands on that is relevant to this endeavour. However, despite having been learning, learning all along the way, I feel that I know only a fraction of what I would like to know.

This book is a distillation of what I have learned thus far—the most satisfactory and profitable way to work, and what pitfalls to avoid. In short, how best to obtain reliable and satisfactory results which have scientific validity. Volume 1 sets the scene, with chapters on aspects of methodology. Volumes 2 and 3 then deal in fair detail with each of a number of grammatical topics.

The reader will find opinions expressed straightforwardly, without demur. Some of the things that are said may go against certain of the current 'fashions'. I do not expect others to agree with everything I say. But all the points made here have validity, and are worthy of serious consideration.

The languages I know best are those that I have worked on myself and published on—the Australian languages Dyrbal (1972, 1973, 1989b), Yidiñ (1977a, 1977b, 1991b), Warrgamay (1981), Nyawaygi (1983), and Mbabaram (1991c), plus Boumaa Fijian (1988), Jarawara from Brazil (2004a), and English (1991a, 2005a, 2005b). If some point can be illustrated from one of these languages then I do so, rather than using data from another language which I know less well. This applies especially to the general discussions in Volume 1. For points which do not occur in these languages, and for further exemplification of points that do, information from many other grammars is used.

Sources are sometimes included in the text but more usually in notes at the end of a chapter. It has not been thought necessary to quote sources for well-described languages such as Latin, French, German, Estonian, Turkish, Hebrew, Mandarin Chinese, Quechua, Swahili, Thai, and the like. Specific references are often not given for the languages I have worked on. If, say, an example is taken from Jarawara, the interested reader can easily consult my comprehensive grammar of that language (Dixon 2004a) to see how the matter under discussion fits into the overall linguistic system of the language. Sources are provided for information from other languages. There is a glossary of technical terms, included at the end of each volume.

There is today a fashion in linguistics—and no doubt in other disciplines as well—of what can be called ‘quotationitis’. That is, attempting to cite every single thing published on or around a topic, irrespective of its quality or direct relevance. Not unusually, quotations are provided from several sources which are contradictory in assumptions and import, without attention being drawn to this. I have used citations sparingly; these only reflect a small proportion of the grammars and general works which I have studied. The present work is conceived of as being like a well-organized garden; I have tried to avoid it degenerating into an impenetrable jungle.

At several places I mention the number of languages currently spoken across the world. The habit has arisen of quoting a figure of well over 6,000, which is the number of ‘language names’ listed in *Ethnologue* (Gordon 2005). This is put out by a missionary body with the main purpose of indicating where there is considered to be need for translation of the Christian Bible. The volume is uneven in scope and reliability, particularly as regards what is a language and what is a dialect (decisions on these questions frequently relate to policies concerning translation teams, and decisions may change as policies change—for one instance of this, see Dixon 2004c: 8). More than 200 languages are listed for Australia (many labelled ‘nearly extinct’ or even ‘extinct’), but 60 would be an optimistic estimate for the number which are still actually spoken (or else well remembered). The actual number of distinct languages currently in use across the world is no more than 4,000, quite likely a fair number fewer.

This book has been envisaged, planned, and written in close collaboration with my colleague Alexandra Y. Aikhenvald. We have discussed every topic, often many times. I have benefited from her grammars of Warekena (1998), Tariana (2003), and Manambu (2008), and from her typological studies (particularly 2000, 2004). I am the one who has written the book (and Aikhenvald would not necessarily agree with every single word in it) but the ideas, analyses, and generalizations are in very many instances our joint work.

Nick Enfield carefully read every chapter and provided the most useful comments, corrections, and suggestions. And I owe a considerable debt to the several score students and colleagues whose grammatical descriptions I have assisted with over the years, having learnt from each of them.

These volumes have been brought to fruition through the help and encouragement of John Davey, linguistics editor *sans pareil*. Of the several publishers I have worked with over almost five decades, Oxford University Press is, in every department, the most efficient and caring. John Davey exudes an enthusiasm which makes one feel valued and wanted, and works in a friendly and unobtrusive way to assist each author in realizing their potential.

And so, I cast my pebble upon the beach.

Abbreviations and conventions, for Volumes 1 and 2

Some abbreviations are used through the book (for example, A, S, and O), others only in chapters where a particular topic is being discussed (for example, RC for relative clause).

There are abbreviations employed in interlinear glossing of examples, such as *ERG* for ergative and *CLASS* for classifier. However, where an example is short, with plenty of room on the line, a full label *ERGATIVE* or *CLASSIFIER* is written out. It would be pedantic (and otiose) to insist on always employing *ERG* and *CL* when there is no spatial limitation which requires abbreviation. My aim, through the volumes, has been to try to be as reader-friendly as circumstances permit.

-	affix boundary
=	clitic boundary
'	stress (or accent)
1	1st person
2	2nd person
3	3rd person
A	transitive subject
ABS	absolutive
ACC	accusative
AN	animate
ART	article
AUX	auxiliary
CA	common argument (shared by main and relative clauses in a relative clause construction)
CC	copula complement
CLASS	classifier
CoCl	complement clause
COMP	complement clause marker
COMPL	completive

CONTIN	continuous
COP	copula
CS	copula subject
CTV	complement-taking verb (Chapter 18)
D	possessed (Chapter 16)
D	specific description in copula construction (§14.4)
DEC	declarative
DEM	demonstrative
DIM	diminutive
du, DU	dual
E	extension to core
ERG	ergative
exc	exclusive
F	focal clause (§3.11)
FEM, F, f, fem	feminine
FIN	finite
FUT	future
G	general description in copula construction (§14.4)
GEN	genitive
IMM	immediate
IMPERV	imperfective
inc	inclusive
INTERROG	interrogative
INTR	intransitive
LOC	locative
MASC, M, m, masc	masculine
MC	main clause
Mf	marker attached to focal clause (§3.11)
min	minimal
Ms	marker attached to supporting clause (§3.11)
NEG	negation
NOM	nominative
NON.FIN	non-finite

NP	noun phrase
nsg	non-singular
O	transitive object
∅	zero
PART	particle
PERF	perfect
PERFV	perfective
pl, PL	plural
POSS	possessive
PRED	predicate marker
PREP	preposition
PRES	present
R	possessor (Chapter 16)
R	specific referent in copula construction (§14.4)
REDUP	reduplicated
REL	relative clause (marker)
REP	reported
S	intransitive subject
S	supporting clause (§3.11)
Sa	‘active’ S, marked like A
sg	singular
So	‘stative’ S, marked like O
SUBORD	subordinate
SVC	serial verb construction
TAM	tense, aspect, and modality
TR	transitive
VCC	verbless clause complement
VCS	verbless clause subject
VP	verb phrase

Basics

1.1 What this book is about

In writing this book I have two aims. To provide an outline characterization of the structure of human language. And to provide a guide for those who wish to pursue the central business of linguistics—describing and analysing natural languages, and then by inductive generalization contributing to the typological theory summarized here.

The focus is on grammar, the organizational nucleus of every language. There is brief discussion of phonology in Chapter 7 (with reference to further sources), and some notes on how a dictionary/thesaurus (or lexicon) should profitably be presented are in Chapter 8. Grammar is built on the relations between items chosen from lexical word classes, so that throughout the grammar chapters there is continued reference to the structure of the lexicon.

1.2 Linguistics as a branch of natural science

There are a number of possible approaches to the study of language. That which is followed here treats linguistics as a natural science, on a par with geology, biology, physics, and chemistry. As Toulmin (1984: 382) states, ‘the task of science is to explain actual events, processes or phenomena in nature, and no system of theoretical ideas, technical terms, and mathematical procedures . . . qualifies as scientific unless it comes to grips with those empirical data at some point and in some way and helps to make them more intelligible’. The task of linguistics is to explain the nature of human language, through active involvement in the description of languages—each viewed as an integrated system—together with explanation of why each language is the way it is, allied to the further scientific pursuits of prediction and evaluation.

In order to master any scientific discipline, one must actively engage in it, commencing at the grass-roots level. Just as a biologist undertakes detailed observation and examination of the behaviour and nature of some animal or plant, so a linguist is concerned with investigation of the complete and complex phenomenon that is a language.

Anyone setting out to properly learn linguistics should choose a language which has not been (or has scarcely been) described, and whose speakers welcome cooperation with a linguist. If the language is still spoken on a daily basis the ideal course is to live (for, say, nine to twelve months) as a part of the community, being exposed to the language, gradually gaining proficiency and observing how it is used. The basic material for study will be texts recorded in the language—legends, narratives of recent events, instructions for planting, hunting, herding, and manufacture, and so on. These are recorded, transcribed (in terms of a suitable phonemic alphabet), and analysed.

Grammatical structures and rules are worked out inductively, on the basis of the textual corpus, from utterances observed as the community goes about its daily business, and from example sentences gathered during the construction of a lexicon. Hypotheses relating to grammatical organization must be worked out and then checked. The checking involves generating predicted sentences on the basis of the putative structures and rules, and putting them to speakers (within a suitable context). If they need to be corrected, the hypothesis will require adjustment, and then a further check must be carried out.

By continuing to work in this way, over a period of three years or more, a complete grammar of the language will be built up, each part relating to other parts within an integrated whole. As in geology or chemistry, description and explanation will be in terms of the established scientific theory of that discipline.

There is constant feed-back between theory and description. Each description is in terms of the established theory, and the theory itself is made up of interrelated inductive generalizations based on the descriptions provided in terms of it. As each new description is completed, it is likely to lead to the refinement or revision of some aspect or aspects of the theory.

Once a linguist has served their apprenticeship, as it were, by producing full documentation of a language, they may move on to the next stage, that of typological comparison. This involves seeking to add to linguistic theory by considering how some parameter occurs across a range of languages, and attempting new generalizations. Here a word of warning is in order. As a general principle, the necessary preliminaries should be mastered before embarking on any advanced task; for example, one should not try to write before acquiring the art of reading, or gibberish may result. Applied to linguistics, one should learn the art of analysing a language, and constructing a grammar, before embarking on theoretical generalizations based on examination of a selection of good grammars (which must be assessed for reliability, internal consistency, and probity).

There are some linguists, of a different ilk, who avoid the travails of field-work and do not themselves produce a grammar, lexicon, and text collection

for a previously undocumented language, but attempt straightaway to work on linguistic theory. This is rather like a biologist who has only observed animals in picture books (or perhaps in a zoo) and then proceeds to statements about the nature and habits of a particular animal, or about animals in general. Too often, people who haven't worked on a grammar themselves can't distinguish—on looking through the literature—between an adequate grammar and a poor one; they simply don't have the grass-roots experience to enable them to know what to look for.

This book deals with linguistics conceived as a branch of natural science, with a single cumulative theory which has recently come to be called 'basic linguistic theory'. There are other approaches to the study of language, which involve the postulation of a profusion of competing 'theories' (the term 'theory' here being accorded a quite different sense). This is reminiscent of the competing 'theories' in disciplines such as economics or literature.

Each of such (oft-labelled) 'formal theories' puts forward a few ideas concerning limited aspects of language, which they seek to confirm by looking at the relevant parts of just a few languages, each considered outside the context of the holistic system to which they belong. 'Formal theories' tend to build on some part of the cumulative basic scientific theory, a different part for each. There is seldom any attempt to write a complete grammar in terms of any 'formal theory'. Indeed, the proponents of such 'theories' state that this is not their aim; they wish to propose or examine a 'formal theory' for its sake alone. Occasional insights into the nature of language may be thrown up, and these are then incorporated into the scientific theory. But 'formal theories'—many emanating from the work of Chomsky and different generations of his students, although a number have come from different directions—tend to briefly fizz before being eclipsed by some new rival. In contrast, the basic linguistic theory outlined below has its origins in the pioneering work of Sanskrit and Greek grammarians between 3,000 and 2,000 years ago. It is being continually enhanced through descriptions of new languages, each revision making it able to characterize a little more fully the essential nature of language as a cultural trait of human beings.

The ways in which a language is viewed by proponents of 'formal theories' are reminiscent of the legend of the blind man and the elephant. One touched the animal's side and decided that it was like a wall, the second touched the tusk and likened it to a spear, while another handled the wriggling trunk which seemed to him like a snake. When one reads what different formalists say about a certain language, each concentrating on a different aspect, it is hard to recognize that the same language is under discussion. In contrast, practitioners of linguistics as a science—working in terms of basic linguistic

theory—have their eyes wide open, and perceive each language as a complete linguistic system.

There are four fundamental tasks for any science—description, explanation, prediction, and evaluation. This book is concerned mostly with the first two. Description deals with *how* a language is organized; for example, whether it has a system of tense or of gender, the nature of such systems, and the ways in which they fit into the overall grammatical fabric. Allied to this we must pursue explanation, and enquire *why*. What is the reason that one language has four genders, another just two, and a third language none at all? Why does one language have three future and five past tenses while another has no grammatical system directly relating to position in time?

In the chapters that follow there will be some hints at prediction. If a language has a gender contrast for the 2nd person singular pronoun ('you masculine' versus 'you feminine'), then it is likely also to have this distinction for the 3rd person singular pronoun ('he' versus 'she'). If there is no grammatical distinction between present and past time for a positive sentence, then neither will there be for a corresponding negative sentence. One may also essay a prediction of what is likely to happen over time: a language with a certain complex of mildly incompatible features is likely to rationalize through a predictable avenue of change. By and large, such time-line predictions fall outside my scope here.

Then there is the charged question of evaluation. It is an accepted procedure to evaluate the worth of different economic or political systems. We have outgrown the mistaken colonialist idea that some languages are significantly more 'primitive' than others. All languages are *roughly* equal in terms of overall complexity. But surely they are not all of *precisely* the same value. Might not some languages be better than others, for certain purposes? Is one language easier to learn than another? Does one language provide a superior framework for deep discussion of kinship relationships, or of subtleties of taste, or for assessing the worth of cattle herds, or for sports commentary, or for philosophical introspection?

The matter of evaluation has been scarcely aired by linguists; indeed, some consider it offensive to raise the topic. But within the context of linguistics as a natural science, such questions must be mooted. They will not be directly addressed in this book; but the parameters discussed here provide the where-withal for future evaluation of different languages in terms of specified criteria.

1.3 Prototypical patterns and variations on them

In essence, the grammar of any language is a network of interactive structures and categories. For example, a clause involves a predicate and a number of

core arguments—one for an intransitive predicate (for example, ‘laugh’), and two for a transitive predicate (such as ‘take’). Typically, the head of a transitive predicate may only be a verb. In some languages the head of an intransitive predicate can only be a verb, in others it may be either a verb or an adjective, and in a further set it may be a verb, an adjective, a noun, or a pronoun. A core argument may be realized by a noun or pronoun, which may be modified by an adjective.

Basic linguistic theory—the theory of linguistics as a natural science—consists in study and comparison of the grammatical patterns of individual languages. As a preliminary example, consider the ways languages have of marking possession of ‘knife’ (an object), ‘mother’ (a kin term), and ‘foot’ (a body part). Many languages have two different grammatical devices for showing possession, depending on what is possessed. For example:

- I. One method for ‘mother’ and ‘foot’, another for ‘knife’. This is found in Amele from Papua New Guinea and Maricopa, a Yuman language from Arizona, among many others.
- II. One method for ‘mother’ and ‘knife’, another for ‘foot’. Languages with this pattern include Nootka, a Wakashan language from southern British Columbia, and Dyirbal, from Australia.
- III. One method for ‘knife’ and ‘foot’, another for ‘mother’. Found in a fair number of languages including Mandarin Chinese and Ewe, from Ghana.

Comparison shows that although each of these languages has only two possessive constructions, basic linguistic theory needs to recognize three categories (using *x* to code one device and *y* for the other):

(1) SYSTEM	I	II	III
objects, such as ‘knife’, ‘canoe’	x	x	x
kin, such as ‘mother’, ‘husband’	y	x	y
body parts, such as ‘foot’, ‘eye’	y	y	x

We would then predict that there should be some languages, of a type IV, with three possessive devices, one for ‘knife’ and ‘canoe’, another for ‘mother’ and ‘husband’, and a third for ‘foot’ and ‘eye’. There are indeed such languages: they include Haida, from northern British Columbia, and Tachelhit, a Berber language, from Morocco. And there are, of course, languages like English which just have one way of marking possession: we say *my knife*, *my mother*, and *my foot*; this is type V.

Why do languages behave in such different ways, one having the same mechanism for objects and kin, another having the same mechanism for kin

and parts, and so on? It depends on how each language community views the world. For system III, kinship is regarded as different from all other kinds of possession. For I, a kin relationship is treated as something inherently possessed, like a body part. For II, a kin relation is treated as something different from ‘foot’ or ‘eye’, which are inherent parts of a person.

But surely, one might say, there are two kinds of kinship link. A blood (or ‘consanguineal’) relation such as ‘mother’ or ‘son’ is as inalienable as a body part like ‘foot’ or ‘eye’. In contrast, a relationship through marriage (called ‘affinal’), such as ‘husband’ or ‘mother-in-law’, is far from immutable, since a person can get divorced. One would expect blood relations to group with body parts, and relations by marriage with objects like ‘knife’. Such a system (type VI) is not common, but it does occur; for example, in Lango, a Nilotic language from Uganda.

The six systems can be summarized:

(2) SYSTEM	I	II	III	VI	IV	V
objects, such as ‘knife’, ‘canoe’	x	x	x	x	x	x
affinal relations, such as ‘husband’	y	x	y	x	y	x
blood relations, such as ‘mother’	y	x	y	y	y	x
parts, such as ‘foot’, ‘eye’	y	y	x	y	z	x

Note that we never get blood relatives marked like objects and affinal relatives shown in the same way as parts. This shows that there must be a semantic basis to a division within kin terms, where this occurs.

This is only one of the parameters concerning possession; others include the nature of the possessor, and the nature of the possessive relation (for example, whether permanent or temporary). A full account is in Chapter 16.

Example (2) presents the prototypical patterns of possession, according to the nature of the possessed item. This volume will summarize such prototypes across the gamut of grammars of human languages. There are, as would be expected, a range of deviations from the prototypical arrangements. To mention just one example, in the Austronesian language Gapapaiwa (spoken in Papua New Guinea), kin terms divide into two sets with respect to possession, but this does not relate to whether the relationship is by blood or through marriage. What we find is that kin who are in authority over one (e.g. grandparent, parent, elder same-sex sibling) are marked for possession like ‘knife’ and ‘canoe’, while those not in authority (including spouse, child, younger same-sex sibling) are marked like ‘foot’ and ‘eye’.

To explore every variant on the prototypes across the several thousand recorded languages would require a hefty volume for each grammatical category and construction type. What I aim for here is a summary of the prototypical patterns, and mention of just a few of the most interesting

deviations from them. If the reader, investigating some previously undescribed language, comes across a grammatical feature not covered here, they should in most cases be able to place it as a new kind of variant on one of the prototypical schemes described in this volume.

There is always a temptation to assume that the most prototypical patterns are those found in one's own language, or in languages with the greatest prestige or the largest number of speakers, or in languages which have been well described and are best known (including what is called 'traditional grammar'). Such temptation should be resisted. English and other well-known European languages are of type V in example (2), having a single type of marking irrespective of the nature of the possessed item. Some of the languages mentioned—with more complex systems, of types I, II, III, IV, or VI—are relatively unknown tongues, spoken by small speech communities (although Mandarin Chinese, which has more speakers than any other language, is of type III).

In fact, the most complex grammatical systems, with the most finely delineated instances of prototypical patterns, are typically found in languages spoken by small tribal groups, whereas many features of languages spoken over a wide region—or across the world, as with English and Spanish—are relatively simple and straightforward. This is further discussed in §1.6.

Besides describing typical language patterns, one must seek for explanation of them: why are things this way and not that? For example, the typical array of stop and nasal consonants is:

(3)		BILABIAL	APICO-ALVEOLAR	DORSO-VELAR
	VOICELESS STOP	<i>p</i>	<i>t</i>	<i>k</i>
	VOICED STOP	<i>b</i>	<i>d</i>	(<i>g</i>)
	NASAL	<i>m</i>	<i>n</i>	((<i>ŋ</i>))

Many languages include all nine phonemes in a 3×3 system. But if any are missing, they are likely to be from the dorso-velar column. Many languages have *p*, *t*, *k*, plus *b*, *d*, *g*, but only two nasals, *m* and *n*. And if a stop is missing, this is most likely to be *g*.

Before trying to identify a reason for these gaps, we can also consider vowel systems. The most common system is—like the consonant array in (3)—symmetrical:

(4)		FRONT	BACK
	CLOSE	<i>i</i>	<i>u</i>
	MID	<i>e</i>	<i>o</i>
	OPEN	<i>a</i>	

A variant on this pattern is to have four vowels, and most frequently this involves having just one phoneme covering both high back, [u], and mid back, [o]. This could be written with *o* (as here) or, equally well, with *u*. The system is:

(5)		FRONT	BACK
	CLOSE	<i>i</i>	
	MID	<i>e</i>	<i>o</i>
	OPEN	<i>a</i>	

The point to note is that variants on the prototypical stop-and-nasal template and on the prototypical vowel template both involve a loss of some distinctions among sounds articulated at the back of the mouth. There is a physiological explanation. The tongue needs to move further and to do more work to pronounce a dorso-velar consonant or a back vowel. Although there is a tendency for phonological systems to be symmetrical, there are asymmetrical systems and in such cases there are likely to be more choices available in an area of articulatory ease and fewer in a region of relative articulatory effort.

Biologists establish taxonomies of plants and animals according to their salient characteristics; for example, with or without a backbone, then shape of jaw, arrangement of teeth, and so on. Why cannot linguists produce a comparable taxonomy of languages? There are two main reasons why not. First, a language is a complex of interwoven features each of which can, to a large extent, vary independently of the others. There is no hierarchy of features in terms of which a satisfying and meaningful taxonomy of languages could be proposed. The second reason is that a language is never static; there is always fluidity and change—between one generation and another, and between one community and another.

A common type of change (but there are many others) is when a language shifts a certain feature from one prototype to another. The commonest vowel system is that in (4) with five members; the next commonest has three members, high front *i*, low *a*, and high back *u*. Languages may shift between these profiles. For example, Proto-Indo-European is reconstructed to have had a standard five-vowel system; this developed into a standard three-vowel system in Sanskrit, when *e* and *o* fell together with *a*. In contrast, Proto-Austronesian had a variant of a three-vowel system, *i*, *a*, and *u* plus central vowel, *ə*. Its descendant Proto-Oceanic developed a prototypical five-vowel system through sequences *ay* and *aw* becoming *e* and *o* respectively (and with *ə* also becoming *o*).

A language can move, in a certain feature, from one prototype to another. Or it can diverge from a standard pattern and then, by a later (and different

type of) change, move back to it. The prototypical syllable structure is (C)V(C), where an obligatory vowel may be preceded and/or followed by a consonant. Some languages only allow (C)V, with no final vowel, others CV(C) where an initial consonant is obligatory, and so on. There are, in a minority of languages, widely divergent syllable structures, such as the CCCVCCC pattern in English (illustrated by *strengths*).

In languages which allow a syllable to begin with a vowel, one generally finds vowel-initial forms greatly outnumbered by those which commence with a consonant. What is almost never found is syllable structure V(C) where there *cannot* be an initial consonant. A rare example of this is in the Australian language Olgolo, which underwent a change whereby the initial consonant of each word was lost: **guda* ‘dog’ became *uda* and *minha* ‘animal’ became *inha*. Such a non-prototypical syllable structure would be predicted to be unstable, and not to persist indefinitely. In fact it is being eliminated in an interesting way—through grammar coming to the rescue of phonology. A specific noun is typically preceded by the appropriate generic noun, as in *inha anbu* ‘animal possum’; a word-final vowel is dropped when followed by a word beginning with a vowel, so that this is said as *inh anbu*. Then the generic noun *inha* reduces to be a prefix *nh-*, attached to the specific noun. We get *nh-anbu*—which is CV(C)CV—and prototypical syllable structure (the easiest to articulate) is regained.

1.4 Grammatical labels

No two languages are precisely the same, in any feature. Although the same labels are used for describing grammatical categories in different languages (if they were not, there would be no science of linguistics) they have a slightly different signification for each language.

The label ‘accusative’ is typically used for an affix marking that a noun phrase (NP) is in direct object (O) function. In Quechua (spoken in the Andes), an accusative suffix goes onto the last word of the NP, but in Latin it goes onto every word in the NP. In Turkish an NP is marked by the accusative suffix only when it has definite reference. Beyond this central—and defining—function of marking O function, an accusative affix may have additional roles in the grammar of an individual language. In Latin, for example, accusative may also mark a length of time—as in *tōtam noctem dormiūi* (‘all:ACCUSATIVE night:ACCUSATIVE I:slept’) ‘I slept the whole night’—and accusative is required after a number of prepositions, including *circum* ‘around’.

The size of a grammatical system can vary between languages, and then so will the reference of labels used to describe terms in the system. Consider the grammatical system of number, which may apply to pronouns, nouns, etc. We

TABLE 1.1. Forms of 2nd person pronouns in languages with different number systems

NUMBER OF INDIVIDUALS REFERRED TO	1	2	A FEW	MANY
Akan (Kwa family, Ghana)	singular <i>wo</i>	plural <i>mo</i>		
Kayardild (Tangkic family, Australia)	singular <i>nyingka</i>	dual <i>kirra</i>	plural <i>kilda</i>	
Longgu (Oceanic branch of Austronesian family, Solomon Islands)	singular <i>oe</i>	dual <i>amurua</i>	paucal <i>amolu</i>	plural <i>amu</i>

can examine forms of the 2nd person pronoun ‘you’ in three languages, set out in Table 1.1. In each language, singular involves reference to just one person. Where there is a dual, as in Kayardild and Longgu, this refers to two people. Longgu has two further ‘you’ forms, *amolu*, referring to a few people (this is called paucal), and *amu*, referring to many people.

Now consider the label ‘plural’ as it is used in these three number systems. The meaning of plural is complementary to the meanings of the other terms in the system. In Akan it means ‘not singular’; that is, more than one. In Kayardild it means ‘not singular or dual’; that is, more than two. In Longgu it means ‘not singular, dual, or paucal’; that is, more than a few. Basically, ‘plural’ is the residue term for everything not included under more specific number statements. (There are some languages with a four-term number system which includes ‘trial’, referring to just three people, rather than paucal; ‘plural’ then refers to more than three. But, cross-linguistically, paucal is much more common than trial in a four-term system.)

One must be aware that the signification of ‘plural’ depends on the size and nature of the number system it belongs to. To say that Akan, Kayardild, and Longgu each has a distinct plural 2nd person pronoun, in contrast to English, which lacks one, would be a misleading statement. One should instead say that, unlike English, these three languages have a number distinction for the 2nd person pronoun, but the size of the system (and thus the referent of the term ‘plural’, used for the residue term in the system) varies. When comparing grammatical terms with the same label between languages, one must always pay attention to the system to which the term belongs, in terms of a holistic view of the grammar.

Labels must be used in a consistent way within a grammar. Consider Jarawara, an Arawá language from Brazil, where there are distinct number systems in different parts of the grammar. These include:

- (a) **PRONOUNS.** For each person, there is a form referring to one individual and another form referring to more than one; for example *tiwa* ‘you (one)’ and *tee* ‘you (more than one)’.
- (b) **SOME VERBS** have different forms according to whether the referent of the intransitive subject (S) is one person, or two, or more than two. For example:

‘lie on the ground’ *-homa-* S refers to one person
 mata -na- S refers to two people
 soo -na- S refers to more than two people

Thus, pronouns have a two-term number system (like Akan), for which the labels singular and plural are appropriate, and verbs have a three-term system (like Kayardild), for which the labels singular, dual, and plural are appropriate. But if this principle were followed, ‘plural’ would be used with two quite different meanings—‘more than one’ for pronouns, and ‘more than two’ for verbs—within the grammar of one language. This would be likely to cause confusion, and should be avoided.

The most appropriate course is to use the terms singular, dual, and plural (here referring to more than two) for verbs, and then singular and non-singular for pronouns. Non-singular covers dual and plural, and refers to more than one. That is:

NUMBER OF INDIVIDUALS REFERRED TO	1	2	MORE THAN 2
2nd person pronoun	singular <i>tiwa</i>	non-singular <i>tee</i>	
S for verbs such as ‘lie on ground’	singular <i>-homa-</i>	dual <i>mata -na-</i>	plural <i>soo -na-</i>

Basic linguistic theory is concerned with comparing similar phenomena between languages and to achieve this it is convenient to apply the same label to similar phenomena in different languages. Two widely used names for types of possession are ‘inalienable possession’, where the possessed item has an inherent connection with the possessor, and the complementary label ‘alienable possession’. It is said, for instance, that each of Amele, Nootka, and Lango has a contrast between inalienable and alienable possession. But these languages have systems of types I, II, and VI respectively

from example (2), so that the labels have rather different scope in the three languages:

	INALIENABLE POSSESSION	ALIENABLE POSSESSION
I Amele	'foot', 'mother', 'husband'	'knife'
II Nootka	'foot'	'mother', 'husband', 'knife'
VI Lango	'foot', 'mother'	'husband', 'knife'

Employment of the terms 'inalienable possession' and 'alienable possession' in this way, in the grammars of these individual languages, is perfectly defensible and appropriate. Inalienable possession always relates to possession of parts, such as 'foot', and alienable possession always to possession of separable objects, such as 'knife'; these can be regarded as defining criteria for use of the labels. Each term may have additional, extended reference in a specific language. (There is a fuller discussion of this in Chapter 16, mentioning further kinds of extension.)

An informed grammar written in terms of basic linguistic theory should describe what happens in the language under study and then relate this to cross-linguistic typological parameters, such as those set out in example (2). That is, in a grammar of Amele, for example, one should say that the label 'inalienable possession' covers body parts (the defining criterion for use of this label) and kin terms are also marked in the same way (mentioning that in many other languages kin lexemes are not grouped with part nouns in this manner).

1.5 Limitations of a language

Languages vary enormously in the degree of complexity they have in a certain area. For example, there are no grammatical genders at all in Turkish or Hungarian, two in French and Hebrew, three in Russian and German, four in the North-East Caucasian language Lak, five in Supyire from Mali, and at least seven (ten by some counts) in Swahili. In the last sections we looked at the variety of ways of marking possession in Haida and Tachelhit—system IV in example (2)—and described the four-term number system in Longgu. A brief sample of large systems associated with some other grammatical categories follows.

- (a) **Case.** Finnish has a generous array of fifteen cases, for marking the function and meaning of an NP; these include specifications such as 'towards the inside of', 'towards the outside of'.
- (b) **Tense.** The language of the Western Torres Strait islands, between Australia and New Guinea, is reported to have three future tenses ('beyond tomorrow', 'tomorrow', 'later today') and five pasts ('just completed', 'earlier today', 'last night', 'yesterday', and 'before yesterday').

- (c) **Evidentiality.** In about one-quarter of the world's languages, every statement must specify the type of source on which it is based, choosing from a system of evidentials. In some languages the system has only two terms (often, whether or not directly observed). But in others there are more choices available. For example, Tucano, from north-west Amazonia, has a five-term evidentiality system: one must specify whether the evidence is (i) visual (seen), (ii) non-visual sensory (heard, smelt, tasted or felt), (iii) inferred from direct evidence, (iv) assumed on the basis of general knowledge or common sense, or (v) reported by someone else.
- (d) **Imperatives.** Tuyuca, spoken in Colombia, has eight varieties of imperative: (i) 'do at a future time'; (ii) 'make sure that something which should be done is being done'; (iii) 'do lest something bad might happen'; (iv) 'do what a third person has ordered you to do'; (v) imperative of invitation; (vi) imperative of imploring; (vii) intimate imperative, used with close friends and animals; (viii) general imperative.

The more choices there are in a grammatical system, the greater the communicative power of the language—put bluntly, the richer the language. Why, then, don't all languages have fifteen cases, eight tenses, five evidentiality values, eight imperatives, seven genders, three ways of marking possession according to the nature of the item possessed, and a four-term number system? Why do we have a high degree of complexity of one grammatical category in this language, a high degree of complexity of another category in that language, but never all kinds of complexity in a single language? Why aren't all languages—especially those whose speakers think of themselves as sophisticated and highly civilized—developing more and more choices in their grammatical systems, so that they should move towards an all-encompassing grammatical profile with maximal choices in all categories?

Why? Because, it seems, the human brain can only tolerate a limited degree of complexity in a grammar. A few grammatical categories may have a substantial array of choices available, but not many (let alone all) categories. It appears that we just cannot cope with an 'ideal grammar', which would provide maximal information on every front.

There is an analogy to the knowledge and specialization of people. One person can be a highly skilled lawyer, familiar with every interstice of legal precedent and practice. One can be a doctor, acquainted with the recognition and treatments of innumerable diseases. Other *métiers* can be journalist, plumber, composer, carpenter, farmer, linguist, electrician. Some may combine one specialism with a smattering of another—a professional lawyer may also be a hobby farmer, a journalist may be able to render first aid, a linguist may

be able to undertake some electrical work. But no one could combine, in full fashion, three or four—let alone all nine—of the specialisms just listed (and there are a myriad more besides).

Can we thus infer that any human language is only a partial means of expression? As grammar goes, this is undoubtedly so. But there is also, of course, the vocabulary. Every language has a stock of at least several thousand words. What is not expressed through grammar can be said, in some fashion or other, through the lexicon. Whereas the language of the West Torres Strait islands has grammatical marking for three future tenses, in English (which has minimal grammatical expression of tense) all one can do is specify using lexemes: *after tomorrow* or *tomorrow* or *later today*. In Tuyuca one could say ‘Sit in the shade!’ using the ‘do lest something bad might happen’ imperative ending, with no need to specify precisely what the unpleasant consequence might be. To express this in English requires two clauses, something like *Sit in the shade so that you don’t get burnt by the sun!* And so on.

One *can* say just about anything in every language. If there is no appropriate grammatical system, then lexemes will be needed. The fundamental difference between languages with distinct grammars is in what one *must* say. In English one can just tell someone *I’ll do it*, with vague and general reference to future time. But in Western Torres, reference to future time *has to* involve choice of the appropriate tense ending, indicating which period of the future is being referred to. This is an *obligatory* specification. Each language has its own set of obligatory specifications, this being what provides its linguistic ‘character’, as it were.

1.6 Why is each language the way it is?

Languages differ in the nature and size of the grammatical categories which they include. It is almost as if there were a bag which contained every known grammatical category, in varying sizes, with each individual language putting in its hand, blindfold, and picking out as many items as it thinks its grammar can handle.

But in fact it could not be like this. It surely cannot be an arbitrary matter that one language has a large tense system in its grammar, another language a small one, and a third language no tense distinctions at all; and so forth. There surely must be—at least in part—some principles explaining why the grammar of Cantonese is the way it is, and similarly for German, and Eskimo, and every other language.

Just as a person’s character stems from a combination of nature and nurture, so does the profile of a language come from the interaction of three main factors. In brief:

(a) **Genetic history.** Each language is the lineal descendant of a succession of ancestor languages; it is likely to retain some (but not all) of the characteristics of its immediate parent. Languages are always changing but they change bit by bit, not all at once. For example, English is the descendant of proto-Germanic, whose own ancestor was proto-Indo-European. Some of the irregularities in modern English are a direct reflex of forms reconstructed for proto-Indo-European (for example, verb forms *sing* and *sang*). Others are a continuation of irregularities introduced by umlaut processes at an intermediate stage (for example, *foot* and its plural *feet*). English has lost some of the distinctions made by its ancestors; for example, different forms for singular and plural of the 2nd person pronoun. And it has introduced some new distinctions of its own; for example, one can say (with a slight difference of meaning) either *He looked at the baby* or *He took a look at the baby* and either *She strolled in the park* or *She took a stroll in the park*.

(b) **Inter-language contact.** Speakers of a given language generally have social relations—involving trade and/or marriage and/or ceremonials—with speakers of neighbouring languages, so that some people acquire a degree of competence in one or more of these adjacent tongues. As a consequence, there is a tendency for the grammar of the language to gradually get more similar—in some ways—to the grammars of neighbouring languages. For example, speakers of Tariana, from the Arawak family, moved into the Vaupés River basin in north-west Brazil, and entered into social relations with speakers of Tucano and other languages of the Tucanoan family, who were already in the region. In the Vaupés area, one must marry someone from a different language group; each Tariana man marries a woman speaking a Tucanoan language, and vice versa. Originally, Tariana had little grammatical marking of evidence, but due to close inter-language contact, it has developed an intricate system with five evidentiality values, just like the systems in the Tucanoan languages (that in Tucano is described in §1.5).

(c) **Environment, life-style, and beliefs.** Any language naturally adapts so that it relates directly to the habitat of its speech community, and the ways in which its speakers live and think.

First, the geographical terrain in which a language is spoken may be reflected in its grammar. Lak, from the North-East Caucasian family, has three demonstratives—*aha* ‘this’ (close to speaker) and two forms for ‘that’ (far from speaker):

- hava ‘that’, on the same level as speaker
- ho ‘that’, higher or lower than speaker

In the mountainous terrain in which Lak is spoken it is communicatively appropriate to specify both the distance of some object from the speaker

TABLE 1.2. Locational suffixes in Dyirbal

-bayji	short distance downhill	-dayi	short distance uphill
-bayja	medium distance downhill	-daya	medium distance uphill
-bayju	long distance downhill	-dayu	long distance uphill

-balba	medium distance downriver	-dawa	medium distance upriver
-balbu	long distance downriver	-dawu	long distance upriver
	-guya	across the river	

	-bawal long way (in any direction)		

(close or far) and also its relative height with respect to the speaker. If the Lak community moved to flat country, the grammatical specification of height for demonstratives would stand a high risk of becoming lost.

Dyirbal is spoken close to the north-east coast of Australia, in a mountainous rain forest region with heavy rainfall and many rivers. A noun phrase generally includes an article-like form (marking the gender of the head noun in the phrase) to which can be added one of a set of forms denoting the location of the referent of the noun with respect to where the speaker is. In central and southern dialects the system has twelve terms, set out in Table 1.2.

The grammar thus enables a speaker to supply information concerning whether something they are talking about is up or down hill or river, and how far up or down. Note that *-u* indicates a long way, *-a* a medium distance, and (just for the hill suffixes) *-i* a short distance.

However, the most northerly dialect, Ngajan, has a more restricted set with just six terms:

-baaji	downhill	-dayi	uphill

-baabu	downriver	-dawu	upriver
	-guya	across the river	

	-baandu long way (in any direction)		

Ngajan is spoken in a tablelands region which is a little less hilly than the region inhabited by other Dyirbal-speaking tribes. In keeping with this, it has lost the short/medium/long distance distinction. Interestingly, the original 'short distance' forms are used for 'uphill' and 'downhill', but the 'long distance' forms for 'upriver' and 'downriver', there having been no 'short distance' choice here. (It will be seen that Ngajan has undergone a phonological

change whereby *ay* and *al* have been replaced by *aa* before a consonant; and it has a slightly different form for ‘long way (in any direction)’.)

The social organization of a community may be a further factor in mediating the make-up of a grammar. For the indigenous peoples of Australia (as for groups in other parts of the world), the kin relationship between two people is criterial for determining how they should behave towards each other and what their reciprocal obligations are. A major distinction typically concerns generation level; this is reflected in the pronominal system of the Lardil language. There are two forms of each dual pronoun, which linguist Ken Hale referred to as ‘harmonic’ and ‘disharmonic’:

	HARMONIC	DISHARMONIC
1st person inclusive dual, ‘you and I’	ŋaku-rri	ŋaku-ni
1st person exclusive dual, ‘he/she and I’	nya-rri	nya-anki
2nd person dual, ‘you two’	ki-rri	nyi-inki
3rd person dual, ‘them two’	pi-rri	rni-inki

The harmonic set is used for two people in the same generation (e.g. two brothers, two sisters, or brother and sister) or two generations apart (e.g. grandparent and grandchild) and the disharmonic set for people one or three generations apart (e.g. parent and child, great-grandparent and great-grandchild). Having generational level as an obligatory specification in the grammar serves to draw attention to concomitant social responsibilities and ways of behaving.

Lardil was spoken by a small tribe of just a few hundred people. Larger language communities often have a strong sense of social hierarchy, and their grammars may incorporate reference to it. For example, whereas English has a single 2nd person pronoun, *you*, Bengali has six, distinguishing singular and plural (as most languages do) and also the relative social status of speaker and addressee(s). There are three choices: an honorific form is used for someone of higher status towards whom respect must be accorded; an intimate form is used with a close friend; and an ‘ordinary’ form is employed in other circumstances:

	2nd person singular pronoun (nominative case)	2nd person plural pronoun (nominative case)
INTIMATE	tui	to-ra
ORDINARY	tumi	tom-ra
HONORIFIC	apni	apna-ra

(There are also ordinary and honorific—but not intimate—forms for 3rd person pronouns. As would be expected, 1st person pronouns have a single form.)

The unusual situation in formal English, with a single 2nd person pronoun, has its historical origin in a system which paid attention to social status. The

original plural pronoun, *you*, came to be used in formal contexts and the original singular pronoun, *thou*, in informal or intimate circumstances (rather like *vous* and *tu* in French today). Then the situations deemed appropriate for the use of *thou* diminished, until finally *you* was employed for all reference to (one or more) addressee(s). Many speakers of English are in fact creating a new 2nd person plural with forms like *yous* and *y'all* (but these are resisted by 'educated' folk, and scarcely used in writing).

The way in which a community goes about its daily living may be reflected in the grammar. Speakers of Yidiñ, in Australia, were hunters and gatherers; there is a single classifier, *mayi*, used to refer to the more than a hundred edible plants which are gathered for food. In contrast, speakers of Jacaltec, from Guatemala, are agriculturalists. They have one classifier, *ixim*, referring just to the garden crop which is their main source of food—corn (or maize); there is another classifier, *te?*, covering all other plants (whether edible or not).

Having some sort of religious belief is common to all human groups. The nature of the belief can affect structural possibilities within a grammar. Kate Burridge has studied an Old Order Mennonite community in Canada which follows an extreme form of Christianity that subordinates self to the 'will of God'. As a consequence, it is believed that an individual should not want something for himself; in keeping with this, the dialect of Pennsylvania German spoken in this community has eliminated the use of *wotte* 'want' with an infinitive complement, as in 'I want to come'.

Sadly, linguists have thus far paid scant attention to why each language is organized in the way that it is; only a few hints have been provided concerning how culture helps determine grammar. Intensive study of this question is a high priority for research, making future work in linguistics such an exciting and attractive proposition. Why, for instance, do about one-quarter of the world's languages have an obligatory grammatical system of evidentiality, which is missing from the other tongues?

Speakers of languages with evidentiality are generally aware of this, especially when interrelating with people who speak a language which lacks the category. For instance, the Tariana know a little Portuguese, for communicating with non-Indians. They comment on what they perceive as a deficiency of Portuguese—that it is vague and lacks precision, enabling one to make a statement without having to specify the type of evidence on which it is based.

A critical parameter of variation between communities is how much people expect to be told by their neighbours. It is reported that when speaking Malagasy the convention is to provide as little information as possible, with vagueness being an accepted norm. Other societies have an opposite attitude,

expecting that maximal information should be provided. Anyone acting in a vague manner is here regarded as stupid or retarded. Dyrirbal society is of this type, and as a consequence there is no general verb 'know'. Rather than a vague statement such as English *I know where the money is*, one must specify how one has acquired such information, utilizing a specific statement such as 'I saw where the money is' or 'My father told me where the money is'. Speakers of Tariana have a similar distaste for under-specification, but whereas Dyrirbal speakers have to use appropriate lexical items, the Tariana language has evolved an obligatory grammatical system of evidentiality, whereby for each statement there is a suffix to the verb indicating whether what is being described was seen or heard or inferred or assumed or reported.

There is unlikely to be a unique explanation for any component of a grammar; there will instead be an intersection of diverse kinds of motivation. One of perhaps several reasons for a grammar including the category of evidentiality is the importance attached to explicitness in the culture. And there may, in turn, be a reason for this. For example, in Amazonian society there is held to be an explicit cause for everything that occurs; nothing happens accidentally or 'naturally'. As a case in point, if anyone dies—other than a very young child or a very old person—some magical means must have been employed, and the victim's relatives will seek out and punish the alleged perpetrator. So as not to be blamed for something they had no responsibility for, a speaker is careful always to be absolutely explicit about what they have done. This is expedited by the grammar requiring a specification of the evidence on which each statement is based.

In lexicon, each language reflects the activities of its speakers. Cattle herders in East Africa have a rich variety of terms to describe different kinds of beast, and the variety of hues they exhibit. A community which focuses on weaving will have specific terms for warp and weft, ribbing, twill, and selvage. In large, highly articulated communities, people will vie for superiority—in sports, and in social activities—and have an array of lexemes in connection with this. But in the languages of many small tribal societies there are no terms 'compete' or 'win' or 'lose', simply because people do not subscribe to such notions. And there may be no verb 'order' in a society where it is not the practice to issue direct commands to others. (There may be other ways of getting people to do things, such as drawing attention to an acknowledged social obligation, or a polite (but firm) request.)

The cultural perception of an ethnic group may, in part, determine what they consider it proper to say and not to say. In Korean, for instance, a statement of physical or psychological state may only be used with a 1st person subject. One can say of oneself 'I am feeling cold/sad/glad/bored' but it is not considered felicitous to attribute such feelings to anyone else. That is, one

should not overtly infer how someone else is feeling. (It is, of course, possible to question someone else: ‘Are you feeling cold/sad/glad/bored?’)

All over the world, speakers conceive of a language as consisting of its vocabulary, with little regard paid to grammar. In the 1920s, Kemal Atatürk established a programme to strengthen Turkish society by using Roman rather than Arabic script, and by eliminating Arabic loan words. A century earlier there had been a similar drive to rid Hungarian of Latin loans. A major aim of the Tariana people in north-west Amazonia is to keep their language free of foreign influence. They have a strict interdict against using any words from the Tucanoan languages of their region. But—due to intermarriage and multilingualism—many Tucanoan structural patterns and grammatical categories (such as evidentiality) have been taken into the grammar, and these pass unnoticed.

I have worked for several decades with speakers of the Girramay and Jirrbal dialects of the Dyirbal language; these have around 80 per cent of their lexicons in common and show just a few grammatical differences. In the 1960s, speakers of the two dialects kept them fully apart. Over the next quarter-century, the mixed Jirrbal-Girramay community adopted a merged dialect, whose grammar was entirely Jirrbal with most lexical items also being Jirrbal plus just a few from Girramay. Bessie Jerry was a Girramay elder, proud of her own tribal heritage. In the 1980s and 1990s, Bessie would teach me the original Girramay words—‘ear’ is *garba*, whereas Jirrbal use *maŋa*; ‘bathe’ is *bujay*, as against *ŋabay* in Jirrbal; and so on. But Bessie put these Girramay words into sentences using entirely Jirrbal grammar; *-ñ* for future tense rather than the original Girramay suffix *-jay*, and Jirrbal *ŋaja* for the 1st person singular pronoun, intransitive subject form, instead of the proper Girramay form *ŋayba*. Like speakers of Hungarian, Turkish, and Tariana—and of just about every other language—Bessie identified Girramay in terms of its vocabulary, paying little or no attention to grammatical forms and constructions.

It is relevant to enquire whether there is any correlation between type of society and the profile of its language. There does appear to be. Languages with the most complex word structure (morphology) tend to be spoken by smallish speech communities made up of only a few hundred or a few thousand (or, at most, a few tens of thousand) people. Some of these languages also have rich possibilities for syntax: many construction types and a wide variety of subordinate clause types. All in all, they have the most intricate and difficult grammars that we know. (Such languages are not easy to learn; but then few people outside the immediate speech community wish to learn them.) They also have a fair-sized vocabulary, with up to 5,000 distinct lexemes, perhaps more in some cases.

Languages that are spoken by all members of a large socio-political group (such as a nation, or a state in a country such as India) tend to have less overall grammatical complexity. That is, they are easier to learn than many small tribal tongues; and in fact many people do have the need to learn such languages. There may be a quite restricted morphology but a fairly complex syntax. And every such regional or world language has a considerable vocabulary. Leaving aside specialist technical registers such as medical and legal terminology, the normal working vocabulary of an educated speaker of English approaches 20,000 words, and that of uneducated people certainly more than half that amount.

European colonizers of the nineteenth century put about the belief that small tribal groups—considered primitive in their material culture—spoke primitive languages with little or no grammar. Yet the Europeans experienced great difficulty in attempting to learn the local languages, whereas the tribal people soon acquired fluency in the language of the invaders. How could this be so?

In fact, the belief that many dark-skinned non-European peoples speak ‘primitive languages’ is wholly in error. Typically, a small ethnic group will have a sizeable lexicon (although nothing like so big as that of a major world language) and an elaborate grammar; it is because of this that the languages are difficult to learn. The invaders’ language was in many cases less daunting, certainly in terms of morphology. In most parts of the world, tribal people were multilingual, adept at learning other languages, and they soon mastered the rudiments of the strangers’ tongue.

Finally, brief mention should be made of creoles. A creole typically grows out of a pidgin, which is a code used for limited communication—typically concerning trade—between groups of people who do not know each other’s language. A pidgin is no one’s native tongue and involves no more than a few hundred words linked by simple grammar. It may come to be used as first language by children and is then speedily augmented in size and complexity, thus becoming a creole, with several thousand words and a certain measure of syntax. But, of the well-documented creoles, none equals the complexity—or the communicative power—of a non-creole language (which will have developed as the vocal emblem of its community over eons of time).

There were originally around 800 languages in Papua New Guinea, most of daunting complexity. To an increasing extent, children are not learning their traditional language, but instead utilize the national creole, Tok Pisin, which is easy to learn. Martin Kumwau—an elderly man who is a fluent speaker of Gala and of three other languages from the Sepik region—bemoans the incursion of Tok Pisin since he considers it inadequate for decent communication, there

being rather few words, each with a vague general meaning. ‘Tok Pisin is not a real language,’ he opines, ‘just a short-cut.’

1.7 Meaning and its organization in a language

Someone has an idea in their mind and attempts to let some other person know what it is, or to record it for their own or someone else’s future reference. This is the central function of language—the communication of meaning. Meaning must be accorded a major focus of attention in any linguistic study.

The ineluctable functions of language are to describe events and circumstances, to outline plans or instructions, to detail social obligations or religious convictions. But besides normal talking, all human groups have more refined uses of language. Recitation in poetry or song will convey a message in an aesthetically pleasing and socially significant manner.

The way in which things are expressed—choice of words and grammatical constructions, mode of pronunciation—may identify the language user’s ethnic or social group, and trigger a particular attitude towards them from listeners. But this is the outside dressing, as it were, to the central role of language—the reason for its existence—the task of conveying a meaning from speaker to addressee(s).

Meaning in the raw is not nicely ordered and arranged. Think of a field strewn with leaves and nuts and stones and twigs and fruit; it needs someone to arrange the useful bits into appropriate piles, and to discard the debris. This is the sort of role which language assumes with respect to meaning. There are two aspects to each language—a lexicon which classifies things, and a grammar which organizes things. But in different languages these tasks are performed in diverse ways.

Consider the section of the human body which extends from shoulder to fingers. In English there are six simple nouns referring to parts of this section: *shoulder*, *arm*, *elbow*, *wrist*, *hand*, and *fingers*. (More complex labels can of course also be provided, e.g. *upper arm*, *forearm*, *little finger*.) In the Australian language Dyirbal, this section of the body is described by seven simple nouns:

banjal	top of shoulder (from edge of neck to point of shoulder)
jurru	bottom of shoulder (upper arm just below point of shoulder)
garrgal	upper arm (below <i>jurru</i>)
buru	elbow
manju	forearm (from below elbow to above wrist)
girman	wrist
mala	hands and fingers

WEST TORRES	YIMAS	JARAWARA
just completed	just completed, earlier today or last night	from just completed up to about two months ago
earlier today		
last night	yesterday	from about two months to about two years ago
yesterday		
before yesterday	two to about five days ago	more than about two years ago
	more than about five days ago	

FIGURE 1.1. Reference of tenses in multiple-past tense systems

It will be seen that names in the two languages do not exactly correspond. *Shoulder* in English relates to *banggal* and *jurru*, *arm* to *jurru*, *garrgal*, and *mangu*, while both *hand* and *fingers* correspond to *mala*. Other languages employ different ways of delineating the body. In Fijian, for instance, there are just two simple nouns, *taba* ‘upper arm and shoulder’ and *liga* ‘forearm, wrist, hand, and fingers’.

Simplified somewhat, a grammar consists of a number of closed systems—categories such as tense, gender, and evidentiality—and a number of construction types, or ways of relating together words into phrases, clauses, sentences, and utterances. Some categories effect a classification, similar to lexemes. This is illustrated in Figure 1.1, by comparison of three languages, each of which have three or more distinct past tenses: the West Torres Strait language (already mentioned in §1.5), Yimas from New Guinea, and Jarawara from southern Amazonia. (The reference of terms in the tense systems is relative. In Yimas, for instance, all one can say with certainty is that the reference of the lowest tense in the column is a further time in the past than the reference of the term above it; the actual time involved varies with circumstances.)

It will be seen that the most recent past form in Yimas corresponds to three past tense choices in West Torres. And the most recent past tense in Jarawara corresponds to four of the past tenses plus part of the fifth in West Torres, and to three and part of a fourth in Yimas.

We can now look at types of relative clause as a brief example of how languages differ in the construction types they utilize. A canonical relative clause modifies the head noun in an NP, in a similar way to an adjective. It serves to restrict the reference of the head noun, also like an adjective:

compare *the tall man* and *the man who lives upstairs*. In some languages—such as Kambera, spoken in Indonesia—all relative clauses have a restrictive meaning. However in others, including English, there are also non-restrictive relative clauses, modifying a noun that already has unique reference (and cannot be further restricted). Suppose that I have two daughters and one son. Sentence (1) includes a restrictive relative clause, identifying which daughter I am talking about.

(1) My daughter who lives in New York is an artist

However, since I have only one son, the relative clause in (2) is of the non-restrictive variety.

(2) My son, who lives in Los Angeles, is a singer

There are a number of differences between the two kinds of relative clauses in English. One is that non-restrictive clauses have contrastive intonation, shown by commas in writing, as in (2). In Persian, the suffix *-i* is required on the verb of a restrictive relative, but not on the head of a non-restrictive clause.

Volumes 2 and 3 of this work are devoted to a detailed examination of how different grammars organize their resources in varying ways. Relative clauses are discussed in Chapter 17.

1.8 A grammar as an integrated system

A grammar is in some ways like a machine, an assemblage of cogs, levers, belts, pulleys, and rollers, connected together to function as one integrated system. Move one component and this will, in some way, affect every other part. Take away one bit (making necessary adjustments for its absence) and the role of each other part is likely to be affected. This analogy has some validity but ultimately it fails in that a language is not mechanical; it is a naturally evolving entity, rather than something purposely constructed.

A more appropriate model might be a densely knit forest, with all manner of plants, vines intertwined around limbs, ferns sprouting from trunks, birds and insects spreading seeds, moisture evaporating from leaves and returning as rain on the soil, to provide nurture and advance the cycle of activity. Here each component depends on others, though not in a rigid way. Excise a large tree (or it may fall unbidden) and others will move in to take its place.

A grammar is in some ways like the machine model, in some ways like the jungle model, and in some ways like neither. But the analogies should serve to make the most important point—that every grammar is an integrated system. Each part relates to the whole; its role can only be understood and appreciated in terms of the overall system to which it belongs.

There are some linguists who do just a little work on a language that is not well described, looking at perhaps a single construction type. This is bad science. One cannot appreciate the role of relative clauses in language X without relating and comparing them to other kinds of subordinate clauses, to the ways of marking syntactic function, and so on.

The pioneer linguist Ferdinand de Saussure criticized scholars who studied the history of a part of a language, dissociated from the whole to which it belongs. He insisted that linguists should study the complete system of a language at some point in time, and then examine how the entire system changes over time. Saussure's pupil Antoine Meillet (1926: 16) is responsible for the aphorism: '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. (Proponents of formal theories, who look at isolated bits of language for some particular issue, naturally contravene this fundamental principle.)

The statement of a grammar is a product of analysis. And every analytic decision has to be justified by criteria relating just to *this* grammar, not to meaning (considered apart from this particular grammar) or to what happens in the grammars of other languages.

Within every grammar a number of major word classes can be posited. Their recognition must be based on grammatical criteria from the language under analysis. The nature of the criteria is likely to depend on the structural profile of the language. For Latin, we recognize three major word classes, with the following properties:

class A, inflects for case and number

class B, inflects for case, number, and gender

class C, inflects for tense, aspect, mood, person, and number

For English, we also recognize three major word classes and here the criteria are:

class X, takes suffix *-ing*

class Y, may be immediately preceded by an article and need not be followed by another word

class Z, may be immediately preceded by an article and is then followed by another word (either one from class Y or another word from class Z)

Now the lexemes belonging to each of these classes show a certain range of meaning. They also have typical behaviour in filling functional slots within a clause. It is because of a measure of similarity of meaning and function that

we may identify word classes between languages, and use the same label for them:

Noun—classes A and Y. May be head of a noun phrase which can be in subject or object function within a clause. The class includes words referring to concrete objects (and their parts), such as ‘tree’, ‘stone’, ‘star’, ‘woman’, ‘foot’, ‘water’, ‘axe’.

Adjective—classes B and Z. Typically, modifies a noun. Includes words relating to states, typically dimension (such as ‘big’, ‘little’), age (‘new’, ‘old’), value (‘good’, ‘bad’), and colour.

Verb—classes C and X. Occur as head of a predicate. Includes words referring to actions, such as ‘jump’, ‘sit’, ‘burn’, ‘eat’, ‘laugh’, ‘talk’, ‘see’.

Note that the criteria employed are different for the two languages. Latin has a rich morphology but no strict ordering of words within a clause. English has rather little morphology but fairly strict rules of ordering. (In English some—but not all—nouns take plural suffix *-s*, and some—but not all—adjectives have comparative and superlative forms.)

A very important point to note is that although the word classes have similar semantic content between languages, the full ranges of meanings they cover are never identical. The central members—as exemplified above—are likely to correspond (although there is no guarantee that every single one will). But there can be considerable variation among non-central members. For example, the idea of needing to eat is expressed through noun *hunger* in English, by verb *ēsurio* in Latin, and by adjective *ḡamir* in Dyirbal. (Interestingly, English has a derived adjective, *hungry*, formed from the noun; and Latin also has an adjective, *ēsuriens*, derived from the verb.)

Kin relationships such as ‘mother’ and ‘father’ are nouns in most languages but in some—for example, the Yuman languages of southern California—they are expressed by verbs ‘be mother of’ and ‘be father of’. (These words are, after all, describing a relationship between parent and child.) The number ‘two’ is an adjective in many languages but a verb in others (for instance, Jarawara).

The moral of all this is that it is not possible to decide which class a word belongs to in a given language solely on the basis of its meaning. If this were the case then the word for ‘wanting to eat’, or for a male parent, or for ‘two’ would be in the same word class for every language, which they are not. In one Amazonian language we find the unusual circumstance of ‘good’ being a verb, ‘be good’. It takes the same inflections as verbs such as ‘see’, ‘laugh’, and ‘jump’, having quite different grammatical behaviour from adjectives (such as ‘big’, ‘little’, ‘new’, ‘old’, and ‘bad’). Someone who was actively working on this language once told me that ‘good’ *must* be an adjective because ‘everyone

knows that in every language “good” is an adjective, since it describes a state’. This person simply hadn’t mastered the basic principles of linguistic analysis.

It does appear that every language has open classes of words which can be felicitously named noun, verb, and adjective, although the defining criteria vary between languages, as do their full semantic and functional ranges. (These topics are discussed in detail in Chapters 11 and 12 of Volume 2.) Most languages have complement clauses, which can substitute for an NP in object slot. (Compare English *I know [the truth about Mary]_O*, with an NP as O (transitive object) argument, and *I know [that Mary is a secret agent]_O*, with a complement clause in this slot.) But, as described in Chapter 18, some languages lack complement clauses per se, instead utilizing a variety of other construction types as ‘complementation strategies’.

Just as with word classes, one must be careful to analyse construction types on the basis of internal criteria for the language under study. I read one grammar recently where a range of quite different construction types were all labelled ‘complement clauses’. This was puzzling until I realized that they were all the translation equivalents into the language of complement clauses in German, the native language of the author of the grammar. A complement clause construction in one language is not necessarily translated by a complement clause construction in another; this part of that grammar is thus deeply flawed.

What a linguist must avoid is translating the sentences of some exotic language into English or French or Spanish or German or Chinese, and then analysing the translations. These translations are simply an aid to understanding. In every aspect of grammatical work, analytic decisions must be made within the language under study, based on criteria internal to the system of the language.

1.9 Grammar and meaning

The world teems with a myriad of contrasts and distinctions, but a grammar provides a rather limited scheme of organization. The consequence is that one system in a grammar may be used simultaneously to represent a number of distinct contrasts of meaning. This can be illustrated by noun classes in Dyirbal, and then by complement clauses in English.

Every noun in Dyirbal belongs to one of four noun classes (or genders), marked by an article-like determiner. According rough labels to the classes, the determiners are—in absolutive case form—*bayi* masculine, *balan* feminine, *balam* edible plants, and *bala* neuter. The first step is to establish the

TABLE 1.3. Examples of noun class membership in Dyirbal

class I, <i>bayi</i>	class II, <i>balan</i>	class III, <i>balam</i>	class IV, <i>bala</i>
men	women		parts of the body
kangaroos, possums	dingo		meat
most fishes	some fishes		
some birds	most birds		
moon	sun, stars		
storms, rainbow			wind
	anything connected with fire or water		stones, mud
some spears	some spears		some spears
	some trees	all edible fruit and vegetables	most trees and vines with no edible parts
	fighting ground		place, hill

noun classes, and which class each individual noun belongs to; the criterion here—internal to the grammar—involves co-occurrence with a determiner. On completing this step, we can study the membership of the four classes, as summarized in Table 1.3.

Class III is fairly straightforward—non-flesh foodstuffs. Membership of the other three classes appears quite heterogeneous—most animals and fishes are in class I but some are in class II; most birds in class II but some are in class I; some spears in class I, some are in class II and others in class IV; most trees with non-edible parts are in class IV but some are in class II; and so on. The idea occurs that perhaps there is no overall general principle to class membership. But this would mean that a child would have to learn the class of each noun on an individual basis, like a cipher. Languages do not work in this way. There is a semantic basis to each part of every grammar. (There may well be some exceptions, irregularities that have to be learnt one by one, but these are always a fairly minor encumbrance.)

The key to solving this puzzle consists in *not* commencing with the membership of each noun class and trying to see what is common to the list (in fact, nothing is). The alternative—and rewarding—course of action is to begin with examination of semantic distinctions in the world, and cultural perceptions of the language community, and see how these are mapped onto the grammatical category of noun classes.

The principles of noun class assignment in Dyirbal can largely be explained in terms of (a) certain basic concepts associated with the classes; and (b) two culturally motivated rules for transferring class membership.

The basic concepts associated with the classes are:

Class I (determiner *bayi*): animateness; (human) masculinity

Class II (determiner *balan*): (human) femininity; water; fire; fighting

Class III (determiner *balam*): non-flesh food

Class IV (determiner *bala*): none (this is a residue class, available for everything else)

The first transfer rule is: if some noun has characteristic X, on the basis of which its noun class membership would be expected to be decided, but has—through belief or legend—some association with characteristic Y, then it will belong to the noun class corresponding to Y (rather than that corresponding to X).

Humans are always specified for sex; other animate beings are not. Thus all (nouns referring to) human males (for example ‘boy’, ‘Aboriginal doctor’) are class I, and all human females are class II. Animals and fishes are prototypically class I, by virtue of their animacy. However, birds are believed to be the spirits of dead human females and are, by the transfer rule, class II. Certain birds are excepted from this (and are not included under the generic term *balan dundu* ‘bird’). These birds have, as individuals rather than as a class, a role in dreamtime creation legends: thus the three species of willy wagtail are believed to be men, and are class I. The spangled drongo is in legend the bringer of fire (from the clutches of the rainbow-snake); fire is in class II and so is this bird. And similarly in half a dozen other instances.

Anything connected with water and fire (including light and the stars) is in class II. But the moon and sun are, in legend, believed to be husband and wife. By the transfer rule, the moon is class I, for masculine, and the sun class II, for feminine. Wind, having none of the characteristics ‘animateness’, ‘fire’, ‘water’, or ‘fighting’, is in the residue class, IV. But storms and the rainbow are men in legend, and are thus class I.

Things concerned with fighting are, as a rule, class II—most fighting implements and the fighting ground itself. Thus fighting spears (which are also used for hunting game such as kangaroos) are class II. But multi-pronged spears, used solely for spearing fish, are (like fishing lines) in class I, the same class as fish. And big short spears—used for neither fighting nor fishing—are, like the sticks used for digging up yams, class IV.

The second transfer rule is: if a subset of a set of nouns has some particular important property which the rest of the set does not have, then the subset may be assigned to a different noun class from the rest of the set, to mark this property. The property is most often ‘harmfulness’.

Fishes are generally class I, but two particularly dangerous fishes (which can harm a bather), the stone fish and the toad fish, are—by the second rule—in

class II (note that these fishes are not included under the generic term *bayi jabu* ‘fish’). Trees, bushes, and vines with no edible parts are in class IV, except for two harmful plants, the stinging tree and the stinging nettle vine, which are in class II. In each of these instances, the harmful subset is in class II, whereas the rest of the set is in class I (for fishes) or class IV (for plants). Most birds are in class II; however, hawks—the only birds which eat other birds—are, by the second rule, in class I.

This account has listed only some of the nouns involved; similar explanations apply to others. The class-concept correspondences and the two rules do provide a simple and efficient explanation for the general organization of Dyrbal nouns into the four noun classes. There are, however, some assignments which are at present without explanation (for example, why is dingo in class II?). This is what might be expected. For any part of a grammar, it is generally possible to provide a principled semantic explanation for the way it is; but there are usually just a few points for which explanation is lacking (irregularities, or exceptions). These may be items which originally had a semantic basis, but then lost this as the language evolved. Future change is likely to eliminate the irregularities, producing a more rational system. (But, just as some regularities are being normalized, others are likely to be evolving. Since a language is always in a state of change, it is never likely to completely tidy up the semantic rationale underlying its grammar.)

We can now consider complement clauses in English. There are four basic types, one marked by clause-initial *that*, one by *-ing* on the complement clause verb, and a couple by *to* before the verb. They are illustrated in

- (3) Mary remembered [that she had fed the cat]
- (4) Mary remembered [feeding the cat]
- (5) Mary remembered [to feed the cat]
- (6) Mary remembered [the cat to be greedy]

Each complement clause has a distinctive meaning, as can be seen from (3–6). A *THAT* clause, as in (3), refers to an activity or event as a single unit, without any reference to its internal composition or time duration. In contrast, an *ING* clause, as in (4) refers to an activity as extended in time, relating to the way in which it unfolds. In (3), Mary just remembers the fact of feeding the cat; in (4) she recalls the details of this—opening the packet of cat food, emptying it onto the dish, spilling some on the floor, and so on.

Sentence (5) is an instance of a Modal *to* complement clause, referring to someone becoming involved in the activity; it has a similar meaning to a *THAT* clause which includes a modal (here, *Mary remembered that she should feed the cat*). In a Modal *to* construction, main and complement clauses can have

TABLE 1.4. Sample of verbs taking the four main types of complement clause in English

THAT	ING	Modal TO	Judgement TO
admire	admire	forget	consider
assume	consider	hate	forget
consider	dislike	know	imagine
dislike	enjoy	learn	know
forget	favour	like	learn
hate	forget	love	remember
imagine	hate	remember	
know	imagine		
learn	like		
like	love		
love	remember		
remember			
suppose			

different subjects, with that of the complement clause then being introduced by *for* (for example, *His mother remembered for John to fill in the entry form*). A quite different type of complement clause, a Judgement TO type, is illustrated in (6); here the subject of the main clause offers a judgement concerning the subject of the complement clause (these must be different); no *for* can be included in a Judgement TO clause.

The next step is to list the verbs which take each of the four varieties of complement clause. The full lists would each run to several hundred items; a sample is provided in Table 1.4. Some verbs occur in all four columns of Table 1.4, some just in three or two or one column. Just looking down each column, there is no one feature common to all verbs in the list. This is the same conclusion we reached on looking down the lists of members of the noun classes in Dyirbal, in Table 1.3. If one tries to look for a link between these grammatical properties and the semantics of their members and approaches this from the grammatical side, nothing emerges. The revealing course, as with Dyirbal noun classes, is to start by considering the meanings of verbs and see how these are mapped onto the grammar, with respect to the meanings of complement clauses.

As described in §1.11, the words in the lexicon of a language naturally fall into a number of sets, which can be called 'semantic types'. All the words in each type have a common element of meaning, and they all share certain grammatical properties. The sample of verbs in Table 1.4 comes from two semantic types associated with the verb class, THINKING and LIKING. We can

first consider the THINKING type. There are a number of subtypes, each with slightly different meanings and complement clause possibilities.

THINKING type

- (a) Verbs which refer to someone's mind just focusing on some person, thing, state, or happening; they include *consider* 'think about some actual or possible state of affairs (and its consequences)' and *imagine* 'think of something as if it were true'. Verbs in this subtype take THAT, ING, and Judgement TO complement clauses; for example, *I imagined [that Mary won the prize]*, *I imagined [Mary's winning the prize]*, *I hadn't imagined [Mary to be so clever]*.
- (b) Where there is some doubt as to whether what is being thought about is true. Verbs include *assume*, *suppose*; they only take a THAT complement.
- (c) Where someone has in mind, or tries to get in mind, something about the past. Verbs include *remember* and *forget*, accepting all four varieties of complement clause.
- (d) Where someone is aware of some fact, or body of information, or method of doing something. Verbs include *know* and *learn*, taking THAT, Judgement TO, and Modal TO (but not ING) complement clauses.

It will be seen that all verbs in the THINKING type occur with a THAT complement clause; the subtypes vary as to which (if any) other types of complement clause they take.

LIKING type

All verbs in this type may occur with an ING complement clause referring to some habitual or durative activity. Some verbs may also take a THAT clause referring to the fact that something happens. A smaller group may also take a Modal TO clause, referring to the potentiality of something happening. It is hard to draw firm distinctions within this type (as was done for THINKING verbs). We can roughly distinguish three sets:

- (a) Verbs with the most general meanings—including *like*, *love*, and *hate*—which can take THAT, ING, and Modal TO complement clauses. For example:
 - (7) Mary likes (it) [that John plays the clarinet]
 - (8) Mary likes [John('s) playing the clarinet]
 - (9) Mary likes [John to play the clarinet]

TABLE 1.5. Complement clauses co-occurring with a selection of THINKING and LIKING verbs

	THAT	ING	Modal to	Judgement to
THINKING type				
(a) consider, imagine	✓	✓	–	✓
(b) assume, suppose	✓	–	–	–
(c) remember, forget	✓	✓	✓	✓
(d) know, learn	✓	–	✓	✓
LIKING type				
(a) like, love, hate	✓	✓	✓	–
(b) dislike, admire	✓	✓	–	–
(c) enjoy, favour	–	✓	–	–

Sentence (7) could be used if Mary thinks every boy should play an instrument, (8) if she likes listening to him play it, and (9) if she is pleased for him to play it because it takes his attention away from video games.

- (b) Verbs—such as *dislike* and *admire*—which have a slightly more restricted sense and would not generally be used to refer to some potentiality of action. They can be used with THAT and ING complement clauses but scarcely with the Modal TO variety; a sentence such as **I dislike [John to play the clarinet]* is not felicitous.
- (c) Verbs—including *enjoy* and *favour*—which relate to an unfolding activity and are basically restricted to an ING complement clause.

The complement-clause-taking possibilities of these THINKING and LIKING verbs are set out in Table 1.5. This array codifies—and our discussion explains—what verbs (with which meanings) occur with what complement types (with which meaning). There are, as may have been noticed, further matters of detail. For example, LIKING verbs often take *it* before a THAT clause.

Tables 1.4 and 1.5 have dealt with a small selection of those verbs which take complement clauses in English. Others come from the semantic types ATTENTION (verbs such as *see* and *hear*), DECIDING, SPEAKING, ANNOYING, ACTING, HAPPENING, COMPARING, RELATING, BEGINNING, TRYING, HURRYING, DARING, WANTING, POSTPONING, MAKING, HELPING, SEEM, and MATTER. (A full account is in Dixon 2005a.) But this sample well illustrates what is involved, and the way in which explanation can be provided for the lists in Table 1.4.

Each of these illustrations shows how a large number of meaning contrasts are mapped onto a restricted set of grammatical choices. The four noun classes in Dyirbal code masculine/feminine, animate/inanimate, edible/inedible, flesh/non-flesh food, harmful/non-harmful, whether something is

related to fighting, and all manner of cultural associations. The four main varieties of complement clauses in English relate to differences of meaning across a score of distinct semantic types of verbs. For example, compare the meaning difference between *I like* [*to eat mangoes*] (but scarcely ever get the opportunity) and *I like* [*eating mangoes*] (and indulge in this to excess), and that between *I tried* [*to play the piano*] (but couldn't get the hang of it at all) and *I tried* [*playing the piano*] (for a couple of years, but then decided I had better things to do with my time).

It is appropriate to summarize the story thus far.

1. The communication of meaning is the main reason for the existence of language. Describing and explaining how meaning is encoded in language is the prime task of linguistics.
2. Different languages express meanings—within their grammars and lexicons—in different ways. One cannot, from the meaning of a word, say what its grammatical status is in any language.
3. A grammar is an integrated system in which every part ‘holds together’ with every other part. The statement of a grammar is a product of analysis (not something god-given, which can be uniquely discovered) and each analytic decision must be justified by explicit criteria internal to the grammar. The hallmark of a good linguist lies in considering alternative solutions to a problem, assessing the pros and cons of each, and deciding which is the most appropriate solution for a particular purpose. (Other views of ‘linguistic analysis’ are discussed in Chapter 4.)
4. Once the grammar has been constructed, the meanings associated with its categories and construction types are investigated. In many instances, a variety of diverse meaning distinctions are coded into a single grammatical category. Regarded from within the grammar, the meaning basis of a grammatical construct will not be evident. Looked at from the outside, it can be perceived how a number of different types of meaning contrasts are superimposed on a single system in the grammar.

In essence, a grammar is an abstract system of interlocking elements. However, it must be mapped onto what is often called ‘surface syntax’—a sequence of sentences, each made up of clauses, in turn made up of phrases, these being made up of words, which may be made up of morphemes, each with a phonological form, which has phonetic realization. These constituents must, perforce, be pronounced in a certain order. But this is simply the *realization* of an underlying set of relations. It is the underlying relations themselves which reveal the basic working of the grammar, and it is this which must be the prime object of study for linguists. The following section explains the underlying

basics of grammar, and how these may be realized. It also shows how it is a severe error to attempt an analysis of surface structure, or to posit underlying structures which have the same basic nature as surface structures.

1.10 The basics of grammar

An effective way of learning how to do something is by seeing how not to proceed. I will here illustrate a number of unrewarding approaches to linguistic analysis, in each case explaining how to avoid a particular pitfall.

- (a) English has a considerable number of phrasal verbs, lexemes consisting of a simple verb form plus a preposition, where the meaning of the phrasal verb cannot be inferred from the meanings of simple verb and of preposition. The phrasal verb *hand over* ‘deliver’ can be used in either of two constructions, with no essential difference in meaning:

(10) John handed the documents over

(11) John handed over the documents

That is, the object NP, *the documents*, can come before or after the prepositional part, *over*, of the phrasal verb.

Now some linguists have suggested that (11) should be taken as the underlying order. The reason for this is that in (11) the two parts of the phrasal verb occur together and make up a continuous constituent. Then the variant in (10) is derived from (11) by moving the preposition to the right over the object NP, making *hand—over* into a discontinuous constituent. The principle appears to be that things should be as neat as possible in underlying structure, although they can be messed up, in specific ways, in surface structure.

But if the object is a pronoun rather than a full NP, only the first alternative is possible; one can say (10′) but not (11′):

(10′) John handed them over

(11′) *John handed over them

Under the interpretation just mentioned, for a pronominal object the surface representation is necessarily different from the postulated underlying structure, in that the preposition *must* be moved to the right. This surely casts doubt on the usefulness of such an approach.

Now consider another phrasal verb *go over*, meaning ‘examine’ or ‘rehearse’. We find:

(12) *John went the documents over

(12′) *John went them over

(13) John went over the documents

(13') John went over them

Here the form of *go* and the preposition *over* must be continuous, as in (13) and (13'); sentences such as (12) and (12'), in which the phrasal verb is discontinuous, are unacceptable.

This poses severe difficulties for the proposed analysis, which maintains that (11) is the underlying structure for *hand over*. True, the underlying structure for *go over* would be (13), the same as the surface structure. But why should it be that the *over* of *hand over* optionally moves to the right over a non-pronominal object and obligatorily over a pronominal object, whereas the *over* of *go over* may never move, over any variety of object?

There is in fact an alternative analysis which explains these facts in perspicuous fashion. It involves recognizing a number of different kinds of phrasal verb; each type being mapped onto surface structure in a distinctive way:

- (i) The 'Vp - ' type, such as *go over* - (and many others such as *take after* -, *count on* -, *set about* -, etc.), where the preposition (p) must follow the verb (V), and the object, shown by -, immediately follows the p.
- (ii) The 'V - p' type, such as *hand - over* (and very many others, including *take - off*, *hand - on*, *set - up*) where the canonical surface structure is for p to follow the object, as in (10) and (10'). However, the p can be moved to the left, over a full NP, as in (11). The p may not be moved to the left over a pronoun, so that (11') is unacceptable. This is because an object pronoun is typically a clitic attached to the preceding verb—/hánd=ðəm/ in (10')—and it is not permitted to move a preposition so that it intrudes into a verb-plus-object.pronoun.clitic sequence.

It is without doubt more satisfactory to say that a preposition may move to the left over a non-pronominal NP, for a phrasal verb of type (ii), than—under the analysis first mentioned—to say that the preposition may move to the right over a non-pronominal object and must move after a pronominal object for *hand over*, but never moves for *go over*. Statement of a general rule is always to be preferred, rather than ad hoc specifications for individual lexemes.

There are in fact pairs of phrasal verbs, involving the same basic verb and preposition, which differ just in that one is type (i) and the other type (ii). For example:

TYPE (i)

turn on X 'attack X'

see through X 'understand the true nature of X'

TYPE (ii)

turn X *on* 'excite X'

see X *through* 'ensure that X (e.g. a job) is satisfactorily completed'

And there are four other kinds of phrasal verb. We find (iii) the ‘Vp’ type, taking no object, e.g. (*the rain*) *set in*, and (iv) the ‘V – p –’ type, where one object comes between V and p and a second object after p, e.g. (*John*) *held (it) against (Mary)*. We also encounter phrasal verbs similar to types (i) and (iv) but with a double preposition: (v) ‘V pp –’, as in (*John*) *took up with (a new girlfriend)*; and (vi) ‘V – pp –’, as in (*John*) *put (his failure) down to (nerves)*.

The error in the first analysis described for (10–11) was an assumption that there should be an underlying structure the same as one of the possibilities for surface structure, with elements arranged in a fixed order. And that if there are two alternatives in surface structure—one with a continuous and the other with a discontinuous phrasal verb—then the continuous alternative should be taken as ‘underlying’, since it is simpler. When further facts were examined, this was seen to be the less preferred analysis.

But the major fault was in assuming that an underlying representation should involve a sequence of elements in fixed order, as must be the case in surface structure (in order for the sentence to be coded into phonological form, and spoken). Underlying a sentence is a network of abstract elements, interwoven with each other in terms of grammatical dependencies (and not arranged in any sequential order).

- (b) In English, subject, predicate, and object must occur in a fixed order, as in *Caesar warns Brutus*. One could say *Brutus warns Caesar* but the meaning is quite different; one could *not* say **Caesar Brutus warns* or **Warns Caesar Brutus*, etc.

But in Latin subject is shown by nominative case-ending and object by accusative case (nominative *Brutus* becomes *Brutum* in the accusative). In

- (14) *Caesar Brutum monet* ‘Caesar warns Brutus’

the ordering of words is not needed for showing grammatical relations, since the function of each word is shown by its form. In fact, the words in (14) can be permuted in any of the six possible orders, with no essential change in meaning:

- (15) *Caesar Brutum monet* *Caesar monet Brutum* *Monet Caesar Brutum*
Brutum Caesar monet *Brutum monet Caesar* *Monet Brutum Caesar*

The underlying grammar of the sentences in (15) is that the transitive verb *monet* ‘warns’ has Caesar as its subject (A) argument and Brutus as its object (O) argument. This is exactly the same underlying grammar as the English sentence *Caesar warns Brutus*. The two languages differ in the way these underlying relations are mapped onto surface structure. In English subject and object are shown by their ordering before and after the verb, there being no

mark on the noun itself to show function. In Latin, grammatical relations are shown by case-endings, and the words themselves may be in any order. (The order actually used in any instance will depend partly on the organization of the discourse within which the sentence occurs, and partly on the whim of the speaker.)

There are linguists who believe that each sentence should be assigned an underlying structure which should have a similar profile to surface structure. Such people have typically worked mostly (or entirely) with English or some similar language where word order is fairly fixed in surface structure. They posit an underlying structure with fixed order of elements for English, and also for all other languages, including those like Latin which allow any order in surface structure.

For Latin it is typically said that the underlying order is subject–object–verb (since this is the statistically most common order in surface structure). There then has to be a specification that the words can in fact be *scrambled into any order*. But there is in reality no need to say anything about ordering in Latin, in either underlying representation or in surface structure. To insist that there is an underlying ordering but that it need not be followed seems rather like a legislature passing a law and then saying that no one need observe it.

There is a continuum ranging from languages with fairly fixed ordering, such as English, to those like Latin where virtually all ordering possibilities are permissible. Many languages lie part-way along the continuum, with some ordering restrictions but fewer than in English. There are, in fact, more languages towards the Latin end of the scale than towards the English end (albeit that most of the other best-known languages are like English).

These two examples—of phrasal verbs *hand – over* and *go over* – in English, and of how one may say ‘Caesar warns Brutus’ in Latin—have been included to illustrate the point that the underlying grammatical characterization of any piece of language is a set of interrelated grammatical relations (as will be fully illustrated in Chapter 3). These will be mapped onto surface structure, and the words in surface structure must be placed in sequence in order that they can be spoken. But it is unnecessary, mistaken, and unrevealing to posit a linear ordering of elements in the underlying representation for any language, whether or not there are any fixed ordering constraints in surface structure.

- (c) Just as lexemes such as *hand – over* can have discontinuous mapping onto surface structure, so may a grammatical element.

A considerable number of languages, across the world, require the verb in a main clause to include one of a set of bound pronominal markers referring to the subject of the clause (there may be a further set of bound markers

referring to the object). For instance, Tiwi—spoken on the Tiwi islands, just off Darwin in north Australia—has a system of eight subject prefixes to the verb, including:

- (16) 1st person singular, ‘I’ *nguw-*
 3rd person singular masculine, ‘he’ *aw-*
 3rd person singular feminine, ‘she’ *amp-*

Now consider the following sentences with intransitive verb *-apa-* ‘eat’.

- (17a) *nguw-apa* ‘I eat’
 (17b) *ngiya* *nguw-apa* ‘I eat’
 (18a) *aw-apa* ‘he eats’
 (18b) [*awarra kiyijini jarrangini*] *aw-apa* ‘that small
 that(m) small(m) buffalo(m) buffalo eats’
 (19a) *amp-apa* ‘she eats’
 (19b) [*angilawa arikulanga mutika*] *amp-apa* ‘my big cat eats’
 my big(f) cat(f)

In Tiwi, as in other languages with bound pronouns, a verb can make up a complete sentence—as in the (a) examples here—since it includes a pronoun providing some information about the subject. If further information about the subject is required, the basic pronominal data is expanded by an NP preceding the verb, as in (18b) and (19b). In (17a), the identity of the subject is fully specified by the pronominal prefix *nguw-*; the free form pronoun *ngiya* ‘I’ may be added, in the NP slot—as in (17b)—to further emphasize this.

The interesting feature of the (b) sentences here is that the subject is shown twice, by an NP and by a pronominal prefix. Some linguists attempt to analyse surface structure and pose the following question:

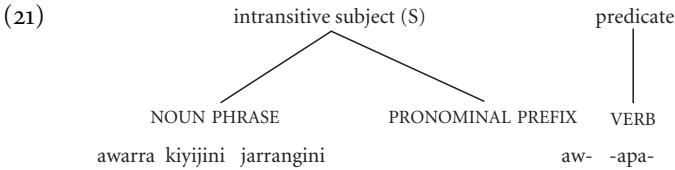
- (20) What is the main marker of the subject in these (b) sentences—the NP or the verbal prefix?

Two different answers are possible for this question:

- (i) The bound pronoun is the main marker of the subject, since it is obligatory. The basic information it conveys can optionally be expanded by a full NP, as in (18b) and (19b).
- (ii) The NPs are the main markers of subject since they provide the fullest information. The prefix to the verb is a secondary realization, agreeing with the NP in person, number, and gender. If, however, there is no NP present, then the bound pronoun is agreeing with an implied specification which may be inferable by listeners (perhaps from the previous discourse).

Each of the answers which can be provided for (20) seems sensible. How, then, do we choose between them? The answer is that we don't. That is, we don't have to. This is because (20)—involving the analysis of surface structure—is not the sort of question which it is useful to ask.

What we have, for (18b), is a sentence whose underlying grammar consists of intransitive predicate 'eat' (marked by present tense), its subject argument being 'that small buffalo'. These underlying elements must be mapped onto surface structure, in a way that accords with the conventions of the language. In a language like English, with no bound pronouns, the subject is realized simply as an NP. In a language like Tiwi, the subject has discontinuous realization, as an NP and as a bound pronoun. Thus, for (18b):



Under this approach, it is neither necessary nor appropriate to enquire whether the main marker of intransitive subject is (i) *aw-*—optionally augmented by an NP—or (ii) *awarra kiyijini jarrangini*—obligatorily cross-referenced by a pronominal prefix to the verb. In (21), the S argument simply has double realization in surface structure. One is by a bound pronoun which—in terms of the morphological structure of the language—is an obligatory component of the verbal word. The other is, like in English, a full NP. If minimal specification of the argument is deemed sufficient, then just the pronominal prefix may be used (just as in English one may simply show the S argument by a pronoun, as in *He eats*).

The lesson here is not to attempt to analyse surface structure; this is rather like trying to assess the physical fitness of an athlete from the clothes they wear. Two languages may have similar surface structure but rather different underlying grammatical make-up, or vice versa. The crux of any language's grammatical character lies in the interweaving relations that underlie it; how they are mapped onto surface structure is a significant—but always a secondary—matter.

- (d) One of the first lessons which a linguist must learn is not to be misled into thinking that sentences which appear similar on the surface are examples of the same construction type. Compare:

(22) John began to paint the wall

- (23) John remembered to paint the wall
 (24) John went to paint the wall

These sentences differ only in the identity of the verb which comes between *John* and *to paint the wall*. In fact, this difference is absolutely critical. Both *begin* and *remember* take a Modal *to* complement clause in object function, as in (5) and (9) of §1.9. The object could, alternatively, be an NP such as *the task*: *John began the task* and *John remembered the task*.

However, when we examine the meanings of these two sentences, (22) refers to an act of painting and (23) to an act of remembering, not an act of painting (there is no guarantee, on hearing (23), that John actually did any painting). As will be discussed in §1.11, *begin* is a Secondary verb, which provides semantic modification for some other verb. Syntactically, *begin* is the main verb in (22) and *paint* the verb in the object complement clause; but semantically *paint* is the focus of attention, with *begin* providing ancillary information about it. (If (22) were to be translated into Dyrbal, *banga-* ‘paint’ would be the main verb with derivational suffix *-yarra-* ‘start to do’ added to it, giving verb stem *banga-yarra-*.) In contrast, *remember* is a Primary verb, being both main verb, at the syntactic level, and the focus of attention semantically.

Whereas each of (22) and (23) consists of a transitive clause with a subordinate clause embedded within it, as O argument of the main verb (*begin* or *remember*), and with *John* as A (transitive subject) argument, (24) involves the intransitive clause *John went*, with *John* as S (intransitive subject) argument of the intransitive verb *went*, linked with a second clause *John paint the wall* by (*in order*) *to* (the second occurrence of *John* is omitted). The full sentence is:

- (24') [John_S went] (in order) to [paint the wall]

In order to can be shortened to *to*, giving (24) which has similar arrangement of words to (22) and (23). Compare this with the structure of the other two sentences:

- (22/3') [John_A began/remembered [to paint the wall]_O]

Note also that *in order* can *not* be inserted before *to* in (22) or (23), confirming that quite distinct structures are involved.

In summary, sentences (22–4) have similar sequences of words, but represent totally different structures, with distinct meanings.

Another example of the same type—sentences which are in fact radically different in both grammar and meaning, but appear similar at the most superficial level—is provided by:

- (25a) I consider John to be clever
 (25b) I consider John clever
 (26a) I want the house to be clean (when I return)
 (26b) I want the house clean (when I return)
 (27a) John retired when he was rich
 (27b) John retired rich
 (28a) He licked the plate so that it was clean
 (28b) He licked the plate clean

In each of the (b) sentences, a transitive-verb-plus-object or an intransitive verb is immediately followed by an adjective. But the relationship between verb and adjective is quite different in each instance, as can be seen from the following discussion, based in each case on the (a) alternative.

- (25) includes a Judgement *to* complement clause in O function, as illustrated at (6) in §1.9. Some, but not all, verbs which take a Judgement *to* complement may drop the complementizer *to*, plus following copula verb *be*, as in (25b).
- (26) involves a Modal *to* complement clause, as in (5) and (9) of §1.9. A small set of verbs (quite different from those which can omit *to be* from a Judgement *to* clause) may omit *to* plus *be*, as in (26b). Whereas only copula *be* may be omitted from a Judgement *to* construction, all three varieties of *be* can be omitted from a Modal *to* clause—copula *be* as in (26b), passive *be* as in *I want the varmint (to be) executed at dawn*, and imperfective *be* as in *I want John (to be) doing some homework when his father gets home from work*.
- (27) includes an adverbial clause of time; it can omit *when he was*. Another example is *Mary entered the house (when she was) angry/in tears*.
- (28) illustrates a resultative construction, which is open to reduction. Examples of this are pretty much fixed expressions, including *knock (him) unconscious*, *shoot (him) dead*, *squash (it) flat*, *sweep (it) clean*, and *paint (it) blue*.

Some linguists have used the term ‘secondary predicate’ for some or all of the occurrences of an adjective after a verb (plus object) as in the (b) alternatives of (25–8). This is scarcely a useful course of action. It implies that there are some significant grammatical and semantic similarities between the four constructions, whereas in fact there is only the most superficial similarity of verb (plus object) being directly followed by an adjective. Alternatives (a)

demonstrate the differences, which have disparate interrelations of underlying grammatical elements.

- (e) In §1.9, we saw how a single system in grammar may code a variety of different contrasts of meaning. The four noun classes in Dyirbal deal with at least a dozen meaning contrasts and the four main complement clauses in English express diverse meaning differences across a score or so of semantic types of verbs. In similar manner, a number of entirely distinct kinds of grammatical specification may share a morphological realization.

We can again look at Latin, where each noun has a given gender and inflects for number and case, and each adjective inflects for gender, number and case. Consider the noun phrase:

- (29) *dominōs bonōs* ‘good masters’ (in O function)

This includes:

- *dominōs*, plural accusative form of ‘master’, a masculine noun
- *bonōs*, plural masculine accusative form of adjective ‘good’

Latin is a language of fusional type (see §5.5), where the *-ōs* ending is a portmanteau for plural number (out of the system singular and plural), masculine gender (from masculine, feminine, and neuter) and accusative case (rather than nominative, vocative, dative, or ablative; or genitive). But these three grammatical elements have different statuses in the grammar.

- **Gender** is an **inherent feature of the noun**. An adjective agrees in gender with a noun it modifies.
- **Number** is a **referential feature of the NP**. Plural marking (which in Latin goes on both noun and adjective) indicates that the NP refers to more than one ‘good master’.
- **Case** marks the **function of the NP in the clause**; here, accusative shows that ‘good masters’ is the O argument of a predicate. In Latin, case goes onto most words in an NP; in other languages it is marked only on the head word, or only on the last word, etc. (In Japanese, O function is shown by particle *o* following the NP.)

At an earlier stage, grammatical marking of gender, number, and case would have had separate realizations. But, as Latin evolved over time, the three elements merged into one portmanteau ending. This conflation, a matter of surface structure, should not be taken to imply any underlying association between the three categories.

The fusional character of Latin is also evident in verb inflection. For example, the ending *-et* in *monet* ‘warns’—used in examples (14–15)—encodes five distinct bits of grammar: present from the tense system, indicative from mood, passive from voice, third person from the person system, and singular from the number system, the latter two being partial realizations of the subject argument. But there is no greater association between these categories in Latin than there would be in a language in which they are all marked entirely separately, perhaps on different types of word. In this and other instances, how grammatical information is coded in surface structure is unlikely to imply anything about the underlying grammatical character of a language.

- (f) Just as diverse grammatical elements may be coded onto a single affix in surface structure (with this having little implication for the underlying representation of a sentence) so a number of elements with diverse meanings and functions may be included in one paradigmatic system.

A clause consists of a number of phrases. Each phrase has internal constituency; for example, the head may be modified by an adjective or a possessive. Languages which use affixation to indicate grammatical relations typically have two kinds of markers:

- (i) Marking relations within a phrase, such as genitive affix for possessive relation. For example, Dyrirbal uses genitive suffix *-ɲu* or *-ɲunjin* to mark alienable possession (which here covers possession of objects and of kin, since Dyrirbal is of type II in examples (1–2) of §1.3). Thus:

(30) *yara-ɲu* *midin* ‘man’s possum’
 man-GENITIVE possum

- (ii) Marking relations within a clause, typically by case affixes. Dyrirbal has absolutive case (for intransitive subject, S, and transitive object, O, functions), with zero realization, and ergative case, shown by *-du* after a stem ending in *n* (for transitive subject, A, function). A case-ending goes onto every word in an NP, including the possessive word (which already bears the genitive suffix, here in longer form, *-ɲunjin*), as in:

(31) [*yara-ɲunjin-du* *midin-du*]_A
 man-GENITIVE-ERGATIVE possum-ERGATIVE
 *gajin*_O *baja-n*
 girl:ABSOLUTIVE bite-PAST

The man’s possum bit the girl

(The four words in this sentence may be permuted into any order, with no essential difference in meaning, since their functions are fully specified by case-endings.)

Genitive, on the one hand, and cases such as absolutive and ergative (or nominative and accusative), on the other, mark different kinds of relations. It is natural and logical that they should be in different grammatical systems, and that genitive should be followable by a case suffix, such as by ergative in (31). Georgian, Quechua, many other Australian languages, and indeed many languages from other parts of the world, behave like Dyrbal in this respect.

However, some of the best-known (and most revered) languages pattern in a different way. Latin, Greek, and Sanskrit, for example, combine markers of function within a clause (nominative, accusative, dative, ablative) and the marker of possessive function within a phrase (genitive) in a single surface-structure system. That is, genitive is mutually exclusive with the cases; a noun cannot be marked for both. In a sentence such as ‘The man’s dog bit the girl’s cat’, ‘dog’ will be marked as nominative, ‘cat’ as accusative, and both ‘man’ and ‘girl’ as genitive (not as genitive-plus-nominative and genitive-plus-accusative respectively, since genitive cannot co-occur with a case). One can only distinguish between ‘The man’s dog bit the girl’s cat’ and ‘The girl’s dog bit the man’s cat’ by placing each genitive next to the noun it modifies (resorting to word order, where function-marking suffixes are not fully adequate).

The great Indian grammarian Pāṇini distinguished the two distinct types of markers within this system. He recognized a set of six *karakas*, for marking the function of a phrase in a clause, but did not include genitive as a *karaka*, thus recognizing the distinct functional properties of the *karakas* (or cases) and genitive. The earliest grammarians of Greek and Latin did *not* make such a distinction, treating genitive as a case inflection, on a par with nominative, accusative, etc. Sadly, many linguists down to the present day have followed this uninformed practice. And they then express surprise at finding that in a language like Georgian or Dyrbal genitive can be followed by a case such as dative or ergative (calling this ‘double case’). (In some languages, a certain affix can have two functions, marking both a clausal and a phrasal function; one typical dual-function is dative/genitive. The fact that grammatical elements with two functions exist—just like lexemes with two quite distinct senses—in no way obscures the fact that phrasal and clausal relations are different matters, and must be clearly distinguished.)

This demonstrates how analysis which is confined simply to consideration of surface structure may obscure the underlying grammatical relations for a sentence. Similar conflation is encountered with verbal affixation. As with nouns, a language may economize on how many distinct verbal systems it accommodates within surface structure, having a single system whose terms relate to two or more distinct grammatical categories. Tense indicators may

be included in the same system as the marker of imperative mood; but this should not be taken to imply that imperative is a type of tense. And sometimes evidentiality suffixes appear in the same system as mood markers; this does not mean that evidentiality is a kind of mood (although some linguists have arrived at such an erroneous inference).

In summary, one is unlikely to achieve an understanding of the basic grammatical organization of a language through analysis of surface structure.

We saw, first, that it is scarcely useful to posit a level of underlying structure which is similar in nature to surface structure, with words organized in a fixed order, the underlying structure perhaps differing from surface structure just in being a little neater, without any discontinuous constituents. Words and phrases must occur in an order in surface structure (in order for a sentence to be spoken); the ordering is fairly fixed in some languages but rather fluid in others. For all languages, the underlying grammar involves a network of relationships. Ordering of elements is not a factor at this level; it comes in only through the manner in which underlying elements are mapped onto surface structure.

An underlying element may be realized at two places in surface structure; for instance, a predicate argument can be shown by an NP and by a pronominal affix to the verb, as in (17–19). It is neither relevant nor useful to reflect on which of these should be considered the main realization of the argument.

A number of quite different constructions may be (often, after shortening by omission of some basic elements) superficially similar in surface structure. This carries no implications that there is anything in common to their underlying grammatical representations, as was illustrated by clauses with *to*, in (22–4) and by those with an adjective directly following a verb (plus object, if transitive), in (25–8).

Many languages organize their surface morphology through a number of disparate categories (such as gender, number, and case) being accorded portmanteau realization, or with a number of rather different markers being combined into a single surface system (as case and genitive are in languages such as Latin, Greek, and Sanskrit). These surface-structure associations have little consequence for underlying grammatical organization.

Grammar includes various kinds of process, which apply to lexemes. A lexical root (such as the verb *organize* in English) can undergo derivation, producing the noun *organization*. This then undergoes the inflectional process appropriate to nouns, number marking; we can get the plural word *organizations*. Suitably processed lexemes are related together in phrases, then forming clauses, sentences, and chunks of discourse. The next section considers and contrasts lexicon and grammar.

1.11 Grammar and lexicon

A language is made up of two independent but interlocking parts—grammar and lexicon. The grammar is a little like a city centre—well-traversed thoroughfares, feeding into each other, replete with signs and signals and short cuts. The lexicon is somewhat akin to a parking lot—full of vehicles which will leave as needed, to engage in traffic within the city.

The wherewithal of grammar consists in small systems, such as gender, case, tense, and types of complement clause. Each system is closed; that is, new members may not (save in exceptional circumstances) be added. The terms in a system may be exhaustively listed, each being fully defined by the exclusion of all others. In the three-term number system of Kayardild (see Table 1.1 in §1.4), ‘dual’ can be specified as ‘neither singular nor plural’. English has seven personal pronouns. Suppose that I am thinking of a pronoun in English. It is not (quoting subject forms) *we* or *he*, *she* or *they*, *it* or *I*. What is it? It must be the second person pronoun, *you*, which can be defined as being complementary to the other six.

One grammatical system may depend on others. Gender is found only for 3rd person singular in English pronouns. For Tucano (see §1.5), there is a system of five evidentiality choices in past tense, just three (omitting assumed and reported) in the present, and no evidentiality specification at all in future tense. In Amele (Gum family, Papuan region), there are three past tenses and two futures within positive polarity, but just one of each for a negative clause. (See §3.19.)

Every grammatical system has limited size so that its terms can be—and should be—exhaustively listed within a statement of the grammar: all the pronouns, prepositions, articles, interrogatives, noun classes, every type of complement clause, each of the possessive constructions, ways of forming a causative, and so on.

Whereas a grammar involves closed systems, a lexicon consists of open classes—typically noun, verb, and adjective. We found it possible to specify the pronoun *you* by saying that it was not any of the other terms from the English pronominal system (not *I*, *he*, *she*, *it*, *we*, or *they*). This would not be possible for a lexeme. (I’m thinking of a noun and it is not *aardvark*, *abacus*, *acacia*, *adenoids*, ... What is it? Can’t be done.) Lexical classes typically have large membership on which no upper limit may be placed. While I’m attempting to list all the nouns I know, new ones will be coming into being—created from within the language or borrowed from other tongues. The task would have no end.

A grammatical form will be fully specified within the grammar (for example, *she* is 3rd person singular feminine subject pronoun), a lexeme partially

so. *Dog*, in English, is a countable noun (it can take plural *-s*). But exactly the same grammatical profile is provided for *cat* and *horse* and *crocodile*. It is the role of lexical entries to distinguish between these various count nouns. And similarly for adjectives such as *red*, *blue*, and *yellow*, for verbs such as *ask*, *request*, and *demand*, for adverbs such as *mainly*, *mostly*, and *chiefly*. (Some remarks on the ideal nature of a lexicon are in Chapter 8 below.)

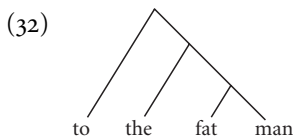
The description of a language has two parts. The grammar deals—in as much detail as is considered necessary—with the underlying categories and structure (with a chapter on their phonological realizations). The lexicon, or dictionary, lists as many as possible of the lexical forms, which can slot into grammatical constructions. Grammar and lexicon are essentially separate components, with considerable cross-referencing between them.

A dictionary exists to deal with lexemes, those forms which are not uniquely specified by the grammar and require definitions to tell them apart. It is neither necessary nor appropriate to include in a dictionary grammatical forms, which are uniquely defined within the grammar. Some of the earliest dictionaries followed this practice, listing just lexemes and excluding fully grammatical items such as *what* or *that* or *the* or *to*. Then it became the custom to list in a dictionary every word, even those which are fully specified within a grammar of the language. Nowadays some dictionaries even include affixes, like *un-* (as in *untie*) and *-th* (as in *truth*). They list grammatical forms, but without providing information concerning the composition of the grammatical system to which the item belongs, although it is the contrast with other terms in its system which characterizes the grammatical form's function and meaning.

Sadly, dictionary and grammar—at least for the major languages—tend to be compiled by separate groups of scholars, with different aims and methods. Ideally, dictionary and grammar should be produced in concert, with ample cross-referencing. If it is considered useful to include grammatical forms in the overall alphabetical list, all that is required is a reference to those sections of the grammar in which they are fully discussed.

Just as grammatical forms do not require 'definitions' in a dictionary, so a clear division should be made between lexemes and grammatical elements in, say, parsing a sentence. An unfortunate trend in modern studies of a language like English is to treat each orthographic word equally, taking no account of whether it is a lexeme or a fully grammatical unit.

For example, the phrase *to the fat man* (from the sentence *He gave an apple to the fat man*) is typically assigned a 'tree structure' something like:



What we have here is two lexemes, *fat* and *man*, marked by the definite article, *the*, and by the preposition *to*. Although written as separate words, the two fully grammatical items are pronounced as proclitics (‘=’ indicates a clitic boundary):

(33) /tə=ðə=fæt mæn/

The is not a constituent of the phrase; it is a grammatical form stating that *fat man* has definite reference. In similar fashion, *to* is a marker of the function of *the fat man* in the clause, indicating that it refers to the recipient of an act of giving.

In some languages, definiteness is shown by an affix, rather than by a clitic (written as a separate word) as in English. It would then be clear that the definite marker is not a lexical-type constituent of the phrase, in the way that *fat* and *man* are, but the realization of a grammatical category.

To the fat man is a noun phrase, marked by preposition *to* (which in this context indicates benefactive function). Some linguists call *to the fat man* a ‘prepositional phrase’, with a binary split into constituents *to* and *the fat man* (and some go further, and say that *to* is the ‘head’ of this ‘prepositional phrase’). But in Latin, for instance, ‘to the fat man’ would be *vir-ō obēs-ō*, where the *-ō* ending on both *vir-* ‘man’ and *obēs-* ‘fat’ marks masculine singular dative (Latin has no grammatical category of articles). One surely wouldn’t call *vir-ō obēs-ō* a ‘case phrase’ (although this would be the logical extension of calling *to the fat man* a ‘prepositional phrase’). And one surely wouldn’t pick out the repeated ending *-ō* as an immediate constituent of *vir-ō obēs-ō* (and as the head of this phrase!); neither should *to* in English be treated as a lexical-type constituent. The English phrase is most appropriately represented by something like:

(34) BENEFACTIVE RELATOR_(to) [*fat man*]_{DEFINITE(the)}

There is further discussion of prepositions in §5.4 and §5.6.

Similar comments apply to clause linkers such as *and*. Some linguists represent the phrase *cats and dogs* as having three constituents (treating *and* on a par with *cats* and *dogs*). Others prefer binary splits and require two constituents; there is then a problem—is it [*cats*] [*and dogs*] or [*cats and*] [*dogs*]? What we

have here, in fact, is two lexemes, linked by *and* (which is not a constituent, or part of a constituent, but rather a grammatical marker).

Some concepts are always dealt with through the lexicon rather than in the grammar, in every language—the contrasts between *cat* and *dog*, between *laugh* and *cry*, between *white* and *red*, and so on. Other types of information are always the province of grammar—marking a sentence as interrogative or imperative, showing what is subject and what is object, and other things of this nature.

But there are concepts which may be coded within a grammar in one language but are shown only by lexemes in another. The Australian language Yidiñ—like many languages from Africa, and elsewhere—has verbal suffixes *-ɲali-* ‘go and do’ and *-ɲada-* ‘come and do’. They can be illustrated with verb *wuna-* ‘sleep’ and imperative inflection *-n* in:

- (35) *wuna-n* ‘sleep!’
 wuna-ɲali-n ‘go and sleep!’
 wuna-ɲada-n ‘come and sleep!’

That is, Yidiñ does by choice from a grammatical system what English requires lexemes, *go* and *come*, to achieve. (The ‘go and do’—or ‘do while going’—suffix is further illustrated in §4.9.)

Cross-linguistically, a number of what can be called ‘secondary concepts’ may be recognized. These are items which are coded within the grammar in some languages but dealt with through lexemes in others. They include ‘try’, ‘start’, ‘continue’, ‘cease’, ‘finish’. English, a language with a rather small set of suffixes, has all of these as verbs. It was shown in (d) of §1.10 that in (22) *John began to paint the wall*, *begin* is the syntactic main verb, taking a complement clause in O function; but semantically the complement clause verb *paint* is the focus of attention, with *begin* providing ancillary information about the activity. As mentioned in §1.10, the secondary concept ‘begin’ is expressed through a verbal suffix *-yarra-* in Dyirbal; added to verb *baɲga-* ‘paint’, we get *baɲga-yarra-* ‘begin to paint’.

Other secondary concepts include ‘want’ and ‘make’, which are again dealt with through lexemes in English but by grammatical affixes in many languages. For example, Luiseño, a Uto-Aztecan language, has a desiderative suffix *-viču-* which can be added to lexical root *ɲée-* ‘leave’, giving *ɲée-viču-* ‘want to leave’. Causative suffix *-ni-* may also be used with this verb, yielding *ɲée-ni-* ‘make (someone) leave’. A verb may accept both suffixes—*ɲée-viču-ni* ‘make (someone) want to leave’. And, indeed, there can be two instances of the desiderative flanking the causative suffix, in *ɲée-viču-ni-viču-* ‘want to make (someone) want to leave’.

As a language develops over time, new elements will enter the grammar, often developments from lexemes. Yidiñ has lexemes *gali-* ‘go’ and *gada-* ‘come’. They would have been used with another verb, as in *wuna-n gali-n* (‘sleep-IMPERATIVE go-IMPERATIVE’) ‘go and sleep’. The two words merged, giving *wunaŋgali-n*, which reduced to the present-day *wuna-ŋgali-n*, where *-ŋgali-* is now one of the two terms in a grammatical system of derivational suffixes to a verb. Such grammaticalization of lexemes is a pervasive tendency as a language evolves over time. A lexical item, from an open class, may develop into a grammatical element; it is likely gradually to lose the old lexical meaning, and may take on a wholly relational role.

The noun *side* in English has a long history, originally meaning ‘the long part of a thing’. In Middle English there developed *beside* (a single word) as a preposition within the grammar, later becoming *besides*. The original meaning ‘by the side of’ took on a more abstract sense ‘in addition to’, as in *Besides a gun, a soldier should also carry chocolate*. Then *besides* also took on the role of clause linker (similar to *moreover* and *however*), as in *I haven’t the time to see that film and, besides, I don’t like sloppy love stories*.

The fact of lexemes being grammaticalized is indisputable. Some tense affixes developed from time lexemes, and case markers from things like body-part terms. However, linguists argue concerning the margin of what should be included under this label.

In order to produce an acceptable sentence, one must make a choice from a number of obligatory grammatical systems, depending on the language. The organization of a grammar determines what one *must* say. As mentioned in §1.5, in Tucano a speaker is obliged to state the evidence on which a statement is based by choosing one of the five terms from the evidentiality system (seen, heard, inferred, assumed, reported). In the Western Torres Strait language, a statement about the future must specify whether the time reference is to ‘later today’, ‘tomorrow’, or ‘beyond tomorrow’. In other languages, such types of information may be provided through lexemes, but as an optional matter.

To some extent, the types of categories in a grammar both reflect and motivate the way in which its speakers view the world about them. But they do not limit this. Estonian has no grammatical category of gender, yet speakers of Estonian are fully aware of differences between the sexes, and can—if they wish—add a noun ‘man’ or ‘woman’ to a basic sentence such as *tema sööb* ‘he/she eats’ in order to specify the physical gender of the person referred to.

We can now briefly examine the recurrent classes of lexemes.

- All languages have a large open class of **nouns**, with at least several thousand members, to which new items can be—and steadily are being—

added. There are always some ‘concrete’ nouns referring to objects, such as ‘girl’, ‘tiger’, ‘hill’, ‘stone’, ‘water’, and ‘head’. Many languages also have abstract nouns, some basic (such as *beauty* and *hunger* in English), others derived from adjectives and verbs (*ugliness*, *thought*); in other languages such concepts are rendered only through adjectives and verbs.

- There is always a class of **verbs**, generally also an open set with at least several hundred members. However, there are languages with a smaller verb class of only about a hundred items, sometimes even fewer. Such languages typically have many complex expressions, each including a verb and another stem, which may be a noun or adjective (or a preposition, similar to phrasal verbs in English). Yawuru, spoken in north-west Australia, has only about eighty simple verb roots, which take affixes to mark tense, aspect, and mood, plus person and number of subject and object. About a dozen of the verbs may be used with a range of non-inflecting ‘coverbs’. For example, *-ga-* ‘carry’ occurs in many complex verbal expressions, including:

- (36) η anjbi -ga- ‘carry, holding under the arm or by the side of
 the body’
 η anjdja -ga- ‘carry in the mouth (as a dog does)’
 wirrp -ga- ‘smash, hit hard’
 mardalj -ga- ‘make noise, be noisy’

It will be seen that the first two complex expressions expand on the basic meaning of *-ga-*, while the last two have an entirely different sense (much like phrasal verbs in English). In languages of this type, there is generally a very large number of complex verbal expressions—effectively, an open class.

- It is likely that a class of **adjectives** can be recognized for every language, although there are two main parameters of variation.

The first relates to grammatical profile. In some languages, adjectives have similar properties to nouns; Latin is an example, where adjectives inflect for case and number—like nouns—and also for gender—in agreement with the noun they modify. (For such languages, some linguists treat adjectives as a subclass of nouns.) In other languages, adjectives share grammatical properties with verbs; in Chinese, for example, an adjective may occur in the same functional slot as an intransitive verb. (Some linguists treat adjectives as a subclass of verbs, in such languages.) Then there are languages—like English—where adjectives have rather different grammatical properties from both nouns and verbs. And others—including the Berber languages of North Africa—whose adjectives share properties with both nouns and verbs.

The second parameter of variation is size. English and many other languages have an open class of adjectives, with hundreds of members (to which new items may be added). Other languages have a small, closed class, with from half a dozen to a hundred or so members.

It is far from the case that all the members of each lexical class have the same grammatical properties. We can usefully recognize a number of ‘semantic types’ within each class; the members of a given type will have similar meanings and shared grammatical properties.

For adjectives the major semantic types are DIMENSION, AGE, COLOUR, VALUE, PHYSICAL PROPERTY, and HUMAN PROPENSITY. Small closed adjective classes tend to draw their members from the first four types; for example, there are just eight adjectives in Igbo, from Nigeria, two in each of the critical types—‘large’ and ‘small’ (DIMENSION), ‘new’ and ‘old’ (AGE), ‘black, dark’ and ‘light, white’ (COLOUR), and ‘good’ and ‘bad’ (VALUE). Slightly larger classes may include some PHYSICAL PROPERTY terms (such as ‘unripe’ and ‘heavy’). Only in adjective classes with at least a few score members would we expect to find HUMAN PROPENSITY terms (such as ‘clever’ and ‘jealous’).

Within a given language, the types are likely to have different grammatical properties. For example, in English the prefix *un-* may be used with many adjectives from the HUMAN PROPENSITY type, with some from VALUE and with a few from PHYSICAL PROPERTY, but with none from DIMENSION, COLOUR, or AGE. Just adjectives from the HUMAN PROPENSITY type may be followed by a preposition-plus-NP explaining the nature of the quality, as in *clever at mathematics/at solving puzzles* and *jealous of his rival/of Mary’s winning the prize*.

For the noun class there are also significant semantic types. Only a HUMAN noun can be subject of a SOCIAL CONTRACT verb, such as *appoint*, *convert*, and *arrest*. The object of *experience* is generally a STATE NOUN, such as *hunger* or *a whipping*.

The class of verbs, in any language, covers a wide range of meanings, so that it is here really useful to recognize distinct semantic types, each with its own set of semantic roles. For verbs of the GIVING type, there are three roles: Donor, Gift, and Recipient, as in *John gave a book to Mary*. For the SPEAKING type, four roles must be recognized: Speaker, Addressee, Message, and Medium, as in *John told Mary a joke in French*. For LIKING verbs the roles are Experiencer and Stimulus, as in *John likes jazz*. For ATTENTION verbs (including *see*, *hear*, *find*, *witness*) the roles are Perceiver and Impression, as in *Mary witnessed the fight*.

Every semantic type, in a given language, employs a convention whereby each semantic role is associated with a certain syntactic function. In English

all of Donor, Speaker, Experiencer, and Perceiver correspond to transitive subject (A) function. There is little in common between someone who transfers possession of something, someone who utters something, someone who experiences a certain feeling, and someone who (possibly unwillingly) receives a sense impression. What is common to these roles is that all are associated with the same syntactic relation. That argument whose referent could initiate or control the activity (if anything could) is recognized as fulfilling function A.

A further useful distinction is between Primary verbs, which directly refer to an activity or state, and Secondary verbs (the lexical realization of Secondary concepts), which effect semantic modification of some other verb. And there is a further division. Primary-A verbs have all of their arguments realized as NPs; these semantic types include MOTION (such as *run*, *throw*), REST (*sit*, *hold*), AFFECT (*hit*, *twist*), GIVING (*give*, *present*), CORPOREAL (*eat*, *laugh*), and half a dozen more. Primary-B verbs may have either an NP or a complement clause as one argument, e.g. *I chose Mary* or *I chose to marry Mary*.

In §1.9, we examined the complement clause possibilities for two Primary-B types, THINKING and LIKING. All THINKING verbs can take a THAT complement and all LIKING verbs accept an ING complement clause, with other verbs from the two types showing other possibilities, relating to their specific meanings. Other Primary-B types include ATTENTION, DECIDING (e.g. *choose*, *elect*), COMPARING (*resemble*, *compare*), and RELATING (*depend on*, *imply*).

As mentioned earlier, Secondary concepts may be realized as verbal affixes in languages with an extensive morphology, but tend to be expressed as lexemes in languages with little morphology, such as English. A Secondary verb functions as a main verb, but—as illustrated above—it effectively provides semantic modification for the verb of its complement clause. Secondary verbs in English include the semantic types of BEGINNING, TRYING, WANTING, MAKING, and HELPING.

Chapters 11 and 12 discuss in more detail the three major word classes, and criteria for distinguishing between them.

Sources and notes

1.3. Sources for the types of possession mentioned here are: Amele in Roberts (1987); Maricopa in Gordon (1986b); Nootka, Haida, in Sapir (1917); Ewe in Ameke (1996); Tachelhit in Aspinion (1953); Lango in Noonan (1992); Gapa-paiwa in McGuckin (2002).

Vowel development from proto-Austronesian to proto-Oceanic in Dahl (1976: 14–19). Discussion of developments in Olgolo in Dixon (1982: 207–10; 2002: 468).

1.4. 2nd person pronouns in Akan from Ikoro (1997: 27); in Kayardild from Evans (1995: 202); in Longgu from Hill (1992).

1.5. Genders in Lak from Khaidakov (1963); in Supyire from Carlson (1994). Tense in the Western Torres Strait language from Bani and Klokeid (1971). Evidentials in Tucano from Aikhenvald (2004). Imperatives in Tuyuca from Barnes (1979, 1984).

1.6. Demonstratives in Lak from Khaidakov (1966). Harmonic and disharmonic dual pronouns in Lardil from Hale (1966). Pronouns in Bengali from Onishi (1997). Classifiers in Jacaltepec from Craig (1986).

Information on Pennsylvania German from Burridge (2002). On Malagasy from Keenan and Ochs (1979). On Korean from Sohn (1994: 99).

Kemal Atatürk on Turkish language reform—see Bazin (1983). Hungarian eliminating Latin loans—see Fodor (1983).

1.7. Useful cross-linguistic discussion of body part terms is in Majid, Enfield, and van Staden (2006).

Tense in West Torres Strait language from Bani and Klokeid (1971); in Yimas from Foley (1991). Relative clauses in Kambera from Klamer (1998); in Persian from Comrie (1989).

1.8. Halpern (1942) describes kin terms as verbs in Yuman languages.

1.9. Noun classes in Dyirbal—Dixon (1982: 178–83). Note that Lakoff (1987) repeats information from Dixon (1972, 1982) but does not add to it. Complement clauses in English—Dixon (2005a).

1.10. Information on Tiwi from Lee (1987: 173, 175, 101, 223, 230). In Tiwi, bound pronouns are used for both subject and object (not illustrated here). But in some languages they only represent subject. This produces difficulties for solution (i)—do we have the bound pronoun as the main realization of subject but an NP as main realization of object? This would surely be a rather heterogeneous technique of analysis.

In Latin, as mentioned under (e) below, one verbal suffix encodes tense, mood, voice, plus person and number of subject. One would surely not want to say that this portmanteau suffix constitutes the major realization of the subject argument.

And there are languages with bound pronouns that are optional. In some one can have a bound pronoun or a free pronoun (in an NP) but not both, posing difficult decisions for analysis (i). All this is avoided by working in terms of underlying arguments, and then specifying the range of realizations of each.

Concerning genitive and cases, Plank (1995) repeats the time-worn misconception that genitive is a case. His paper (and following chapters in the

same book) do provide more examples of languages in which genitive can be followed by a case (as well as a couple of examples of genuine 'double case').

For the mistaken idea that evidentiality is a type of mood see Palmer (1986), and the rebuttal in Aikhenvald (2004).

1.11. Evidentiality and tense in Tucano—Barnes (1984). Tense and polarity in Amele—Roberts (1987). Information on Luiseño from Langacker (1972: 76–7). Verbs in Yawuru—Hosokawa (1991) and Dixon (2002: 185).

2

Principles to Follow

2.1 Writing a grammar

Writing up a grammar for publication is a quite different matter from analysing the data and gradually building up a full understanding of the language. The way in which the grammar is organized should not (except coincidentally) reflect the way in which the linguist worked and the order in which analytic decisions were reached.

When commencing work, the linguist will record, transcribe, and analyse texts, uncovering bit by bit the grammatical regularities and irregularities of the language. Gradually, over a period of months, the overall structural scheme of the language will emerge, as the linguist is able to relate together bits of patterns from different areas.

As illustration, consider a language with a standard five-vowel system and CV syllable structure. There are only a few suffixes, but some have alternative forms. These include:

- The past tense suffix on verbs is *-he* but, after a verb root ending in *i*, this *i* assimilates to *e*; for example, *madi* ‘climb’, *made-he* ‘climb-PAST’.
- Nouns take accusative suffix *-ho*. When this is added to a root ending in *u* or *i*, these root-final vowels become *o* and *e* respectively. For example, *bamu* ‘dog’, *bamo-ho* ‘dog-ACCUSATIVE’; *gadi* ‘cat’, *gade-ho* ‘cat-ACCUSATIVE’.

First of all, the linguist notes these instances of vowel lowering while studying the forms of verbs and of nouns, and then combines the two sets of alternations:

VERB	-i	+	-he	→	-e-he
NOUN	-u	+	-ho	→	-o-ho
	-i	+	-ho	→	-e-ho

The next step is to seek out an inductive generalization:

high vowel at end of verb or noun root (<i>i</i> or <i>u</i>)	is replaced by	corresponding mid vowel (<i>e</i> or <i>o</i>)	before	suffix commencing with <i>h</i> followed by a mid vowel (<i>e</i> or <i>o</i>)
---	----------------------	--	--------	--

Now if this were a general rule, across the language, we would expect to get a verb which ends in *u* and which would raise this to *o* before past tense suffix *-he*. At this stage of study, the linguist has many verb roots ending in *a*, *e*, *o*, and *i* but none with final *u*. Then such a verb does turn up (in a new text or in conversation) and it does behave like this. The prediction is confirmed and the inductive generalization established. Rather than stating morphemic alternations separately in the noun and verb chapters, the linguist can now include this general rule of assimilation in the chapter on phonology.

As work on the language proceeds, results from different areas will be related to each other; for example, the morphological structure of nouns, and the syntactic principles of noun phrase structure and clause structure. Suppose that each noun may take one of four suffixes, which may be roughly labelled:

-ho accusative -nu genitive
-bi comitative -ga locative

A noun in subject function takes no suffix at all; this can be called nominative case, with zero realization.

However, these suffixes have disparate roles within the grammar of this particular language. Two indicate function within the clause—accusative for the object argument of a predicate, and locative for a peripheral argument ('at', 'on', or 'in'). The other two indicate function of an element within a noun phrase, as in:

- | | | | | | |
|-----|--------------|------|-----|------|----------------|
| (1) | dama-nu | bamu | (2) | dama | bamu-bi |
| | man-GENITIVE | dog | | man | dog-COMITATIVE |
| | man's dog | | | man | man with a dog |

As exemplified under (f) in §1.10, for some languages—typically, those with a more extensive morphology than the one under consideration here—there may be distinct morphological systems for marking phrasal and clausal functions, so that one kind of suffix could be followed by the other kind. For example, 'man-ACCUSATIVE dog-COMITATIVE-ACCUSATIVE' for '(I_A saw) [the man with a dog]_O'. But in the language under study, *one* system of nominal suffixes combines *two* syntactic roles—and must be perceived to do so. The first stage is recognition of the set of suffixes which can follow a noun; the next is to work out that there are two systems of suffixes, with distinct syntactic functions.

Once the linguist has achieved a reasonable understanding of all areas of grammar and phonology (perhaps at the end of the first long period of fieldwork), it is time to step back in order to view the structure of the language as a whole, within the perspective of general typological theory. And then to devise a plan of how best to present this structure in the course of writing up the grammar.

There is similarity to the methods of an investigator of some criminal or nefarious political activity. They will gradually uncover bits of the puzzle. Only when all the bits have been assembled, to form a whole—and they have achieved a complete understanding of who did what, and why, and the causes and consequences—will they start to plan how to write it up, as a book or article. The order in which things are described in this account will bear no relation to the order in which the investigator came to uncover them. So it is when writing a grammar.

A good grammar will flow. It should be written in clear style, avoiding pedantic or obscure prose. The linguist should be able to take pleasure in composing it; if they do, the reader is likely to study it with pleasure as well as with profit. Sometimes people say: ‘I consulted one of your grammars for a particular point and just kept on reading. I couldn’t put it down.’ This, to me, is a high compliment.

Each grammar requires different organization. If the facts concerning X are needed as a criterion for Y, then the chapter on X must precede that on Y. For instance, in Fijian the object of a transitive verb can be an NP, as in ‘He is eating breadfruit’. But an object noun can be incorporated into the verb, making the whole construction intransitive: ‘He is breadfruit-eating’. The justification for this analysis lies in the position and meaning of modifiers which follow the verb, but precede an object NP. Consider first a regular transitive sentence:

- (3) erau_A 'ani-a ruarua
 2du.SUBJECT eat-TRANSITIVE+3sg.OBJECT both
 [a uto]_O
 ART breadfruit

The article *a* marks the head of its NP as a common noun, rather than a proper noun or a pronoun. Suffix *-a* to the verb marks it as transitive with a 3sg object. The modifier *ruarua* ‘both’—which must immediately follow the verb—could relate to either subject or object argument; that is (3) could mean either ‘They are both eating breadfruit’ or ‘They are eating both breadfruit’.

Now consider (4), in which the noun *uto* ‘breadfruit’ is incorporated into the verb (which lacks a transitive suffix); the composite form *'ana-uto* is intransitive.

- (4) erau_S 'ana-uto ruarua
 2du.SUBJECT eat-breadfruit both
 They are both breadfruit-eating

The fact that *ruarua* here follows *uto* shows that this noun is now part of the verb. The modifier *ruarua* can relate to either subject or object; in (4) it relates only to the subject—so that this sentence is unambiguous—showing that there is no object and the sentence is intransitive.

In writing a grammar of Fijian, it is thus sensible to describe the positioning and meaning of verb modifiers such as *ruarua* before the chapter on noun incorporation. There is nothing worse (and more user-unfriendly) than coming across a criterion in chapter five which refers to something not yet introduced—so that one has to flip ahead to chapter eight in order to understand the nature of the criterion, and then back to chapter five.

Working out how to organize a grammar is often not an easy matter. One realizes that a full discussion of topic A presupposes knowledge of B, a discussion of B requires prior knowledge of C, and C requires A—a full circle. Where to start? In a couple of cases I have solved this by commencing the grammar part of a description with one or two short chapters which provide an overview of the whole (chapters 4–5 of Dixon 1988, on Boumaa Fijian, and chapter 3 of Dixon 2004a, on Jarawara). These survey the basic facts concerning topics A, B, C, and so on. Later chapters then expand on each aspect of the overview; there is enough about B in the overview for the needs of the chapter on A, and so on.

Many grammar-writers attempt to follow these basic principles; they want linguists to be able to read their grammars in the way that they themselves read the grammars of other scholars. But a few flout such a convention. For example, Newman's otherwise excellent grammar of Hausa (2000) has the eighty chapters all arranged in alphabetical order! Thus, for example, '28, Focus' is between '27, Expressions of contempt' and '29, Frequentatives' while '72, Topicalization' is flanked by '71, Tone and intonation' and '73, Universals and generic relatives'. The chapter on phonology (54) is nowhere near that on tone and intonation, being instead nestled between '53, Numerals and other quantifiers' and '55, Pluractional verbs'. I tentatively infer that Newman himself is not in the habit of sitting down and perusing a grammar from first page to last, and thus doesn't consider catering for those who do. (There were two series of grammars, the *Handbook of Amazonian Languages* and the *Lingua* (later *Croom Helm*, later still *Routledge*) *Descriptive Series*, which imposed a fixed format. They are not at all easy to read and—some of the authors have commented—were far from easy to write.)

The ‘traditional’ way of presenting a description of a language is: phonology, then morphology, then syntax, then discourse structure (with plentiful cross-references back and forth). For some languages it is possible to vary the order. In the 1972 grammar of Dyirbal I put the main syntax chapter before the bulk of the morphology and the phonology. But, ahead of the syntax, I did include a chapter called ‘Word classes’ which outlined the inflectional morphology; this was, of course, needed to understand the syntax. Different languages require different strategies of description but, whichever technique is chosen, a main aim should be to ensure that it is internally consistent and easy to both read and refer to.

The example sentences provided in a grammar must in every instance be clearly analysed and glossed, so as to be immediately understandable. Four basic principles help achieve this:

- (a) Ensure that all multi-word constituents are within square brackets.
- (b) Label all verbal arguments. For core constituents, a good convention is a subscript indicating syntactic function, A (transitive subject), S (intransitive subject), or O (transitive object).
- (c) Where possible, show morpheme boundaries by hyphens.
- (d) Provide a gloss for each morpheme.

These principles are exemplified in (3–4) above, and in many other examples throughout this chapter (and, indeed, book).

To take an example from English, the sequence of words *the African helps remember* is ambiguous between two syntactic structures and meanings:

- (5) [The African helps]_S remember
- (6) [The African]_A helps [remember]_O

In (5) *African* is a modifier to noun *helps* (with similar meaning to *assistants*), which is plural head of the intransitive subject NP to verb *remember*. In (6) *African* is singular head of the transitive subject NP to verb *helps*, whose object is a complement clause (*to*) *remember*, with the *to* here omitted. *Remember* could be provided with its own object; for example, *what happened*.

In other languages, especially those in which subject and object typically occur on the same side of the verb—and also when there is greater freedom of phrase order and of word order than in English—bracketing and labelling are often critical for a reader’s understanding.

In a grammar, the linguist should thoroughly describe the organization of the language—generalizing as much as is appropriate, and relating their account to the typological parameters of basic linguistic theory (built up by inductive generalization on the basis of many previous reliable grammars). There should

also be as much explanation as possible. Why—when a certain affix is added to a certain class of words—does the vowel become shortened? Because the vowel would then be in an unstressed syllable and long vowels must always bear stress. (There should be reference back to the chapter on phonology, where this will have been discussed.)

Explanation may also have an historical basis. This can be exemplified from English. As I learned the language (growing up in England in the years immediately following the Second World War), one could say *larger* and *smaller*, *largest* and *smallest*, *bigger* and *biggest* but not **littler* or **littlest* (*smaller* and *smallest* had to be used instead). Why was this? Well, in Old English, the adjectives *micel* ‘big’ and *lytel* ‘little’ had irregular comparative and superlative forms:

	PLAIN	COMPARATIVE	SUPERLATIVE
‘big’	micel	māra	mǣst
‘little’	lytel	lǣssa	lǣst

The form *micel* dropped out of use (being replaced by *big*), but its comparative, *more*, and superlative, *most*, were retained as general periphrastic markers for adjectives which do not take *-er* or *-est* (and for some that do). The comparative, *less*, and superlative, *least*, of *little* took the same path, becoming dissociated from the adjective *little*. As a consequence the lexeme *little* had no comparative or superlative forms. Only within the last few decades—and particularly in dialects outside Britain—have the regular comparatives *littler* and *littlest* come into use, filling in this gap after a period of almost a thousand years.

One should not invoke historical changes that have taken place as criteria for analysis of the present-day system of a language. But history may provide explanation for an irregularity or a gap, or just some particular pattern of grammatical marking. In Jarawara, for example, many body-part terms have distinct f(eminine) and m(asculine) forms. These include:

f	m	
nok̩	nok̩	‘eye’
tame	teme	‘foot’

It can be seen that the difference in gender is shown by the final vowel for some body-part terms, such as ‘eye’, and by the first vowel for others, including ‘foot’. Why should this be?

By comparing Jarawara with related languages from the Arawá family, I have been able to reconstruct the forms of body-part nouns in the shared ancestor language, proto-Arawá, and the changes between it and modern-day Jarawara. In proto-Arawá, each body-part term had a fixed form (**noko* for

‘eye’, **tama* for ‘foot’, marking reconstructed proto-Arawá forms with *) with f and m being shown by suffixes *-*ni* and *-*ne* respectively. There were then the following changes:

f	m
* <i>noko-ni</i> > <i>noki</i>	* <i>noko-ne</i> > <i>noko</i>
* <i>tama-ni</i> > <i>tame</i>	* <i>tama-ne</i> > <i>teme-ne</i> > <i>teme</i>

First of all, m suffix *-*ne* triggered a change by which preceding *a*'s in the word became *e*, **tama-ne* > *teme-ne*. The final -*ne* was then dropped from the m form of some words, *teme-ne* > *teme*. (However, -*ne* was retained on others; for ‘blood’ we have f **ama-ni* → *ame*, and m **ama-ne* > *eme-ne*.)

After this, the f suffix *-*ni* was lost, -*o-ni* becoming *i* (**noko-ni* > *noki*) and -*a-ni* becoming *e* (**tama-ni* > *tame*). Note that this change followed that which gave **tama-ne* > *teme-ne* for the m (otherwise we would have got **tame* > *teme* for the f form, which didn't happen).

All this explains the odd-seeming alternatives in the modern forms. I must admit that when working with the language I may remember the two forms for each body-part term, but can't always immediately recall which is f and which m. What I do is think of the proto-Arawá forms (which were totally regular) and apply the historical changes in order to get present-day forms. (This only takes a couple of seconds.)

If the historical route by which some irregularity in the modern language developed is known—or can be reconstructed—then this should be included in the grammar, as an explanation of how the modern language is as it is, in this respect.

In summary, the plan of a grammar is only decided on once every part of the structure of the language is more or less understood. The way the grammar is written will not relate to the order in which bits of the puzzle were worked out. The aim is to produce a logically coherent account, where the prerequisites for each statement of analysis have been included at an earlier stage in the exposition. Each language has its own particular character and it is this that determines the optimal order of presentation.

Clarity should be a major consideration, both in general exposition and in the presentation and labelling of examples. Hand in hand with description there should be explanation; see §1.6 and §4.9. An apparent irregularity may be explainable in terms of a constraint in some other part of the grammar. Also, the way in which the language developed (insofar as this is known, or can be reconstructed)—while not valid as a criterion for preferring one analysis over another—can provide an explanation for why things are the way they are.

2.2 Always go back to the primary sources

It is a joy to peruse a first-rate description of a language. However, not every grammar achieves a reasonable level of reliability. Someone who has themselves written a grammar knows what to look for in another's work—consistency, plausibility, clarity, explicitness.

When pursuing typological studies—when essaying a further contribution to basic linguistic theory—the linguist will consult a wide variety of sources. One must take care to use only good and reliable descriptions, to read them carefully in order to understand exactly how a particular category works in the language under consideration, and to quote accurately.

Not everyone follows these principles. The 'armchair typologists'—who have not themselves undertaken fieldwork and written a grammar—tend to be the poorest. To quote one of very many examples, I was recently reading a linguistic monograph put out by a high-profile publisher and encountered two references to my work, both erroneous. First, I was cited as using a terminology which I abhor and have always argued against; and then there was a quotation which cited the right page numbers but the wrong book. Checking a little further, I followed up a phrase quoted from Maori, 'the chief's pig'. There were two errors in this one phrase, an accent missed off a noun and the subordinate possession marker given as *o* in place of *a* (the point being made in the original is that this type of possession uses marker *a* rather than *o*). And so on. No doubt many things are quoted correctly in this and similar works; but the fact that there is a significant batch of errors makes the work unreliable as a source of typological generalization.

Some of the grammars I have written get referred to a good deal. About 30 per cent of citations involve an error of some kind. Quite often what happens is that someone doesn't bother to go back to the original monograph but instead copies information from a secondary source which may have included one or more mistakes. The errors are thus perpetuated, and perhaps compounded.

When Winston Churchill (1951: 616) was writing *The Second World War*, he found that his memory of events did not always exactly correspond with the actual record of what had happened. 'It was only when I got home and searched my archives that I found the facts as they have been set out here. I am reminded of the professor who in his declining years was asked by his devoted pupils for his final counsel. He replied, "Verify your quotations".'

In scientific work, one should never rely on secondary sources. If Gikam-Faker quotes Matisoff as saying something about Lahu which is relevant to a topic you are working on, do *not* just rely on what Gikam-Faker says; go to the library and consult Matisoff's 1973 grammar, to get full and correct information from the horse's mouth, as it were.

My policy of always checking every secondary source yields some interesting results. Joseph H. Greenberg was revered by his students as someone who had an enormous number of facts about languages in his head. Ah!, but were they correct? In discussing coincidental similarities between languages, he states: ‘*man* means “man” in Korean.’ A nifty example, which I wanted to use. But one should always check in a primary source, so I phoned a Korean colleague. Sorry, he didn’t recognize it. I photocopied the page from Greenberg and sent it over. Definitely not in the modern language so he looked up dictionaries of Middle Korean. No, nothing at all. (There is a Sino-Korean bound root *nam-* ‘man’, which requires a suffix; this may possibly relate to the error.) People continue to quote this piece of nonsense; those who don’t follow the golden rule of always checking back in a primary source. The more this—and other—errors appear in print, the more veracity they appear to obtain.

As incorrect statements in the secondary literature get further propagated, the error may intensify. In 1952, Archibald A. Hill examined the oft-repeated tenet that Cherokee has a dozen or more terms for washing parts of oneself, or of something else, but no general term ‘wash’, this lack showing that in a ‘primitive language’ there are a multiplicity of specific but no general terms. Hill demonstrates that there are in fact two verbs each with a general meaning—*-wo* ‘bathe’ and *-e* ‘wash’—and these can take an incorporated noun (e.g. ‘face-wash’, ‘clothing-wash’), or a classifier, or a reflexive marker, yielding complex (but analysable) verb forms which were taken as unanalysable by an 1820s observer, with this dictum being repeated many times since—even in linguistics texts of good repute—as an instance of ‘primitiveness’. The people quoting didn’t think to check with any of the good publications on Cherokee since the 1820s!

Possibly the most pervasive myth is that the Eskimos don’t have one general term ‘snow’ (as do speakers of ‘sophisticated languages’ like English) but instead dozens, or hundreds—depending on the degree of generosity of the reporter—of specific terms. In fact, the number of terms for ‘snow’ in Eskimo is two (there are many complex forms based on these, as indeed in English we have *snowdrift*, *snowstorm*, *snowflake*, *snowfall*, *snowbank*, *snowball*). In 1986, linguist Laura Martin published a scholarly rebuttal of the popular belief about the myriad words for snow in Eskimo, and Geoffrey K. Pullum has shouted her dictum as loud as only he knows how (in a 1989 article and a 1991 book, entitled *The Great Eskimo Vocabulary Hoax, and other irreverent essays on the study of language*). But the myth is now in the public domain and surely nothing will arrest its piquant appeal. However, it does concern language, and it is important that anyone who calls themselves a scholar of linguistics should recognize this and other fables for what they are.

Misrepresentation can proceed further, in ways that would be unbelievable if not attested. In 1978, a linguist by the name of Janice Jake published a short paper ‘Why Dyirbal isn’t ergative at all’ (in fact, the identical paper was published in two different journals). The argument hinged on a sentence she provided, the meaning she assigned being ‘man told woman to scrape beans’. All of her information on Dyirbal came from my 1972 grammar in which this sentence does *not* appear; Jake had simply manufactured it. It is, in fact, not a grammatical sentence in Dyirbal; it could be interpreted—as many ungrammatical sentences can be—but could then only mean ‘man told woman [to do something] and [he] scraped beans’. (And there were other things made up by Jake in her paper, besides various errors of copying.)

As has been shown, a good deal of published work on linguistics is sloppy and unreliable. The Jake paper is an extreme example; but in well over half the papers and books coming out nowadays there are substantial errors of fact and/or interpretation. Yet the discipline is important and worthwhile. Linguists who themselves undertake fieldwork and write comprehensive grammars learn the value of accuracy and achieve the ability to distinguish the reliable from the unreliable in other people’s work. And, as a matter of habit, they follow Churchill’s advice always to check back in primary sources.

2.3 Unfashionable issues

People tend to follow one another in what they say, do, and think. If there is a habit of doing things in a certain way, then people will do them in this way. So it is with language description.

The basic medium of linguistic communication is speaking. Everyone who uses a language speaks it, and always has (for—probably—hundreds of thousands of years). Relatively recently, languages have been written down. Writing systems vary in their efficiency. Those devised and implemented by professional linguists are generally excellent. But many languages have a ‘traditional’ writing system which is less than fully adequate. For a linguist to use an inadequate traditional orthography is a lazy way out; unfortunately, some people do follow this practice. For example, Fijian has a contrast between long and short vowels, as can be seen in *maacawa* ‘interval of time, week’ and *macawaa* ‘worthless’. But this is not generally shown in the orthography, speakers writing *macawa* for both. One linguist who worked on Fijian simply used the everyday spelling. When questioned about this, the reply was, ‘when people are telling me how to spell words they don’t mention vowel length’. Maybe not, but one does have an ear to listen (and a linguist should have a trained ear).

Save for this feature, Fijian orthography is satisfactory. The writing systems used for Spanish and Portuguese are also pretty good. Not perfect, but a linguist only has to modify them a little for full reliability (for example, specifying whether a particular instance of ‘x’ indicates *ks* or *z* or *s* or *f* (‘sh’)). English orthography is, in contrast, utterly inadequate as a representation of the modern-day spoken language. Hundreds of examples could be quoted; to mention just one, /ɛə(r)/ in British English is variously shown as *ear*, *are*, or *air* (as in *wear*, *pear*, *glare*, *pare*, *fair*, and *pair*).

I know of only one textbook on English—and that is directed at foreign learners—which uses a phonological writing system. All others employ the everyday orthography. Now if a student working on a previously undescribed language from Africa or New Guinea or Amazonia submitted their dissertation in terms of an orthography as inadequate as that used for English, it would be failed. But virtually every discussion of any point in English grammar follows the lazy person’s option. Sure, it’s familiar, but it is insufficient as the basis for scientific enquiry.

Quite apart from the representation of vowels and consonants, standard English orthography does not properly represent what is a word. *The* is written between word spaces but in fact—unless stressed, which happens rather rarely—it is a clitic, making up one word with what follows; for example, *ðə=mán* (where ‘=’ indicates a clitic boundary), written as *the man*. And this lack of attention to the phonological reality of the language impedes understanding of how the grammar works.

We can once more consider phrasal verbs, briefly mentioned under (a) in §1.10. One may say either *Tom made the story up* or—moving the preposition *up* before the object—*Tom made up the story*. Now substitute pronoun *it* for *the story*. We can say *Tom made it up*, but not **Tom made up it*. Why not? There is a phonological reason for this grammatical constraint. *It*, like other object pronouns, is not a word in its own right but rather a clitic which attaches to the previous word—*Tom made it up* is /tóm méid=it ʌp/. As mentioned in §1.10, the rule is that a preposition can be moved to position between a verb and an object NP (like *the story*) but it may not intrude into the middle of a verb-plus-object.pronoun.clitic sequence since this constitutes one phonological unit. Working in terms of a phonological representation helps the linguist to perceive things which would not be apparent if viewed through the everyday orthography.

In some languages lexical words—and often, certain grammatical distinctions—can be differentiated by tones; these are often (although not always) marked in the orthography. In all languages, grammatical functions and meanings are—in part—marked by sentence stress and intonation. Punctuation marks may provide a little information, but never anything like

enough. Consider the English sentence *We should go*, said in three different contexts (several more could be provided):

- (7) (a) 'We should go?
A question, expressing surprise (the speaker could continue: *why should it be us that have to go?*). Sentence stress on *we* and rising intonation on *go*.
- (b) We should 'go
A straightforward expression of obligation, which will be fulfilled (*it is our duty and we will do our duty*); sentence stress on *go* and level intonation.
- (c) We 'should go
A statement of obligation that is likely not to be fulfilled (*but in fact I don't feel like going*); sentence stress on *should* with slightly falling intonation on *go*.

In an examination of the grammar of English (or of any other language), it is desirable to state the pragmatic import of each sentence discussed, in something like the way shown here.

Looking back on the linguistics of a hundred years ago, we can note various inadequacies—little idea of the phoneme, not much in the way of a typological theory of syntax, no incisive understanding of topics like transitivity and ergativity. What will linguists of a hundred years hence say when they look back on us? How could those linguists at the beginning of the twenty-first century—they may remark—have discussed the structure and meaning of sentences without stating the full textual context in which each occurred? How could they have expected to gain any insights into the organization of a grammar through quoting everything in an imperfect orthography, rather than employing accurate phonological transcription? All this and more besides. For instance, how could they have paid no attention to the elusive matter of sound symbolism, which tells us so much about the relation between language and the world it is used in?

It is today well recognized that a small number of expressions in each language are onomatopoeic. For example, certain names of birds and other creatures reflect the sound they make. In Dyrbal *biyilbiyil* describes a bird called in English *pee-wee* (or *magpie lark*, *Grallina cyanoleuca*), the names in the two languages reflecting slightly different impressions of the bird's call. But, leaving aside onomatopoeia, it is generally believed that the relation between sound and meaning is arbitrary. Note—people say—the varying names for an equine quadruped: *horse* in English, *cheval* in French, *Pferd* in German. Surely—they

continue—this shows that there is no connection between the sounds which make up a name, and the reference of the name.

Arbitrariness may prevail in parts of a language. But in certain areas there is without doubt an association between sound and meaning. In §1.6, the forms in Dyirbal for short, medium, and long distance uphill and downhill were given, being distinguished just by choice of final vowel: *a*, *i*, or *u*. I've asked all manner of people which vowel they would associate with each gradation of distance. Everybody plumps for *i* to indicate short distance, and the majority choose *u* for long distance, leaving *a* for mid-distance. 'You know this part of the grammar already,' I tell them, 'nothing to learn.' A general intuition is that a front high vowel should be associated with something close to the speaker. This applies also to demonstratives: compare *this* and *that* in English, *nih* and *núh* in Khmer, *ii* and *uu* in Telugu.

And sound associations may also apply to lexemes. It has been suggested that a high front vowel such as *i* is naturally associated in people's minds with a little or small object and a high back vowel such as *u* with a big or large one. However, this appears to be contradicted by English—*big* and *little* involve the same vowel and *large* and *small* similar ones (*a:* and *ɔ:* respectively). Maybe there is more to it than just vowels.

I distributed a questionnaire to a freshman linguistics class (in their first week). They were asked to guess which of two forms means 'big' and which 'little' in each of a selection of languages. The questions were repeated another year with the order of forms reversed. Fairly consistent results were obtained with samples of 38, 91, and 36 people. Overall, there was 70 per cent selection of the right term. In Tzotzil—a Mayan language from Mexico—83 per cent correctly identified *muk'ta* as 'big' and *bik'it* as 'little' one year, 73 per cent the next, and 88 per cent the third year. For Gumbaynggir, from south-east Australia, 74 per cent correctly chose *barway* as 'big' and *dyunuy* as 'little'. For Burushaski, a language isolate from Kashmir, 65 per cent gave *uyum* for 'big' and *lukon* for 'little'. Of the fourteen languages tested, correct identifications were made for eleven, with an average score of 77 per cent. For three languages the choices were correct for less than half of the respondents; only 42 per cent identified *ti'u* as 'big' and *mea'me'a* as 'small' for Rotuman, an Austronesian language from the Pacific. For these three the average correct score was 43 per cent, giving 70 per cent for the fourteen languages overall.

Analysis of the results shows that it is not just vowels—consonants also play a role. 'Big' is associated with sounds articulated on the periphery of the mouth: semi-vowel *w*, bilabials *b*, *p*, and *m*; dorso-velars *g*, *k*, and *ŋ*, and back vowels (often also involving lip-rounding) such as *u* and *o*. 'Little' is associated with high front vowels *i* and *e*, and with palatal and alveolar consonants such as *y*, *d*, *t*, and *dy* (*j*).

These figures are certainly significant. In the majority of the languages tested, most respondents were able to perceive which word meant ‘big’ and which ‘little’. I then tried a different lexical pair, words for ‘sharp’ and ‘blunt’ in nine languages. The overall score here was 61 per cent correct, but with wider variation than for the dimension adjectives. Six languages recorded a very high count, with an overall average of 81 per cent; for example, 87 per cent identified *hux* as ‘sharp’ and *muxul* as ‘blunt’ in Tzotzil. But for the other three languages scores ranged from only 8 per cent recognizing *muniṗoy* as ‘sharp’ and *sulik^w* as ‘blunt’ in Yurok (from California) to 31 per cent correctly identifying *hirom* as ‘sharp’ and *ṗfot* as ‘blunt’ in Burushaski. It seems that there is a significant degree of sound symbolism underlying ‘sharp’ and ‘blunt’ in some languages, but not in others.

There can be no doubt concerning a sound symbolic basis for many meaning contrasts in grammar and in lexicon. No proper theoretical framework has yet been devised to deal with this; as a result the topic is often not even mentioned in textbooks on linguistics. There is not enough to say anything scientific about it; but there is surely too much to ignore. The topic must be extended beyond a collection of anecdotes if we are to make headway in an understanding of language as the pre-eminent cultural tool of the human race.

What each fieldworker should do is note all instances of sound symbolism which they perceive (rather than ignoring them, as many have in the past). They would thus be helping to provide a store of basic information which can be input to attempts at inductive generalization, which may lead to the postulation of a theoretical model to describe and explain the phenomenon.

2.4 Avoid the fashion fads

It is wise to utilize a wide array of descriptive tools, rather than constrain linguistic description within a procrustean jacket. Nikolai Trubetzkoy, in his outstanding work *Principles of phonology* (published in 1939, just after the author’s death), explained how some contrasts are bilateral and others multilateral. For example, the opposition between *d* and *t* is bilateral since no other phoneme shares their common features (apico-alveolar, stop). But *b*, *d*, and *g* enter into a multilateral opposition since everything common to two of these sounds (voiced, stop) is also shared by the third. It is, of course, appropriate and useful to distinguish between oppositions involving just two terms, and those with more.

Trubetzkoy’s colleague Roman Jakobson developed the idea that all oppositions should be binary. This meshed well with the post-Bloomfieldian doctrine—put forward in the 1940s and 1950s—that a sentence should be divided into two immediate constituents, each of these into two further

constituents, and so on. Chomsky and many of his followers continue the binarist approach. In an intellectual endeavour, one *can* do almost anything. If one tries hard enough, a multilateral opposition *can* be reduced to a succession of binary splits. For example, quite a few languages have a three-term system of demonstratives according to whether the thing referred to is near, mid-distant, or far from the speaker. There are three ways of restating this in terms of two binary splits:

- near versus non-near, and then mid-distant versus far within non-near
- far versus non-far, and then near versus mid-distant within non-far
- mid-distant versus non-mid-distant, and then near versus far within non-mid-distant

How should one choose between the three alternatives? Or rather, *why* should one force oneself to choose? What can be gained by denying that there is a straightforward opposition between three terms? That is, beyond satisfying the strictures of a theoretical model which demands that all oppositions must be binary, thus making the language structure appear homogeneous. But languages are not neat and tidy; they involve all manner of different kinds of arrangement which should be described in the most insightful way. The fad of binarism can be an impediment to this.

The most insidious fad which has infiltrated linguistics during past decades is the idea that every language has an underlying structure involving a fixed order of phrasal constituents (often mislabelled ‘word order’), and that the ordering of elements is one of the (or is the) most fundamental typological feature(s) of a language. The fad has two parts, which arose at about the same time from quite different directions.

Modern linguistics has its roots in study of the classical languages of Europe, which allowed a remarkable freedom in the order of clausal constituents, and also of words. This was illustrated under (b) in §1.10, where it was shown that in the Latin sentence *Caesar Brutum monet* ‘Caesar warns Brutus’, subject, object, and verb can occur in any order, with no substantial difference in meaning. The basic constituents of a clause can, of course, consist of more than one word, as in:

- (8) [Hominēs obēsī]_A [cēnās
 person+NOM+PL fat+NOM+MASC+PL meal+ACC+PL
 magnās]_O edunt
 big+ACC+FEM+PL eat
 Fat people eat big meals

(Note that *edunt* fuses the verb root ‘eat’ with six grammatical specifications: present, active, indicative, plus 3rd person plural subject.)

There are three parameters relating to ordering, in any language:

- (a) The order of words in a phrase. Some languages, such as English, require words to be in fixed order within a phrase; here *fat* must precede *people*, and *big* must precede *meals*. In contrast, Latin allows *hominēs* and *obēsī* to occur in either order, and also *cēnās* and *magnās*.
- (b) The order of clausal constituents (phrases) within a clause. There can be a fixed order, as in English where the subject (here *fat people*) must precede and the object (*big meals*) must follow the verb. In contrast, Latin allows the subject, *hominēs obēsī*, the object, *cēnās magnās*, and the verb, *edunt*, to occur in any order; it has free ordering of phrasal constituents within a clause.
- (c) The order of words within a clause. Some—but not all—languages which allow freedom of words in a phrase and freedom of clausal constituents within a clause go one step further, and permit words to occur in any (or almost any) order within a clause. (Which words together make up one phrase is likely to be inferable from their bearing the same inflection.)

It is important not to confuse the parameters. There has arisen the habit of referring to the ordering of clausal constituents as ‘word order’. Following on from this, a language which has free order of phrasal constituents in a clause—but not free order of words in a clause—is said to have ‘free word order’. No distinction can then be made between a language with free constituent order, under (b) above, and one which really does have free word order, under (c).

Some of the same American linguists who espoused binarism (including the Chomsky school) paid especial attention to English and other languages with fixed constituent order. They maintained that this surface-structure requirement in English implied that the underlying (or ‘deep’) structure should also involve a fixed order of elements. And that languages which do not have a fixed order in surface structure should nevertheless be assigned one for their underlying structure. As mentioned in §1.10, this leads to the unproductive—and mildly ludicrous—scenario of positing a fixed order in underlying structure (and having to decide what this should be) and then saying that phrasal constituents—and, in some languages, words—can be scrambled into any order in surface structure. (A special terminology came into play; languages like English with fixed order were said to be ‘configurational’ and those like Latin and Dyirbal with free order were labelled ‘non-configurational’.)

The emphasis on order of elements in surface structure was accentuated through typological studies (from 1963) by Joseph Greenberg. He examined

‘the order of meaningful elements’ in a sample of thirty languages (not well chosen, there being none from New Guinea, the USA, or Canada). Greenberg first examined the ordering of subject (his S), object (O), and verb (V). He did not distinguish between transitive subject (my A) and intransitive subject (my S), arguments which do not always fill the same slot in surface structure; for example, some languages have preferred sequences SV and OVA, or VS and AVO.

Of the six possible permutations for his S, O, and V, Greenberg found that only three were common—VSO, SVO, and SOV. He then put forward a number of ‘implicational universals’, including: ‘Languages with dominant VSO order are always prepositional. With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional. In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes.’ This was an interesting typological study but with a limited sample and relating only to surface structure. (Later work turned up a fair few exceptions to Greenberg’s ‘universals’.)

But the classification of languages into VSO-type, SOV-type, etc. became an obsession with linguists. A language would be referred to as having ‘SVO word order’, although ‘S’ and ‘O’ (and often ‘V’) relate to clausal constituents which may be phrases rather than words. (Greenberg himself did not misuse ‘word order’ in this manner.) It came to be believed that the most fundamental question to be asked about any language was ‘what is its word order?’ Most other (surface-structure) properties should then—it was believed—be inferable—whether employing prepositions or postpositions; whether a genitive constituent preceded or followed the head of its noun phrase; and so on.

Ethnologue—mentioned in the preface—is the only purported full listing of the languages of the world. Typically, *Ethnologue* provides just one piece of typological information concerning a language, the ‘basic order’ of clausal constituents: Korean is SOV; Maasai (from Kenya) is VOS. Macushi, a Carib language from Brazil, is given as OVS, despite the excellent grammar of this language specifying that the ‘basic orders’ are OVA (although AOV also occurs frequently) and SV.

There are two difficulties concerning this modern-day fixation on constituent/word order. The first is that it means different things in different languages; the second that it conveys nothing of significance concerning the underlying character of a language.

Some languages, including English, do have fixed constituent order which plays a major role in marking which NP is in which function. But, as mentioned in §1.10, other languages have no such constraint. Phrases can occur in any order in a clause and/or words in any order in a phrase; plus, sometimes,

words in any order in a clause. The actual order of elements in any particular clause may relate to which piece of information is being highlighted, and it can be—at least in part—at speaker's whim.

However, many 'word order' typologists like to classify each and every language as SOV, SVO, VSO, OSV, OVS, or VOS. How to decide which, if there is in fact no fixed order? It may be taken as the order which involves least morphological marking; or the statistically most frequent order in texts; and so on. Dyirbal allows any order of words in a clause, save that certain particles (including negation) must precede the verb. Some orders are, of course, more common than others in texts. If both A and O are pronouns, then AOV order predominates, but if A and O are both NPs with a noun as head, then OAV is the most frequent order (in intransitive clauses SV is more common than VS). How then should Dyirbal be categorized? As 'OSV', one linguist decided, since the order of nouns is taken to be more crucial than the order of pronouns. But why should this be? No reason is given. Apparently, the end justifies the means. (Any means?) Like every other language, Dyirbal *has* to be placed in a 'word order' category. It will be seen that this so-called 'word order' typology may be of little value.

There is no doubt that the order of elements in surface structure is an important feature for some languages. But it must be realized that it is of minor interest for others. In *A grammar of Boumaa Fijian* I wrote as follows.

The question as to what is the basic order of transitive subject (A) and transitive object (O) NPs is one often posed by linguists nowadays. It is not a very important or even a very real question as far as Fijian is concerned . . . About 70 per cent of clauses in texts are intransitive, with a further 5 per cent being agentless passives. Further, of those clauses which are transitive, only a small fraction (10 per cent or less) have explicit A and O NPs. That is, overall only about 2 or 3 per cent of clauses are likely to have A and O NPs. And some of these—perhaps one in four—may have one of the NPs [topicalized and] fronted before the predicate, giving APO or OPA order [where P stands for predicate].

Of that small number of clauses that do have A and O NPs after the predicate in my textual corpus, about half have the order PAO and half the order POA. In one story the narrator twice used a clause with a predicate head *tau-ra* 'hold, possess' and two explicit core NPs, subject (A) *o Boumaa* and object (O) *a drano* 'the lake'. On first mention he said *sa tau-ra o Boumaa a drano* 'Boumaa held the lake' (PAO order) and then, eleven lines later *sa tau-ra a drano o Boumaa* 'Boumaa held the lake' (POA order). (In neither instance had either of the NPs, *a drano* or *o Boumaa*, occurred in the immediately preceding part of the text.) We conclude that, on a text count, Fijian cannot be said to have either AO or OA constituent order.

But even if every language did have a fixed or preferred order of clausal constituents (which is not the case), a typology based on this would be

of only marginal interest for basic linguistic theory. Scholars who view the description of languages as a branch of natural science have primary interest in the underlying grammatical relations and construction types, and the way in which these are linked together. How they are accorded surface clothing is of secondary interest.

2.5 Avoid sloppy terminology

A vital requirement for all scientific work is a set of terminology which is clear and unequivocal. To use terms in a sense other than that normally associated with them, or without their being properly defined, can cause confusion. As just mentioned, the use of ‘word order’ to refer to order of phrasal constituents is an extreme example of sloppiness. It will be worthwhile to mention a few more typical instances.

(a) **Sentence and clause.** The unit ‘sentence’ is clearly identified in writing as what comes between two full stops (or periods). In the analysis of spoken language, a sentence boundary is often recognizable on prosodic criteria; for example, in Jarawara the final syllable of a sentence is nasalized and bears rising intonation.

A sentence involves a main clause—which can stand by itself as a monoclausal sentence (a ‘simple sentence’)—and, optionally, a number of subordinate clauses. A complement clause can function as one argument of a main or subordinate clause. A relative clause is a modifier within an NP which itself functions as an argument of a clause. Consider the following sentence:

- (9) (a) [On looking out of the window]_{ADVERBIAL CLAUSE}
 (b) [[[my best friend]’s mother [whom he loves]_{RELATIVE CLAUSE}]_{NP:A} decided [that she would drive to work]_{COMPLEMENT CLAUSE:O}]_{MAIN CLAUSE}
 (c) [since it was too wet to walk]_{CONSEQUENTIAL CLAUSE}

Here the main clause is (b), a transitive clause with predicate *decided*, which has complement clause, *that she would drive to work*, as its O argument and NP *my best friend’s mother whom he loves* as A argument. This involves a relative clause, *whom he loves*, and a genitive phrase, *my best friend’s*, modifying the head noun *mother*. The main clause is flanked by two subordinate clauses, *on looking out of the window*, and *since it was too wet to walk*.

All this is quite clear. But in recent decades some linguists have failed to distinguish between sentence and clause, using ‘sentence’ when ‘clause’ appears to be intended. One reads that ‘a sentence consists of an NP and a VP’; what is meant is that a main clause (and certain types of subordinate clause) involve

a predicate slot (which can be realized by a VP) and an argument slot (which may be realized by an NP).

Complementation was discussed, for English, in §1.9. It was seen that a clause may function as an argument in clausal structure, as in *Mary_A remembered [that she had fed the cat]_O*. Yet some people describe complementation as ‘a sentence functioning as a clausal argument’. This is loose and misleading use of labels.

‘Sentence’ and ‘word’ are familiar terms in everyday discourse, while ‘clause’ is more restricted, more the technical language of linguists. A linguist must employ ‘clause’, rather than simply using ‘sentence’ to cover both units.

(b) **Subject.** In logic, every proposition consists of a subject and a predicate. Subject is often defined as ‘the entity about which something is affirmed or denied’.

Leaving aside copula clauses (discussed in Chapter 14), every language has two varieties of verbal clause:

- transitive, with two core arguments—transitive subject (A) and transitive object (O)
- intransitive, with one core argument—intransitive subject (S).

In a transitive clause, that argument whose referent could initiate or control the activity (if anything could) is recognized as being in A function. And if there is something which is saliently affected by the activity, the argument referring to this will be in O function.

Linguistics grew up linked with logic, in the study of Greek and Latin, which are strongly nominative-accusative languages—A and S are here marked in the same way (by nominative) and O differently (by accusative case). But in an absolutive-ergative language, S and O are marked in the same way (by absolutive) and A differently (by ergative case). Within a cross-linguistic typology, one *has* to recognize A, S, and O as universal syntactic relations. (These letters are simply abbreviations; one could always use ‘intransitive subject’ in place of S, for example, just as one could always write out ‘United States of America’ in place of ‘USA.’)

My experience is that even in a language with strongly accusative character—such as English—there is significant advantage in working with A, S, and O. To mention one simple example, the nominalizing suffix *-ee* originally referred to something in underlying O relation, such as *employee*, *nominee*; it has recently been extended to also apply to an underlying S, as in *escapee*, *retiree* (but not to an underlying A).

A, S, and O are the basic relations. As a secondary step, A and S are grouped together as ‘subject’. In even the most ergative language, there is some need for

this grouping. For instance, imperative always allows A or S to be 2nd person (sometimes only 2nd person is permitted). And in the variety of reflexive which involves a reflexive pronoun, this will always be in O or a peripheral function, with the A or S antecedent maintaining its full form. ‘Subject’ is a relevant and useful notion, to be used in addition to A, S, and O. But to always refer to ‘subject’—without a distinction between A and S—is to sacrifice a degree of explicitness, descriptive clarity, and scope for explanation.

(c) **Ambitransitives.** In some languages, every verb has a fixed transitivity; some only occur in transitive and the remainder only in intransitive clauses. In other languages there are a number of verbs which can occur in either kind of clause. These are termed ‘ambitransitives’ (‘labile’ is an alternative label found in the literature). However, it is not enough to say that a verb is ambitransitive; one must specify whether it is of type S = A or of type S = O. Consider the following sentences in English:

- (10) (a) John_S has eaten
 (b) John_A has eaten lunch_O
- (11) (a) Mary_S tripped
 (b) John_A tripped Mary_O

That transitive argument which corresponds to the intransitive argument, S, varies; it is A for the S = A ambitransitive verb *eat*, in (10), and O for the S = O ambitransitive *trip*, in (11).

Some writers of grammars mention ambitransitive verbs without specifying of which type(s); this is only half the story. The reader may then try to ascertain which verb is of which type by examination of the glosses in English (or Spanish or whatever); this is not a good thing to have to resort to, and can yield wrong results.

Some languages have further kinds of ambitransitives. A verb, which can be transitive, may be used with one core argument in an intransitive clause, and this has a reflexive sense (effectively A = O, becoming S). Compare *Mary_A hid [the money]_O* and *Mary_A hid herself_O* with *Mary_S hid*. Other basically transitive verbs may be used intransitively (then requiring a non-singular subject) with a reciprocal sense; compare *Mary_A hugged John_O* and *John_A hugged Mary_O* with *[John and Mary]_S hugged*.

(d) **Active/stative.** In most languages the intransitive argument, S, is either marked in the same way as A, a (nominative-)accusative system, or in the same way as O, an (absolutive-)ergative system. However, there are languages in which some verbs mark S like A (this can be called Sa) and others mark S like O (So). The Sa verbs are often called ‘active’; they typically (but not exclusively) refer to volitional activities such as ‘go’ and ‘enter’. The So verbs are then called ‘stative’; they predominantly refer to non-volitional activities such as ‘fall’ and ‘weep’. This can usefully be called a ‘split-S’ system, since the

sole core argument of an intransitive clause has split marking, according to the subclass of intransitive verbs it relates to.

However, there is a further group of languages, of fluid-S type. That is, some intransitive verbs must mark their core argument as Sa, some as So, but others allow either of these possibilities, with Sa generally referring to a volitional and So to a non-volitional sense of the verb. For example, in the North-East Caucasian language Batsbi (or Tsova-Tush), the same verb can mean ‘slide’ (a volitional activity) with Sa marking and ‘slip’ (non-volitional) when marked as So. Other verbs of this type include ‘lose weight’ and ‘get drunk’, each of which can be accorded a volitional or a non-volitional interpretation.

As mentioned, split-S languages are often referred to as ‘active/stative’. But so are fluid-S languages, which is inexplicit and confusing. If the labels ‘active’ and ‘stative’ are to be retained (rather than Sa and So) then further designations along the lines of ‘split active/stative’ and ‘fluid active/stative’ are required (although ‘split-S’ and ‘fluid-S’ are considerably more wieldy).

(e) **Misuse of ‘ergative’.** As mentioned before, in some languages S and O are marked in one way (this is called ‘absolute’) and A in a different way (this marking is termed ‘ergative’).

Now in the early 1980s, a couple of proponents of Chomsky’s brand of formal linguistics examined verbs like *open* in English. One can say [*The window*]_S *opened* and also *John*_A *opened* [*the window*]_O, showing that *open* is an S = O ambitransitive verb. But they didn’t work in terms of A, S, and O, and were apparently unfamiliar with the literature on transitivity. They had heard the term ‘ergative’ but seem not to have studied standard works on the topic. They said that, in these examples, the functions of *the window* were ‘the ergative set’. Pullum (1988: 583) referred to this as ‘a truly crackbrained terminological revisionism’. I don’t believe that this was a deliberate ‘revision’ of the meaning of ‘ergative’. The most likely explanation is that these formalists didn’t understand the accepted meaning of ‘ergative’ and applied the term wrongly. First, ergative is normally applied when core argument A is marked differently from S and O (which are marked in the same way). It is not used in connection with an S = O ambitransitive verb. But if it were to be, then in the examples under consideration *John* (and not *the window*) should be ergative, with *the window* being absolute.

(f) **Predicate.** As mentioned under (b), in logic a proposition consists of a subject (what is being talked about) and a predicate (what is being said about it). Some linguists have taken over this terminology—binarists, in particular, since it provides a simple division of a sentence into two parts. Thus, in a sentence like

- (12) John sold the books that he didn't like to the shop on the corner in which his mother used to work to help pay off his credit cards so that he wouldn't be sent to debtor's prison

John would be the subject and everything else the predicate. Under a slightly more sensible approach, if one said that a sentence consists of a number of clauses and the main clause consists of subject and predicate, the predicate would be just *sold the books that he didn't like to the shop on the corner in which his mother used to work*.

Most linguists use the term 'predicate' in a rather different way. A clause has a predicate element, generally filled by a verb (although there may be other possibilities; see Chapter 11). Each predicate requires a number of arguments; in English these are realized by an NP or a complement clause. (In other languages they may relate partly to NPs/complement clauses and partly to bound pronouns; see (c) in §1.10.) An extended transitive predicate with verb *sell* in English requires three arguments, in A function—*John* in (12)—in O function—*the books that he didn't like*—and an indirect object in E function (see Chapter 13)—*the shop on the corner in which his mother used to work*.

Within the logic-type use of 'predicate', just one argument is picked out as 'subject' the other arguments being subsumed within the predicate. But which argument should be subject? In an accusative language like English, an S or A argument is obligatory in every clause and is felicitously taken to be subject. But in an ergative language it is most likely that the S argument is obligatory in an intransitive but the O argument in a transitive clause, with the A argument being omissible. Does one then take O as the subject of a transitive clause? Linguists paying fealty to binarism agonize over this problem.

Those who use 'predicate' in the linguistic sense, as *requiring* various arguments but not *including* any of them within itself, face no such quandary. A transitive predicate has two core arguments, A and O. In every language there are certain similarities between A and S and others between O and S; the nature of each set of similarities varies from language to language.

(g) **Infinitive and finite.** Finally, two terms should be mentioned which are used by grammarians in many different ways and can be the source of much misunderstanding. First, 'infinitive', a term going back to Latin grammarians, who employed it to refer to the form of a verb which—unlike the verb form in a main clause—was not marked for the person and number of a subject, and did show tense. This was a nominalization, which functioned as an indeclinable neuter noun, which can function as head of an NP.

In English, preposition *to* plus the root form of a verb was labelled 'infinitive', on the basis of translation equivalence from Latin. In fact *to* and *demolish* do not form a constituent in a simple sentence like *John wanted to demolish*

the building (*demolish* and *the building* do). Then the form without *to*, here *demolish*, came to be called the ‘bare infinitive’. But an infinitive in Latin was a nominalization which had some of the syntactic possibilities of a noun. The bare verb form *demolish* is not a nominalization; it cannot be preceded by an article (the nominalization of this verb is *demolition*).

In view of the great variety of current uses of ‘infinitive’, a number of grammar-writers find that the most satisfactory course is to shun the term. For those who feel a need to employ ‘infinitive’, an explicit definition is required, in terms of the grammar of the language under study.

The term ‘finite’ appears not to have been used in Greek and Latin grammars but was coined relatively recently, as a faulty back-formation from ‘infinitive’. The initial use has been traced to Lindley Murray’s *English grammar*, first published in 1795. In the fourth edition of 1798 we find ‘finite verbs are those to which number and person appertain’. This is complementary to the definition of ‘infinitive’, as lacking number and person marking; ‘finitive’ would have been a more congruent label.

The term has since been used in a wide variety of different ways. The most satisfactory definition is perhaps that in *The concise Oxford dictionary of linguistics* (Matthews 1997: 129): ‘any verb whose form is such that it can stand in a simple declarative sentence’; but not all uses in the current linguistic literature conform to this. And care would have to be taken in applying Matthews’s definition; for example, in English the verb in THAT complement clauses may be marked for all of the categories open to the verb in a main clause; but many people would prefer not to call a type of subordinate clause ‘finite’.

As with ‘infinitive’, a goodly number of grammar-writers prefer to shun the term. Anyone who cannot resist using ‘finite’ should clearly define it, in terms of the grammatical system of the language they are describing.

There are a number of other terms in current use which are assigned varying meanings by different linguists and are best avoided. They include ‘unaccusative’ and ‘unergative’, discussed in Chapter 13. There is further discussion of terminology in Chapter 5.

2.6 How to learn linguistics

In order to master the discipline of linguistics, one should carefully study the most insightful books and papers and, if possible, attend lectures by consummate practitioners—scholars who have themselves written grammars and also made a significant theoretical contribution. And one should closely study two or three good grammars.

It is not enough just to read. One should also take notes—summarizing what has just been read will assist an understanding of it and provide a

succinct source which can be referred back to on a later occasion. Since 1961 I have made notes on everything significant I have read (not on the poorer publications), in large hard-bound notebooks (these started at A and have now reached W).

A most important piece of advice is to read the classic literature, not just things put out recently. If you wanted to become a student of film, for example, and planned to familiarize yourself with a few hundred movies, it would not do just to watch every one of the films produced during the last few years. Each year sees a handful of outstanding films, and many poor ones. What one would do is attempt to view a selection of the best films from every era since movie-making became an art. So it is with linguistics. It is most certainly not the case that books published this decade eclipse everything which has gone before. Many of the latest books are on fad theories. These may be an instantaneous centre of attention but will fade into oblivion within a decade. Many other current publications, in any year, are simply of poor quality. Rather than basking in a just-off-the-press copy of McAndrew's *Amputated syntax*, much better to spend a few hours with Nida's *Morphology* from 1949; it may be a little dated—just like an old film—but it includes heaps of sound methodology, which is as relevant now as then.

One begins by reading the best textbooks and, if possible, attending a few good and inspiring lecture courses. Then carefully study a grammar or two. Once a basic understanding has been achieved (say, after a linguistics-packed BA or MA course), the student should undertake fieldwork on a previously undescribed (or scarcely described) language (see §2.1 and Chapter 9). Reading of classic texts and of sound grammars should continue, for the whole of a scholarly career.

A further note is in order. Each linguist should attain a sound overview of the whole discipline, not just of their area of specialization. Someone planning to work on a Tibeto-Burman language, for instance, should of course study some of the best grammars of languages from this family. But they should also carefully study grammars of a couple of languages from different parts of the world, which show a quite different linguistic profile.

Some suggestions for reading follow.

(a) Recommended Grammars

1. James A. Matisoff. 1973. *The grammar of Lahu*. Berkeley and Los Angeles: University of California Press. li, 673 pages; includes vocabulary but no texts.

A thorough study by a scholar who plainly knows this language well and also the whole Tibeto-Burman family; includes many keen historical comments and explanations. But not the easiest of grammars to read; there

are no interlinear glosses, and the list of symbols and abbreviations runs to ten pages.

2. Nora C. England. 1983. *A grammar of Mam, a Mayan language*. Austin: University of Texas Press. xi, 353 pages; includes text but no vocabulary.

A model descriptive grammar of a verb-initial language (spoken in Guatemala and Mexico) with ergative-type bound pronouns, and special classes of positionals and of relational nouns.

3. R. M. W. Dixon. 1988. *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press. xix, 375 pages; includes texts and vocabulary.

A user-friendly grammar of an Austronesian language whose phonology and morphology are fairly straightforward, but which has many points of interest. One fault is that it lacks an index (but does include a detailed table of contents).

I have published three other full-length grammars, all a little more challenging to the reader—of the Australian languages Dyirbal (1972) and Yidiñ (1977a) and of a language from the small Arawá family of Brazil, Jarawara (2004a).

4. William A. Foley. 1991. *The Yimas language of New Guinea*. Stanford, Calif.: Stanford University Press. xvi, 490 pages; includes texts but no vocabulary.

A quality grammar of a Papuan language, with a highly synthetic, yet agglutinative, verbal structure.

5. Geoffrey D. Kimball. 1991. *Koasati grammar*. Lincoln: University of Nebraska Press. xxviii, 640 pages; includes texts but no vocabulary.

A sound and cautious grammar of a language (spoken in Louisiana) from the Muskogean family, with split-S marking and switch-reference.

6. Francesca C. Merlan. 1994. *A grammar of Wardaman, a language of the Northern Territory of Australia*. Berlin: Mouton de Gruyter. xxi, 671 pages; includes texts and vocabulary.

A fascinating study of a language with pronominal prefixes to the verb, organized on a different basis from case marking on nominals.

7. Nicholas D. Evans. 1995. *A grammar of Kayardild, with historical-comparative notes on Tangkic*. Berlin: Mouton de Gruyter. xxv, 837 pages; includes texts and dictionary.

A masterly study of an Australian language with multiple case marking—an NP may be marked for its function in a subordinate clause, and for the function of that clause in the main clause, and so on.

8. David E. Watters. 2002. *A grammar of Kham*. Cambridge: Cambridge University Press. xxv, 477 pages; includes texts and vocabulary.

A comprehensive grammar of a Tibeto-Burman language from Nepal, with split-ergativity, transitivity alternations, and mirativity.

9. Alexandra Y. Aikhenvald. 2003. *A grammar of Tariana, from northwest Amazonia*. Cambridge: Cambridge University Press. xxiv, 705 pages; includes texts and vocabulary.

An inspired reference grammar of an Arawak language. The multilingual milieu in which it is spoken has given rise to a combination of genetically inherited bound pronouns and areally diffused case marking.

Aikhenvald has recently published *The Manambu language of East Sepik, Papua New Guinea* (2008), a culturally anchored reference grammar of a language with many fascinating features, including socially defined genders, and proper names which are 'owned'.

10. Nicole Kruspe. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press. xxv, 493 pages; includes texts and vocabulary.

An in-depth description of a language from the Aslian branch of the Austroasiatic family. It features two types of morphological process, a genetically inherited non-concatenative system plus a concatenative system acquired through extended contact with Malay.

11. Elizabeth Zeitoun. 2007. *A grammar of Mantauran (Rukai)*. Taipei: Institute of Linguistics, Academia Sinica. xviii, 551 pages, includes texts and list of affixes, but no vocabulary.

A comprehensive study of an Austronesian language from Taiwan, with rich morphology, fascinating properties of negators (for instance, a polar question will generally include a negator), and a number of techniques for complementation.

12. N. J. Enfield. 2007. *A grammar of Lao*. Berlin: Mouton de Gruyter. xxvi, 567 pages, includes texts but no vocabulary.

A fine grammar of an isolating language (closely related to Thai). Grammatical information which other languages code morphologically (including non-spatial setting) is here shown by particles. The syntax is centred around serial verb constructions.

Nine of these grammars are based on fieldwork in a speech community where the language is actively spoken on a daily basis—Lahu, Mam, Fijian, Jarawara, Yimas, Kham, Manambu, Semelai, and Lao. Others involved working with the last generation of speakers, in a community where the language was not being passed on to children—Dyirbal, Wardaman, Kayardild, Tariana, Rukai (and probably also Koasati). For Yidiñ, I worked with the last three fluent speakers (now all deceased). The next two grammars involve work with the last speaker, and on the available materials of a language long dead.

13. Kirsten Refsing. 1986. *The Ainu language: The morphology and syntax of the Shizunai dialect*. Aarhus, Denmark: Aarhus University Press. 301 pages; does not include texts (none could be recorded) or vocabulary.

Careful grammar of a dialect of this language isolate, spoken in northern Japan. Refsing is well aware of the difficulties of working with a single speaker and takes full account of these.

14. Marie-Louise Thomsen. 1984. *The Sumerian language: An introduction to its history and grammatical structure*. Copenhagen: Akademisk Forlag. 363 pages; includes catalogue of verbs and full information on textual sources.

A first-class summary of what is known (and what is not known) about this isolate language, which ceased to be used about 1,600 BCE. Scrupulous attention to all previous literature on the language, with discussion of points concerning which there is scholarly disagreement.

It is also worthwhile to mention some older grammars, which are models of their kind:

15. Edward Sapir. 1930. 'Southern Paiute: A Shoshonean language', *Proceedings of the American Academy for Arts and Sciences*, Volume 65, part 1, pages 1–296. Does not include texts or vocabulary.

Focuses on morphological processes (with little on syntax); an acute and insightful analysis, well repaying detailed study.

Sapir's other major grammar, of Takelma (1922), is equally engrossing and worthwhile.

16. Mary R. Haas. 1941. *Tunica*. New York: J. J. Augustin. 143 pages; includes text but no vocabulary.

A tour de force, being based on work with the very last speaker (who was able to provide fluent texts) of this isolate language from Louisiana. (In 1977 Mary Haas told me that she originally wrote the grammar at twice the length, but was required by Franz Boas—editor of the *Handbook* in which it was to appear—to condense it, which she did by omitting many examples and some paradigms.)

17. Franz Boas and Ella Deloria. 1941. *Dakota grammar*. Washington, DC: US Government Printing Office. xii, 183 pages.

Boas, who was essentially the founder of descriptive linguistics in the USA, here teams up with native-speaker-cum-linguist Deloria for a clear and insightful description of this Siouan language. All of Boas's other grammars—and his theoretical essays on linguistics—are eminently worthy of study.

18. L. S. Freeland. 1951. *Language of the Sierra Miwok*. Indiana University Publications in Anthropology and Linguistics, Memoir 6. (Supplement to the *International Journal of American Linguistics*, Vol 17, No. 1.) vi, 199 pages; includes texts but no vocabulary.

Effortlessly conveys the grammatical character of this language (spoken in central California), from the Utian family. Grammatical relations are shown by an interesting interaction of case system and pronominal suffixes to the verb.

A small number of grammars have been cast in terms of a formal theory; most could not be recommended. An exception is:

19. Anthony J. Vitale. 1981. *Swahili syntax*. Dordrecht, the Netherlands: Foris. 214 pages; includes neither texts nor vocabulary.

Clear, informative, and accessible account of some of the main features of Swahili syntax. Can be augmented by one of the standard accounts of morphology, such as Loogman (1965) or Ashton (1947).

(b) Books on Linguistics

1. Edward Sapir. 1921. *Language*. New York: Harcourt Brace. ix, 242 pages.

A finely crafted essay on the nature of language—sounds, words, meanings, grammatical processes, how languages develop, how they influence each other. Should be read by all linguistics students in their first semester, and recurrently restudied. (There is no book that I refer to more often.)

Almost everything that Sapir wrote is first-class. For instance, his *Selected writings* (1949), especially ‘Sound patterns in language’, ‘The psychological reality of phonemes’, ‘Language and environment’, and ‘Time perspective in Aboriginal American culture: A study in method’.

2. Leonard Bloomfield. 1933. *Language*. New York: Holt. [Reissued in 1984 by University of Chicago Press.] ix, 566 pages.

A masterpiece, now only slightly dated. The chapters on comparative-historical linguistics remain an unimpeachable introduction. Those on phonology, morphology, and syntax include many excellent insights and examples. One can readily grant allowances for such things as calling tones ‘phonemes’, and for limited ideas about meaning.

3. Nicolai S. Trubetzkoy. 1939. *Grundzüge der Phonologie*. Travaux du Cercle Linguistique de Prague. 7. [English translation by Christine A. M. Baltaxe: 1969. *Principles of phonology*. Berkeley and Los Angeles: University of California Press. xvi, 344 pages.]

If I had to choose one linguistics book for a desert island, this would be it. Of wondrous intellectual depth, combined with a fine cross-linguistic span. Reveals the structured nature of phonology: oppositions, markedness, units, and delimitation.

4. Eugene A. Nida. 1949. *Morphology: The descriptive analysis of words*, 2nd edition. Ann Arbor: University of Michigan Press. xvi, 342 pages.

A classic textbook, explaining the principles of morphological analysis. The procedural aspects need not be taken too seriously, but the principles involved are absolutely solid.

5. Bernard Comrie. 1976. *Aspect*. Cambridge: Cambridge University Press. ix, 142 pages.

Insightful analysis and argumentation, with detailed account of the semantics of aspect in Slavic languages.

Comrie's *Language universals and linguistic theory* (2nd edition, 1989) is perhaps the most reliable and readable introductory textbook on typology.

6. John Lyons. 1977. *Semantics* (two volumes). Cambridge: Cambridge University Press. xiv, 894 pages.

Parts of this are an intellectual feast—well-organized discussion, well-thought-out arguments. Clearly written, including many original (sometimes provocative) ideas. However, not too much attention is paid to unfamiliar languages (a large proportion of the examples are from English and French).

Lyons's other major work, *Introduction to theoretical linguistics* (1968), was a valuable publication at the time, but is now—unlike, say, Sapir's *Language*—somewhat outdated, especially sections on transitivity (and 'ergativity').

7. P. H. Matthews. 1981. *Syntax*. Cambridge: Cambridge University Press. xix, 306 pages.

8. P. H. Matthews. 1991. *Morphology*, 2nd edition. Cambridge: Cambridge University Press. xii, 251 pages.

These are both rewarding texts, providing a synthesis of the field, with incisive discussion of alternative approaches and argumentations.

The concise Oxford dictionary of linguistics by Matthews (1997) is an indispensable tool, with definitions that are informative and to the point (and sometimes also witty).

9. Timothy Shopen. 1985. Editor of *Language typology and syntactic descriptions*. Cambridge: Cambridge University Press. Three volumes x, 399; x, 317; xii, 427 pages.

Contains eighteen chapters covering most recurrent categories and construction types. As would be expected with a variety of authors, the quality varies, but overall this constitutes an indispensable source-book.

A second edition was published in 2007. A number of chapters are by the original authors, some only slightly revised (when a thorough updating was needed), some appropriately revised, and others completely rewritten (opinions will vary on whether the 1985 or 2007 version is the more useful). A number of chapters are by new authors, and are up to date and good; a couple of new chapters have been added. Unaccountably, chapters on causatives and on negation (highspots of the 1985 edition) have been omitted.

10. R. M. W. Dixon. 1994. *Ergativity*. Cambridge: Cambridge University Press. xxii, 271 pages.

An account of the nature of grammatical relations, and how languages may combine ergative and accusative characteristics at both morphological and syntactic levels.

Volumes edited by Dixon and Aikhenvald on ‘word’ (2002), ‘adjective classes’ (2004), and ‘complementation’ (2006)—and by Aikhenvald and Dixon on ‘serial verb constructions’ (2006)—each include, in the first chapter, an outline of the parameters of variation, followed by a series of informed case studies, each dealing with a language on which the author has undertaken extensive fieldwork (and, in almost every case, written a grammar of).

11. Peter Ladefoged and Ian Maddieson. 1996. *The sounds of the world’s languages*. Oxford: Blackwell. xxi, 426 pages.

A thorough and insightful study, describing every type of parameter for speech production, their interrelations, physiological bases, and functional correlates.

12. Greville G. Corbett. 2000. *Number*. Cambridge: Cambridge University Press. xx, 358 pages.

Sound cross-linguistic survey of how a number system may be realized in different areas of a grammar.

13. Alexandra Y. Aikhenvald. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press. xxvi, 535 pages.

A comprehensive cross-linguistic study of classifiers, noun classes, and genders. Consolidates previous work in the field, in terms of an original and overarching theoretical model.

Aikhenvald’s study of *Evidentiality* (2004) provides a theoretical overview of how around one-quarter of the world’s languages include in their grammar some means for specifying the evidence on which a statement is based (seen, heard, inferred, reported, etc.). And her monograph *Imperatives and commands* is due to be published in 2010.

Each student of linguistics should spend some time studying the writing of leading figures in the field. Sapir, Bloomfield, and Trubetzkoy were included above; five others should also be mentioned.

14. Ferdinand de Saussure. 1916. *Cours de linguistique générale*. Paris. 336 pages. [Available in two English translations, as *Course in general linguistics*.]

Notes by his students on Saussure’s lectures, arranged (and edited) into a book. A classic which helped define modern linguistics (although some of the ideas have been misinterpreted).

15. Otto Jespersen. 1924. *The philosophy of grammar*. London: Allen and Unwin. 359 pages.

An impressive synthesis of typological theory at that point in time. Asks important questions and provides thoughtful and thorough answers. It may appear a little outdated but it is still worthwhile examining what Jespersen had to say on virtually any topic.

Jespersen wrote several other important books, most notably his seven-part *A modern English grammar, on historical principles* (1909–49) which stands up against any modern grammar of the language.

16. Antoine Meillet. 1925. *La Méthode comparative en linguistique historique*. Oslo: Instituttet for Sammenlignende Kulturforskning. viii, 166 pages. [English translation by Gordon B. Ford, Jr. 1967. *The comparative method in historical linguistics*. Paris: Champion. 138 pages.]

Simply the clearest and soundest introduction to the principles of historical linguistics. Meillet wrote many other books (quite a few translated into English), all intellectually rewarding.

17. Émile Benveniste. 1966. *Problèmes de linguistique générale*. Paris: Gallimard. 356 pages. [English translation by Mary E. Meek: 1971. *Problems in general linguistics*. Coral Gables, Fla.: University of Miami Press. x, 317 pages.]

A collection of 28 essays, focusing mostly on Indo-European languages but with wide relevance. Each chapter makes an important theoretical point. As with Meillet—and also Jespersen, Boas, Sapir, Bloomfield, Trubetzkoy, and Jakobson—his many other writings are uniformly worthwhile.

18. Roman Jakobson. 1990. *On language* (edited by Linda R. Waugh and Monique Monville-Burston). Cambridge, Mass.: Harvard University Press. xx, 646 pages.

A felicitous selection from the voluminous writings of the scholar who introduced such terms as ‘structuralism’ (in 1929). Includes many classic papers, each a gem of insight and explanation.

(c) Papers on Linguistics

1. Yuen-Ren Chao. 1934. ‘The non-uniqueness of phonemic solutions of phonetic systems’, *Bulletin of the Institute of History of Philology, Academia Sinica* 4: 363–97. [Reprinted as pages 38–54 of Joos 1958.]

Both a fine exposition of the idea of phoneme, and an original and insightful discussion of alternative phonemic treatments for a given set of phonetic data.

2. Morris Swadesh and Charles F. Voegelin. 1939. ‘A problem in phonological alternation’, *Language* 15: 1–10.

Demonstration of how a seemingly irregular paradigm in the Uto-Aztecan language Tübatulabal can be explained by nifty use of morphophonemes.

3. Kenneth L. Pike. 1947a. ‘Grammatical prerequisites to phonemic analysis’, *Word* 3: 155–72; and 1952. ‘More on grammatical prerequisites’, *Word* 8: 108–21.

Masterly demonstration of how different levels of analysis interact. (Arguing against the ‘discovery procedures’ and ‘watertight levels’ of much American work of that period.)

4. Charles C. Fries and Kenneth L. Pike. 1949. ‘Coexistent phonemic systems’, *Language* 25: 29–50.

Describes how there can be distinct phonological systems applying to borrowed and to native forms in a language, and related issues.

5. Dwight L. Bolinger. 1950. ‘Rime, assonance and morpheme analysis’, *Word* 6: 117–36.

An important discussion of just how far morphological analysis may be taken (with allusions to sound symbolism). Bolinger was an outstanding linguist and his many other papers and books are heartily recommended.

6. Edward L. Keenan and Bernard Comrie. 1977. ‘Noun phrase accessibility and universal grammar’, *Linguistic Inquiry* 8: 63–99; and 1979. ‘Data on the Noun Phrase Accessibility Hierarchy’, *Language* 55: 333–51.

Based on a cross-linguistic study of relative clauses, establishes a hierarchy for determining which grammatical relations can be the common argument in a relative clause.

7. Paul J. Hopper and Sandra A. Thompson. 1980. ‘Transitivity in grammar and discourse’, *Language* 56: 251–99.

A searching and well-illustrated presentation of semantic parameters which interact to provide the basis for the distinction between transitive and intransitive verbs in individual languages. Note that transitivity is a syntactic category, which has semantic grounding. It is not useful to talk of ‘semantic transitivity’ (as Hopper and Thompson imply one could), only of the underlying semantics of verbs which are classed as transitive and as intransitive in each language. This is further discussed in Chapter 13.

8. Marianne Mithun. 1984. ‘The evolution of noun incorporation’, *Language* 60: 847–94.

An exciting cross-linguistic study of a phenomenon which lies at the intersection of syntax and morphology.

9. John Du Bois. 1987. ‘The discourse basis of ergativity’, *Language* 63: 805–55.

Shows that, although the topic running through a stretch of discourse is most likely to be in A or S function in each clause, when some new information is introduced into a discourse it is predominantly in S or O function for first mention.

Other important books and papers relating to particular grammatical topics are referred to in the appropriate chapters of Volumes 2 and 3.

Sources and notes

2.1. Fijian examples from Dixon (1988: 228). Modifiers such as *ruarua* are discussed in chapter 8 of that book (page 99) and object incorporation in chapter 18 (pages 226–9). ‘Eat’ is one of the few verbs in Fijian to have slightly irregular form; the intransitive root is *'ana*, which becomes *'ani-* before transitive suffix *-a*.

2.2. The erroneous statement that ‘*Man* means “man” in Korean’ is from Greenberg (1957: 36). This error is repeated in Hock (1991: 557)—with no source quoted but likely to have been taken from Greenberg—and then in Kumar and Rose (2000: 236–7)—quoting Hock. None of these authors followed the basic scholarly principle of checking back in a primary source. Cherokee words for ‘washing’ are discussed in Hill (1952) and Eskimo words for ‘snow’ in Martin (1986) and Pullum (1989, 1991). Dixon (1979) is my reply to Jake (1978a, reprinted as 1978b).

2.3. To the statement that ‘the basic medium for communication is speaking’ needs to be added recognition of ‘sign languages’ as bona fide languages on a par with spoken languages. They are of a quite different nature and demand to be described in terms of their own array of theoretical parameters (not as unusual instances of the parameters appropriate for spoken languages).

The English textbook in phonetic orthography is Palmer (1924, 1939), Palmer and Blandford (1969). (Do as I say, not do as I do. I’ve been as remiss as anyone in not using a fully phonological orthography in discussion of English grammar, and in not fully specifying the pragmatic context for each sentence discussed in every language I’ve written on.)

For sound–meaning correspondences in demonstratives, see Diessel (1999: 151–2) and further references there. For sound symbolism (also called phonaesthesia) see the important work in Bolinger (1950), Ohala (1984), Hinton, Nichols, and Ohala (1994), and further references therein. Note that Diffloth (1994) mentions that in Bahnar, a Mon-Khmer language of Vietnam, *i:* indicates ‘big’ and *a:* indicates ‘small’, an exception to the general sound–meaning correspondence. Berlin (1992: 232–59) provides a classic study of sound symbolism, especially in relation to names of flora and fauna.

Sources of information for the questionnaires on ‘big’ and ‘little’, and ‘sharp’ and ‘blunt’, are: Tzotzil, Haviland (personal communication 1969); Gumbaynggir, Eades (1979: 358); Burushaski, Lorimer (1938: 445–6, 479, 500); Rotuman, Churchward (1940: 264, 331); Yurok, Robins (1958: 275, 292).

Interesting ideas have been put forward concerning the connection between linguistic meanings and colours. See, for example, Reichard, Jakobson, and Werth (1949), Cytowic (1989), and further references therein.

2.4. Greenberg's influential publication on ordering was in 1963; quotation of 'universals' is from page 110. The list of languages used in the sample included 'Berber', but Berber is a sub-group within the Afro-Asiatic family which consists of about thirty languages. Greenberg does not follow normal scholarly practice and list the source used for each language.

Dixon (1994: 49–52) discusses languages which do not have A and S in the same slot in surface structure. *Ethnologue* data from the Grimes (2000) edition pp. 542, 140, 274. Macushi data from Abbott (1991: 24, 26). Mithun (1987) is a scholarly and well-reasoned dismissal of the claim that a basic 'word order' can be recognized for every language.

Fijian quotes from Dixon (1988: 242–3). This discussion goes on to remark that in elicitation the order POA predominated (although PAO occurred); and if either of A or O is a complement clause this will come last.

A language without bound pronouns requires core arguments to be expressed by NPs. Those with bound pronouns have less need of NPs. This is an important parameter which must be taken into account in discussion of types of constituent order.

2.5. Criteria for 'sentence' in Jarawara, see Dixon (2004a: 530). For difficulty in finding clear criteria for 'sentence' in some spoken languages, see Miller (1995).

As examples of the sort of things mentioned here, Craig (1977) talks of 'complement sentences' rather than 'complement clauses'; and Noonan (1985: 42) begins his chapter on 'complementation' saying 'by complementation we mean the syntactic situation that arises when a notional sentence or predication is an argument of a predicate'.

For further cross-linguistic properties of 'subject', see Dixon (1994: chapter 5). Split-S and fluid-S, see Dixon (1994: 70–83). For Batsbi (Tsova-Tush) see Holisky (1987).

The Chomskians who misunderstood the meaning of ergative were Burzio (1981) and Pesetsky (1982). The term 'ergative' first used in modern sense by Dirr (1912); see Dixon (1994: 3).

As examples of the different meanings assigned to 'finite', the current *Oxford English dictionary* defines it as '(of a part of a verb) having a specific number and person', whereas *The Oxford dictionary of English grammar* (Chalker and Weiner 1994: 151) provides the definition 'having tense' and the *Random House dictionary* proffers 'a verb form that distinguishes person, number and tense and also mood or aspect'.

3

Grammar Overview

This chapter provides an outline of recurrent grammatical structures found in human languages. There are a number of fairly rare structural types which are not mentioned, in order to keep the chapter—and book—of manageable size. They are all special variations in the basic template presented here. About two-thirds of the topics briefly considered in this chapter are accorded more detailed description in Volumes 2 and 3.

An observation made in §1.4 deserves repetition. No two languages are exactly the same, in any respect. Although a certain label may be used in description of different languages, it will have a slightly different role in each language. There will, of course, be a common element of meaning and function—which justifies use of the label—but extensions from this are language-particular.

It is not the case that every language shows every kind of construction type. Many languages have a complement clause construction, in which a complement clause fills an argument slot in another clause; for example *I_A heard [that John had died]_O* in English. Some languages lack such a construction type. What do they do instead? There are typically a number of complementation strategies—construction types which have a secondary role covering the meaning conveyed by a complement clause construction in other languages. See §3.10 and Chapter 18.

3.1 Basic units: word and clause

There are two absolutely basic units in terms of which any grammatical description must be framed—word and clause.

(a) **Word.** The ways in which words are joined together, to form clauses, is generally called **syntax**. However, in this book I prefer to reserve the term syntax for study of the organization and interrelation of grammatical elements, rather than just for their linear realization (which is a secondary matter). Study of the structure of words is called **morphology**, discussed in §3.13. ‘Word’ is thus the central unit, the intersection of syntax and morphology.

The criteria for ‘what is a word’ vary from language to language. Some recurrent criteria are phonological: for example, placement of stress, syllable structure, domain of application of phonological rules. These essentially define a ‘phonological word’. Other criteria are grammatical: the parts always occur together, in a fixed order, and have a conventionalized coherence and meaning. These define a ‘grammatical word’. In some languages, phonological word and grammatical word always coincide. In other languages they do so generally (hence the appropriateness of the label ‘word’ for both) but not always. We do find instances where one grammatical word is made up of two or more phonological words, and the reverse, where one phonological word includes more than one grammatical word. There is a full examination of the unit ‘word’ in Chapter 10.

(b) **Clause.** This is the description of some activity, state, or property. As mentioned under (a) in §2.5, the label ‘sentence’ is often used when what is intended is clearly ‘clause’. A simple sentence consists of a single clause; but many sentences include a number of clauses, linked together and/or with one embedded within another (see §3.11).

The grammar of a language has two components, syntax and morphology. Some linguists treat phonology as a third part of a grammar; others regard phonology as distinct from grammar, but linked to it. A feature can be called ‘morphosyntactic’ if it both occurs in a morphological paradigm and marks syntactic function; for example a system of case affixes. (Recently, there has arisen the habit of using ‘morphosyntax’ as an alternative name for ‘grammar’—that is, for morphology-plus-syntax—a usage which is unnecessary and can be misleading.)

3.2 The clause

A clause has syntactic and pragmatic function, and it has a structure. These will be discussed in turn.

(a) **Syntactic function.** A clause which on its own makes up a sentence, which can be a complete utterance, is called a ‘main clause’ (MC). There are various ways of combining clauses, illustrated in Figure 3.1.

- (i) *Relative clause* (RC), which modifies the head of an NP that fills an argument slot in another clause (which can be an MC or another subordinate clause). For example:

- (1) [John [who has studied the language for years]_{RC}]_A speaks German_O well

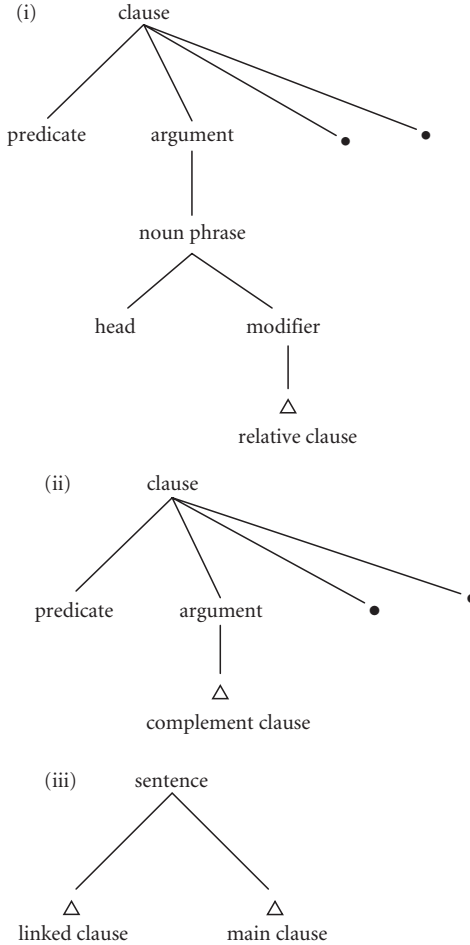


FIGURE 3.1. Types of clause combining

Note: Ordering is arbitrary in these diagrams.

Relative clauses are discussed in §3.4 and Chapter 17.

- (ii) *Complement clause* (CoCl), which fills an argument slot in another clause (again, either an MC or another subordinate clause). For example:

(2) I_A know [that John_A speaks German_O well]_{CoCl:O}

Complement clauses are discussed in §3.10 and Chapter 18. Note that clauses which are included within another clause—such as relative clauses and complement clauses—are said to be ‘embedded’.

(iii) A *clause linked to an MC*, the two forming a complex sentence. These can include temporal associations, as in

(3) [After he_A had studied it_O for years], [John_A could speak German_O well]_{MC}

Or the clause linkage may express contrast, as in (4), or consequence, as in (5):

(4) [Although he_A had not studied it_O for long], [John_A could speak German_O well]_{MC}

(5) [Since he_A had studied it_O for years], [John_A could speak German_O well]_{MC}

Clause linking is dealt with in the discussion of the sentence, in §3.11.

(b) **Pragmatic function.** A sentence generally includes an indication concerning what type of speech act it is; this is called its mood. There are basically three alternatives:

- a *statement*, with *declarative* mood (also called *indicative* mood, these terms often being interchangeable);
- a *command*, with *imperative* mood;
- a *question*, with *interrogative* mood.

Some languages have a morphological system (generally, of inflection on the verb) marking the three moods. However, many languages lack this. Declarative mood is typically left unmarked, as the default speech act. Imperative is often shown by a verbal suffix, and there may be different suffixes for positive and negative imperative, or polite and brusque, or near and distant.

There are two varieties of interrogatives:

- (i) A *content question*, shown by a content question word. In English these are *who*, *what*, *which*, *why*, *where*, *when*, and *how*. The label ‘*wh*- words’ is often used for the content question words in English since all except for *how* begin with *wh*-. (This label would not be appropriate for other languages.) There may, in addition, be morphological marking of a clause being a content question and/or a distinctive pattern of intonation.
- (ii) A *polar question* (sometimes called a ‘yes/no’ question; this is not a good general term since there are languages lacking words ‘yes’ and ‘no’). For example, *Can [your brother]_A speak German_O?* In many languages this is shown simply by intonation (typically, but not invariably, final rising intonation). Some languages have—in addition to intonation,

or instead of it—a polar interrogative affix or particle. English shows a polar question by reversing the order of the subject constituent and the first word of the auxiliary, in addition to rising intonation.

‘Declarative’, ‘interrogative’, and ‘imperative’ are grammatical labels, while ‘statement’, ‘command’, and ‘question’ describe type of speech act. A command typically involves a main clause marked grammatically as an imperative. But there are often other grammatical constructions which can be used to convey a command, with milder pragmatic effect; for example *Would you_A mind opening [the window]_O?* or *I_A wish [that somebody_A—directing gaze at a person—would open [the window]_O]_{CoCl:O}.*

It is important to carefully distinguish mood from modality. Mood—a property of the sentence—deals with speech acts of the three recurrent types declarative, imperative, and interrogative. Modality—which relates to a clause and its predicate—describes semantic distinctions within an irrealis specification. The modal auxiliary verbs in English are typical markers of modality, including *must* (necessity), *should* and *ought to* (obligation), *will* (prediction), and *can* (ability). Some linguists confuse mood and modality, conflating these two quite different types of specification. In fact, mood and modality have totally different meanings and functions, and should always be clearly distinguished. (The adjective ‘modal’ is appropriately used in relation to modality, not to mood.) There is discussion of modality in §3.15.

It was pointed out in Chapter 1 that a single grammatical system or single set of construction types may relate to a variety of semantic distinctions. For example, the four noun classes in Dyirbal code a dozen cultural contrasts, and the four main types of complement clause in English represent differences of meaning across a score of semantic types (§1.9). Under (f) in §1.10, attention was drawn to the fact that in languages such as Latin, Greek, and Sanskrit a single morphological system covers the marking of two different kinds of syntactic function—the function of an NP within a clause (these markers are appropriately called cases) and the function of an NP within another NP (for instance, genitive, which should then not be called a case).

A similar situation can occur with verbal affixes. In Dyirbal, for example, there is a single inflectional system on verbs which includes positive and negative imperative suffixes (mood), future and present/past suffixes (tense within declarative mood), purposive suffix (clause linking), and a suffix marking the verb in a relative clause. Thus, one system includes information about mood, polarity, tense, and subordinate clause type. This does not imply that relative clause marking is a type of mood, nor that purposive is a kind of tense, etc.

Other languages have an inflectional system which includes marking of mood and modality. (This helps to explain why mood and modality are

sometimes confused). And there can be a grammatical system some of whose terms mark mood while others indicate a kind of subordinate clause, such as conditional ('if'). From this it has sometimes been inferred that conditional is a kind of mood; however, inferences of this kind do not provide a fruitful avenue of analysis.

In Latin, what is called subjunctive is marked in a similar way to moods and has, as a result, been called a mood—on the mistaken principle that everything in a given morphological system must have the same syntactic status. The central meaning of subjunctive in Latin is to mark a type of subordinate clause; for example, *veniās*, 2nd person singular present subjunctive of 'come' in *Imperō ut veniās* 'I command you to come'. However, a clause with its verb inflected as subjunctive can, effectively, function as a main clause, often glossed 'let's —'. There are a number of possible analyses of this. One is to say that subjunctive has two functions, one marking a kind of subordinate clause and the other marking a special type of imperative (often called 'jussive' or 'hortative'). This illustrates that the divisions within a grammar are seldom neat and tidy, and that one morphological form may have several roles in the syntax of a language.

We need now to enquire about the interaction between pragmatic function and syntactic function.

- If a sentence has imperative mood, this is marked on the main clause. When, in English, direct speech including an imperative is recast as indirect speech we are likely to get a purposive complement construction. For example, *She_A invited [the youth]_O: 'Come and eat.'* becomes *She_A invited [the youth]_O [to come and eat]_{CoCl}*. This is not in itself an imperative.
- If a sentence is in interrogative mood, this will again be shown on the main clause. Some languages are like English in that one variety of complement clause is marked by a question word, as *She_A asked John_O [whether he_S was going]_{CoCl}* and *She_A asked him_O [who_S was going]_{CoCl}*. This does not imply that the complement clause has interrogative mood.
- Similarly, if a sentence is in declarative mood, this is shown on the main clause. Some subordinate clause may also have the structure of a main clause, plus a marker of that variety of subordinate clause. This applies to THAT complement clause in English; for example, *She_A knew [that John_S was going to resign]_{CoCl:O}*.

(c) **Structure.** Each clause has an internal structure, consisting of a predicate (which typically relates to a verb) and a number of arguments, some of which must be either stated or understood from the context—these are 'core arguments'—and others which are optional—these are 'peripheral arguments'.

The two major clause structures, across the languages of the world, are intransitive, with one core argument, and transitive, with two:

CLAUSE TYPE	PREDICATE	CORE ARGUMENTS
intransitive	intransitive	S (intransitive subject)
transitive	transitive	A (transitive subject) and O (transitive object)

As mentioned in §2.5, that argument in a transitive clause whose referent could initiate or control the activity (if anything could) is recognized as being in A function. And if there is something which is saliently affected by the activity, the argument referring to this will be in O function.

In (b) of §2.5 there was discussion of why we should in all languages distinguish between intransitive subject (S) and transitive subject (A). As mentioned under (f) of §2.5, in logic ‘predicate’ is used for everything in a sentence other than main clause subject (this stance is maintained by many linguists of Chomskian schools, in keeping with a predilection for binary divisions). In contrast, the standard practice in modern-day scientific linguistics is that the predicate does not include within itself the object or any peripheral argument.

The predicate is the nucleus of a clause. The word—generally a verb—that is placed in the predicate slot will determine the number and type of arguments which the predicate takes. In English, *go* is an intransitive verb, occurring only as predicate of an intransitive clause, taking a single core argument which is in S function. *Hit* and *think* are transitive verbs, taking two core arguments in A (transitive subject) and O (transitive object) functions. The meaning of the verb determines the kinds of noun which can fill a core argument slot. For example, the A for *hit* can be a noun denoting a person, or any of several types of animal, or a falling branch or rock, or an idea. The A for *think* is generally a human, with variable extension to higher animals and machines. A predicate including *hit* may involve a peripheral argument referring to an instrument, for example *with a stick*, as in:

(6)	A CORE	TRANSITIVE	O CORE	INSTRUMENTAL
	ARGUMENT	PREDICATE	ARGUMENT	PERIPHERAL ARGUMENT
	John	hit	the vase	(with a stick)

The three semantic roles in this transitive clause can in fact be mapped onto syntactic functions in an alternative way:

(7)	A CORE	TRANSITIVE	O CORE	LOCATIONAL
	ARGUMENT	PREDICATE	ARGUMENT	PERIPHERAL ARGUMENT
	John	hit	a stick	on the vase

That is, we can have:

SEMANTIC ROLES	MAPPED ONTO ARGUMENTS	
	IN SENTENCE (6)	IN SENTENCE (7)
Agent	A core argument	A core argument
Target	O core argument	locational argument
Thing manipulated ('Manip')	(instrumental peripheral argument)	O core argument

Sentence (6) shows the standard syntactic frame for *hit*—only A and O are obligatory, the instrumental argument being an optional extra. Sentence (7) is a rather marked variant pattern, with the Thing manipulated in O function; here the Target is coded through a locational argument and this must be included in the clause.

In some languages (not including English) there is an extended intransitive clause type, parallel to extended transitive. For both the extended clause types, the additional core argument (the 'extension') is labelled 'E'. The extended clause types can be illustrated for Tongan, belonging to the Polynesian branch of the Austronesian family (here ART indicates Article):

(8) intransitive na'e 'alu ['a e fefiné]_S
 PAST go ABSOLUTIVE ART woman
 The woman (S) went

(9) extended na'e sio ['a e fefiné]_S [ki he
 intransitive PAST see ABSOLUTIVE ART woman DATIVE ART
 tangatá]_E
 man
 The woman (S) saw the man (E)

(10) transitive na'e taa'i ['a e tangatá]_O ['e
 PAST hit ABSOLUTIVE ART man ERGATIVE
 he fefiné]_A
 ART woman
 The woman (A) hit the man (O)

(11) extended na'e 'oange ['a e tohi]_O ['e he
 transitive PAST give ABSOLUTIVE ART book ERGATIVE ART
 fefiné]_A [ki he tangatá]_E
 woman DATIVE ART MAN
 The woman (A) gave a book (O) to the man (E)

Tongan has an absolutive-ergative case system: S and O functions are shown by absolutive case (marked by particle 'a) and A function by ergative case

(particle 'e); note that ' indicates a glottal stop. Dative is shown by *ki*. Note that the noun phrases may occur in any order after the predicate, their functions being shown by initial case particles.

Argument profiles for the four clause types are (with case assignment in Tongan):

CLAUSE TYPE/PREDICATE	CORE ARGUMENTS		
intransitive	S (absolute)		
extended intransitive	S (absolute)	E (dative)	
transitive	A (ergative)	O (absolute)	
extended transitive	A (ergative)	O (absolute)	E (dative)

Only a few verbs appear in the extended intransitive frame; they include 'see', 'hear', 'love', and 'hate'. An extended intransitive selects two core arguments, like a transitive verb. But whereas transitive verbs tend to refer to some action which affects the referent of the O argument, an extended intransitive most often describes something which does not affect the second argument ('see' and 'like'). Trumai, a language isolate spoken on the Upper Xingu River in Brazil, also has a set of extended intransitive verbs, which includes 'see', 'hear', and 'smell'.

There is syntactic support for a sentence like (9) being analysed as a type of intransitive rather than as a variety of transitive. The NP '*a e fefiné*' in (9) functions like an S argument—differently from an A argument—in clause linking constructions; these are illustrated in Dixon (1994: 176).

English has a number of kinds of extended transitive constructions: (i) featuring verbs like *give*, *tell*, and *show* which in most circumstances require three arguments to be stated; (ii) the marked construction for verbs like *hit*, illustrated in (7), where again three arguments must be stated; and (iii) verbs *put* and (one sense of) *place*, which require statement of a locational argument.

Most clauses in every language involve either an intransitive or a transitive predicate (with possible subtypes extended intransitive and extended transitive). But there is generally a further clause type, which may not occur as frequently in the language as intransitive and transitive types but is nevertheless a significant syntactic type. This is a copula construction:

CLAUSE TYPE	PREDICATE	CORE ARGUMENTS
copula	copula verb	CS (copula subject) and CC (copula complement)

Transitive and intransitive verbs have referential meaning; one can identify the referent of 'go' or 'hit' or 'think' or 'give'. In contrast, a copula verb has a relational meaning, indicating a relation between CS and CC. The kinds of relation include (with illustrations from English):

RELATION

Identity	[My son] _{CS} is [an engineer] _{CC}
Attribution	[Our bosses] _{CS} are [very happy] _{CC}
Location	[The dog] _{CS} is [in the garden] _{CC}

The predicate of a copula clause is a copula verb. In terms of the old Greek-logic division of a sentence into two parts, subject and predicate, copula-plus-copula-complement would be predicate, sometimes called a ‘nominal predicate’. This is an unhelpful analysis, and should be avoided. The copula complement is an argument of the clause, just like S, A, O, E, and CS. In many languages, the copula subject (CS) has similar properties to an intransitive subject (S). But not in all. In Ainu, CS has the same realization as transitive subject (A), different from that of S. The copula complement (CC) generally shows different properties from other clausal arguments; for example, there are examples of S, A, O, E, and CS being partly realized by bound pronouns, but I know of no instance of this applying to CC.

In English, the two arguments, CS and CC, must be stated for a copula clause. However, other languages have a type of copula construction with just CS, and no CC; this indicates the ‘existence’ of the CS. For example, in Latin one can say *deus est* (lit ‘god is’), meaning ‘there is a god’.

There are languages in which the copula can be omitted in certain circumstances. And others which have no copula at all, just verbless clauses consisting of two NPs—these are verbless clause subject (VCS) argument and verbless clause complement (VCC) argument. They have similar properties to CS and CC in languages with a copula. There is discussion of copula and verbless clauses in Chapter 14, including justification for the analysis presented here.

Peripheral arguments may generally be included in all types of transitive and intransitive clauses (though less readily in copula or verbless clauses). They include spatial and temporal specifications (‘on the roof’, ‘at night’) and also descriptions of implements (‘with a hoe’), benefaction (‘on behalf of the teacher’), and manner (‘with alacrity’).

The distinction between core and peripheral arguments is never a hard and fast one. In English, for example, the verb *put* requires specification of location. One can say *He put the cup on the table/down/here/outside* but not just **He put the cup*. A, S, O, CS or VCS, and CC or VCC are always core arguments. There may be others in specific languages; for example, E in extended intransitives and extended transitives for languages which have one or both of these clause types. And in English a locational argument is obligatory just for *put* (and a couple of other verbs); it is a rather special variety of E argument.

(There has recently arisen a habit of reserving the label ‘argument’ just for core arguments and calling peripheral arguments ‘adjuncts’. This suggests a

definite division, which is not upheld by detailed description of clause structures across a range of languages. The term ‘adjunct’ is best avoided.)

3.3 The two major word classes, Noun and Verb

It has been suggested that some languages lack a distinction between Noun and Verb, there being instead a single undifferentiated class of lexical roots. This claim is discussed in Chapter 11 and shown to be without foundation. Every language that has been studied thus far shows a large Noun class and a separate Verb class.

In §1.8 it was emphasized that word classes must be recognized for each language on grammatical criteria internal to that language. In Latin, a class of words which inflect for case and number is called the Noun class, and in English this label is applied to a class of words which may be preceded by an article and need not be followed in the clause by another word. (Latin has no articles and rather free word order; English has no inflection for case, and number marking only on count nouns.) The Latin class and the English class are accorded the same name, ‘Noun’, since they have similar—although not identical—semantic content, and similar syntactic function. In each language, a noun is typically head of a phrase filling a predicate argument slot.

(a) **Noun.** Each language has, at the least, several thousand items in its Noun class. There are terms referring to many types of fauna and flora, many kinds of geographical, celestial, and meteorological features, all kinds of implements and machines, parts of the human body and of other things, kinship categories, and general terms for professional and social roles (such as ‘doctor’, ‘slave’, and ‘president’). These are all ‘common nouns’, each referring to a class of objects.

There are also ‘proper nouns’—names for people, places, hills, rivers, and so on—each a unique name for its referent. Two places or two people may have the same name but this is a coincidence, and unique identification is provided by adding further information; for example, Paris, France, as opposed to Paris, Texas, and the linguist Paul Newman in contrast to the film actor Paul Newman.

A proper noun may fill an argument slot in a clause, like a common noun, but it almost always has more limited morphological and syntactic properties. For example, a common noun, but not a proper noun, may inflect for number. There are likely to be more possibilities for modification of a common than of a proper noun within a noun phrase (see §3.4).

It was mentioned in §1.11 that a number of ‘semantic types’ can be recognized for each lexical class. For Noun, these include HUMAN, NON-HUMAN ANIMATE, FLORA, ARTEFACTS. Each type has its own potential for occurring in

specific argument slots with particular semantic types of verbs. The discussion in §1.3 described how in some languages the semantic type *KIN TERMS* has its own marking for possession, and in other languages the type (body and other) *PARTS* ('foot', 'eye', 'leaf', etc.) is treated in a special way. A *PART NOUN* may require a possessor to be stated, typically by a pronominal affix; in such a language one cannot say just 'foot' but only 'my foot', 'your foot', 'her/his foot', etc. See §3.4, §8.3.1, and Chapter 16.

Many languages divide nouns into a number of gender or noun classes—see the discussion in §1.9 of the four noun classes in Dyirbal, roughly labelled 'masculine', 'feminine', 'edible plants', and 'neuter'. In other languages nouns may occur with one or more classifiers, and this again relates to semantic types. See §3.16.

(b) *Verb*. All languages have a Verb class, generally with at least several hundred members. However, as exemplified for the Australian language Yawuru in §1.11, there are languages with a limited number of verb roots, sometimes only a few dozen. These combine with a multiplicity of 'coverbs', creating a rich set of compound verbs. (Often, the coverbs have invariable form, all morphological processes—for tense, aspect, and so on—applying to the verb root.)

Clause types were described in the last section. Leaving aside copula clauses (whose predicate is one of a small set of copula verbs), there are two recurrent clause types, transitive and intransitive. Verbs can be classified according to the clause type they may occur in:

- (a) *Intransitive verbs*, which may only occur in the predicate of an intransitive clause; for example, *snore* in English.
- (b) *Transitive verbs*, which may only occur in the predicate of a transitive clause; for example, *hit* in English.

In some languages, all verbs are either strictly intransitive or strictly transitive. But in others there are ambitransitive (or labile) verbs, which may be used in an intransitive or in a transitive clause. These are of two varieties:

- (c) *Ambitransitives of type S = A*. An English example is *knit*, as in *She_S knits* and *She_A knits socks_O*.
- (d) *Ambitransitives of type S = O*. An English example is *melt*, as in [*The butter*]_S *melted* and *She_A melted [the butter]_O*.

The importance of stating whether an ambitransitive verb is of type *S = A* or of type *S = O* was stressed under (c) in §2.5.

When there are additional clause types—such as extended intransitive and/or extended transitive—we may encounter verbs which manifest other

TABLE 3.1. Sample semantic types of the Verb class in English, and their roles

<i>Semantic type</i>	<i>Roles</i>			
AFFECT (for example, <i>hit, burn</i>)	Agent	Target	Manip	
GIVING (for example, <i>give, lend</i>)	Donor	Gift	Recipient	
SPEAKING (for example, <i>speak, tell</i>)	Speaker	Addressee	Message	Medium
THINKING (for example, <i>remember</i>)	Cogitator	Thought		
ATTENTION (for example, <i>see, hear</i>)	Perceiver	Impression		
LIKING (for example, <i>like, love, hate</i>)	Experiencer	Stimulus		

sorts of multiple-clause-type possibilities. For example, being able to occur in either a transitive or an extended intransitive clause, with A = S and O = E, the two possibilities being associated with a meaning difference.

The Verb class is not homogeneous, in terms of the meanings of its members. As briefly mentioned in §1.9 and §1.11, a number of semantic types may usefully be recognized, each with a set of semantic roles describing the participants involved in the activity, state, or property. Dixon (2005a) has a list of about thirty semantic types associated with the Verb class in English, many of which recur in other languages. A sample of six of these is shown in Table 3.1.

The AFFECT verb *hit* was illustrated in (6–7) of §3.2; ‘Manip’ is here employed as label for ‘Thing manipulated’. Examples of the others are:

- (1) John_{DONOR} gave [a book]_{GIFT} [to Mary]_{RECIPIENT}
- (2) Mary_{SPEAKER} told [a story]_{MESSAGE} [to her brother]_{ADDRESSEE} ([in French]_{MEDIUM})
- (3) John_{COGITATOR} remembers [the earthquake]_{THOUGHT}
- (4) Mary_{PERCEIVER} saw [the eclipse]_{IMPRESSION}
- (5) John_{EXPERIENCER} likes ballet_{STIMULUS}

These are all transitive or extended transitive verbs. For each, one role must be related to argument A (transitive subject) in clause structure and another to argument O (transitive object). This provides a further example of how a multitude of varied semantic distinctions—occurring across many semantic types of verbs—are mapped onto a small number of syntactic contrasts, here the core syntactic relations, S, A, and O.

In English, it is the leftmost of the roles in each row of Table 3.1 which is placed in A function. We need to enquire what the principle is underlying this association between semantic roles and syntactic functions. What is there in common between a person wielding an implement, for AFFECT, someone

who transfers their ownership of something, in GIVING, a person whose mouth produces an utterance, for SPEAKING, someone whose mind focuses on something, for THINKING, a person who—either purposefully or involuntarily—receives a sense impression, for ATTENTION, and someone who experiences a certain internal feeling about something, in LIKING?

The principle appears to be as follows: that role which may initiate or control the activity is placed in A function. This is plainly the Agent for an ACTION verb, the Donor for GIVING, the Speaker for SPEAKING, the Cogitator for THINKING, and the Perceiver for ATTENTION. For verbs referring to mental feeling, either Experiencer or Stimulus could hold major responsibility for the state of mind. Indeed, English has two semantic types involving these roles. Verbs of the LIKING type focus on the Experiencer as relating to the success of this mental state. If one hears *Mary*_{EXPERIENCER:A} *likes* [*John's behaviour*]_{STIMULUS:O}, Mary must be paying attention to what John is doing, whereas John may not be aware that he is being observed. In contrast, verbs of the ANNOYING type (including 'offend', 'anger', and 'please') have Stimulus in A role and Experiencer as O. If one hears [*John's behaviour*]_{STIMULUS:A} *pleases* *Mary*_{EXPERIENCER:O}, then the likelihood is that John was making an effort to impress.

If the verbs in a semantic type have just two roles, then that which is not placed in A function becomes O. Thus for THINKING, the Thought role is in O function, and similarly Impression for ATTENTION and Stimulus for LIKING. This much is straightforward. If there are more than two roles, then that which is most saliently affected by the activity, or is most specific, is mapped onto O function.

Verbs in English with more than two roles typically have alternative syntactic frames in which each of the non-A roles can feature as O. Recall (6–7), with *hit*. Typically, the Target is affected by the action and is in O slot, with the Manip as an optional peripheral constituent. Thus, *John*_{AGENT:A} *hit* [*the vase*]_{TARGET:O} ([*with a stick*]_{MANIP}) implies that the action of the stick probably broke or cracked the vase. The alternative syntactic frame is with the Manip as O and Target coded through a locational NP, which must be stated. If one hears *John*_{AGENT:A} *hit* [*a stick*]_{MANIP:O} [*on the vase*]_{TARGET}, it is likely that the stick broke when brought into contact with the vase (which may have been large and made of iron).

Similarly for GIVING verbs in English. The Donor is always A but either Gift or Recipient can be in O slot, as in:

- (6) *John*_{DONOR:A} *gave* [*all his goods*]_{GIFT:O} [*to charity*]_{RECIPIENT:E}
- (7) *John*_{DONOR:A} *gave* [*his favourite student*]_{RECIPIENT:O} [*some books*]_{GIFT:E}

In (6) the Gift has specific reference (*all his goods*) and the Recipient (*charity*) is vague, so Gift is coded as O. In contrast, (7) has a non-specific gift (*some books*) whereas the Recipient is specific (*his favourite student*), and so this is coded as O. (Note that whereas *hit* is a transitive verb and can occur with just two core arguments, *give* is extended transitive, and requires all three roles to be stated, as in (6–7).)

The SPEAKING type in English includes many verbs, with different semantic orientations. For *report*, the Message must be on O function, as in:

- (8) John_{SPEAKER:A} reported [the accident]_{MESSAGE:O} ([to the police]_{ADDRESSEE})

But for *inform*, the Addressee is focussed upon, and is mapped onto O, as in:

- (9) John_{SPEAKER:A} informed [the police]_{ADDRESSEE:O} ([of the accident]_{MESSAGE})

With most verbs from this semantic type, the Medium is an optional extra constituent, as in (2). But it can be placed in O function:

- (10) Mary_{SPEAKER:A} spoke [German]_{MEDIUM:O} ([to the lodger]_{ADDRESSEE})

Here the focus is on the nature of the Medium; the Addressee can be included, as a peripheral argument (but a Message cannot be fitted into this clause).

Secondary concepts were referred to in §1.11; they include ‘begin’, ‘try’, ‘want’, and ‘make’ (in the causative sense). In some languages these are realized through morphological processes, typically as affixes to verbs. In others, including English, they are coded as lexemes, which take a complement clause (see §3.10, Chapter 18, and Dixon 2006a).

3.4 Phrases

(a) **Noun phrase (NP)**. A core or peripheral argument slot in clause structure is filled by a noun phrase. This can consist just of a noun, or have a noun as head, accompanied by a number of modifiers. An NP has the same basic semantic properties and functional possibilities as its head. (An alternative head may be a pronoun or a demonstrative; see §3.7.)

Modifiers of a common noun as head of an NP typically include:

- (i) One or more adjectives (which may include participles derived from verbs, as in *a thinking man*); see §3.6 and Chapter 12. In some languages the adjective may be in comparative or superlative form; see §3.23.

- (ii) A cardinal or ordinal number or a quantifier (in some languages these can be analysed as a subclass of Adjective); for example, in English, *two boys, a third banana, the last runner, every lion*.
- (iii) One or more nouns. The set of nouns which may function as modifier (in addition to their function as head) is generally pretty limited. It may include specification of:
- sex, as in *man child* (as an alternative to *male child*, involving adjective *male*);
 - composition, as in *metal door, stone arch*;
 - purpose, as in *rabbit food, silver polish*.

More complex modifiers of an NP head include:

- (iv) A possessive phrase. This is effectively an NP embedded within an NP; for example [*the best potato peeler from Ireland*]'s POSSESSIVE.NP *new*ADJECTIVE *shoes*HEAD. Something which is alienably possessed is always head of its NP (as is *shoes* in the example just given). When whole–part possession is dealt with by the grammar in the same way as alienable possession, then the part will be NP head, as in *John's nose*HEAD. But in languages which employ a different construction type for whole–part (inalienable) possession, it is sometimes the noun referring to the whole which is head (thus *John*HEAD *nose*). This is discussed in Chapter 16.
- (v) A relative clause, as in:

- (1) I_A saw [the man [who stole the mink]RELATIVE.CLAUSE]O
- (2) I_A saw [the place [where John lives]RELATIVE.CLAUSE]O
- (3) I_A know [the time [when you got home last night]RELATIVE.CLAUSE]O

A relative clause must have an argument which is coreferential with the head of the NP it is modifying. This argument can be in A function in the relative clause, as in (1), or in S, O, E, CS, or CC function. Or it can be in a peripheral function, as in (2–3). Relative clauses are discussed in Chapter 17.

- (vi) An NP, typically referring to the spatial or temporal location of the head. In English, these are marked by a preposition, as in:
- (4) I_A like [the statue [in the square]SPATIAL.NP]O
 - (5) I_A saw [the presentation [at four o'clock]TEMPORAL.NP]O

In English, an NP-plus-preposition modifier may also describe an intended beneficiary, as in:

- (6) I_A saw [the present [for Mary]_{BENEFICIARY.NP}]_O

A demonstrative is a grammatical word which can have pointing (or deictic) reference. In some languages, a demonstrative can function as head of an NP (I_A want *that*_O, with pointing) or as a modifier (I_A want [*that red dress*]_O, with pointing). In other languages, the most appropriate analysis is to say that a demonstrative makes up a complete NP, and can be placed in apposition to an NP with a common noun as head, something like I_A like [[*that (one)*][*(the) red dress*]]_O (with pointing). Here the two simple NPs, *that* and *red dress*, together make up a complex NP which functions as O argument for *like*. Demonstratives are discussed in §15.2.

A number of languages—English being one of them—include in their NPs an article (these have generally developed out of demonstratives). This may be obligatory in certain contexts and indicates definiteness.

A proper noun as the head of an NP is likely to have far fewer—if any—possibilities for modification, when compared to a common noun as NP head. It may be the case that it does not accept an adjective, nor a relative clause, nor a demonstrative.

A major puzzle in understanding the flow of language—for both a learner and a native speaker—lies in identifying an unfamiliar-sounding word as a previously unencountered proper name, rather than trying to relate it to some common noun. A few languages ease this dilemma by including a special marking on a proper noun. For instance, in Fijian, an NP with a common noun as head commences with grammatical particle *na* or *a*, while one which has as its head a proper noun or pronoun—forms which have unique reference—commences with particle *o*. In the tradition of Fijian grammar writing these are referred to as ‘common article’ *na* or *a* and ‘proper article’ *o*, the label ‘article’ being used in a slightly different way from its employment in the grammars of European languages.

An argument slot can be filled by a complex NP, which consists of a coordination of two simple NPs; for example, in English [[*Two small dogs*] and [*that fat cat which lives across the road*]] were having a fight.

(b) **Verb phrase (VP).** A number of grammatical categories are generally associated with the predicate, whose head is typically a verb; these include tense, aspect, modality, evidentiality, direction markers such as ‘coming’ and ‘going’, and secondary concepts such as ‘begin’, ‘try’, ‘want’, and causative ‘make’.

In a language where the verb has a rich morphological structure, these categories may be coded by affixation, or by other morphological processes. In a language of more analytic structure, they may be realized by words modifying a verb (or other item) which functions as head of a ‘verb phrase’ (VP) that fills the predicate slot.

There is much more variation between languages concerning the structure of a VP than there is for NPs. The great majority of languages allow a common noun to be modified by adjectives, numbers, quantifiers, possessive phrases, relative clauses, and so on. Languages with a complex, synthetic verb structure require little in the way of syntactic modification to a verb. English, which has relatively meagre morphology, has a set of modal verbs (*can, must, may, will*, etc.) plus markers of ‘imperfective aspect’ (*be...-ing*) and ‘previous aspect’ (*have...-en*) within its VP. (See Dixon 2005a: 172–7, 209–29.) For example, one can say *He_A [could have been planning] [the heist]_O [for months]*.

Perhaps a third of the world’s languages have what is called a ‘serial verb construction’ (SVC), where the predicate includes two (or more) verbs, each of which could make up a predicate on its own. Although an SVC consists of more than one verb, it is conceived of as describing a single action; for example, in Igbo from West Africa, verbs *-tì*- ‘hit’ and *-wà*- ‘split open’ combine to mean ‘shatter’ in an SVC. There must be a single subject applying to the whole SVC (and there will be other grammatical properties—such as the marking of negation—which differ a little from language to language).

The most common variety of SVC is ‘asymmetrical’ where the major member (which is often, on grammatical criteria, the head) can be virtually any verb, while the minor member comes from one of a number of sets of verbs which effectively provide modification of the head. Typically, these may indicate direction (e.g. ‘come’), aspectual-type meanings (progressive or habitual), or they may code secondary concepts such as ‘begin’ and ‘try’.

The word ‘adverb’ is used in a variety of ways. Adverbs are often derived from adjectives, which modify nouns. It would be reasonable to expect that an adverb should modify a verb. It may be that in some languages adverbs can be appropriately treated as being part of a VP which fills the predicate slot. However, in other languages this is not the preferred analysis. English has sentential adverbs (for example, *only, certainly, possibly*) which apply to a complete clause or sentence, and also manner adverbs (including *exactly, slightly, easily, well*) which relate to the predicate or predicate-plus-other-constituents but are not usefully regarded as part of the predicate constituent. (See Dixon 2005a: 375–445.)

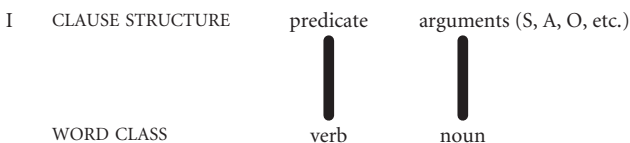
(Those linguists who indulge in binary division of a clause—which they typically call ‘sentence’—often say that a sentence consists of an NP (meaning subject) and a ‘verb phrase’ (meaning ‘predicate’ in the logician’s sense). Their

‘verb phrase’ includes object NP and perhaps other constituents as well. This is a quite different use of ‘verb phrase’ from that followed here.)

3.5 Clause structure and word classes

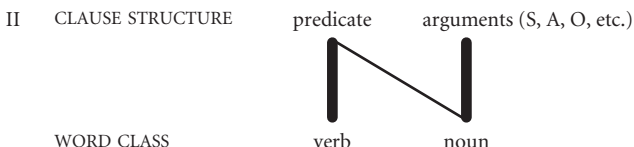
Each slot in clause structure is filled by an appropriate word—or rather, by a phrase, with noun or verb as obligatory head plus optional modifiers. It is useful to focus on the head word of a phrase, and investigate how nouns and verbs relate to elements in clause structure.

The typical association is of verb with predicate and of noun with arguments (whether core or peripheral). That is:



In some languages this is the only association. In Latin—and also in Dyirbal—only a noun (plus possible modifiers) can relate to an argument, and an argument may only be realized by a noun. Similarly for verb and predicate. (There are processes for deriving a verb from a noun and vice versa; for example from Latin *instruō* ‘construct’ can be formed *instrūmentum* ‘equipment’, but this is the creation of a new noun, not use of a verb itself as head of an NP.)

In some languages, noun or verb (or both) may have a secondary function, beyond the primary functions shown in Scheme I. But in every case the primary functions are most important and more frequent. A noun may also function as head of an intransitive predicate:



Hajaú (1963: 67) states that in Nenets (spoken in Siberia, Samoyed branch of Uralic) ‘the substantive [noun] can also be the predicate of the sentence, and in this function it can take verbal person suffixes’. For example:

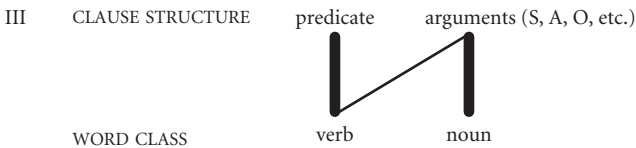
- (7) mañ hāsawa-dm ‘I am a man’
 (8) pydar hāsawa-n ‘You are a man’

Mandarin Chinese shows a different pattern—a noun may only function as argument, whereas a verb may be either predicate or argument. Compare (*le* is here a sentence-final particle ‘currently relevant state’):

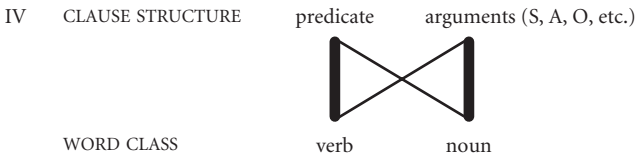
- (9) Daifu lai le
 doctor come PARTICLE
noun verb
 SUBJECT PREDICATE
 The doctor came

- (10) Lai tui le
 come be.right PARTICLE
verb verb
 SUBJECT PREDICATE
 Coming was right (that is, It was right to come)

We now have:



Nootka, spoken on Victoria Island in British Columbia, Canada, is said to have a further pattern of associations:



These are illustrated in:

- (11) [ʔiḥ-ma']_{INTRANSITIVE PREDICATE} [qoʔas-ʔi']_S
 be.large-3sg.INDICATIVE man-ARTICLE
 The man is large
- (12) [qoʔas-ma]_{INTRANSITIVE PREDICATE} [ʔiḥ-ʔi']_S
 man-3sg.INDICATIVE be.large-ARTICLE
 The large one is a man

In (11) verb 'be large' and noun 'man' are in their prototypical functions, as predicate and argument respectively. But a noun can alternatively function as predicate and a verb as its argument, which is illustrated in (12).

It is this kind of apparent interchangeability of noun and verb in argument and predicate slots which has led some linguists to suggest that Nootka might lack a distinction between Noun and Verb. As is explained in Chapter 11, there are other kinds of criteria which distinguish the two word classes. And

it should be noted that noun is most often used in argument slot, only very occasionally as predicate; similarly, a verb is generally employed as predicate, much less often as an argument (as shown by the thick and thin lines in the diagrams).

This variation of associations between clause slots and word classes, in different languages (and there are doubtless further possibilities, in other languages), emphasizes the importance of distinguishing clause structure and word classes. (It is definitely not satisfactory to simply state, say, that a clause consists of NP, with noun as head, and VP, with verb as head, as has sometimes been done.)

Superficial examination of English suggests that some nouns have secondary function as predicate head—for example, *stone* in *The Romans stoned the Christians*—and that some verbs have secondary function as head of an NP—for example, *walk* in *That long walk tired me out*. However, only *some* nouns may be used in a predicate, and only *some* verbs as NP head. One cannot predict exactly which items from these word classes will have an apparent secondary function, nor what their meaning will be in that function. The optimum analysis is to say that we have in English a number of ‘zero derivations’; that is, word-class-changing derivations which have zero marking. They are paralleled by derivations with non-zero marking. Compare noun *market* and verb *market* with noun *hospital* and verb *hospital-ize*. Also compare verb *witness* and noun *witness* with verb *observe* and noun *observ-er*.

All of these issues are further discussed, and exemplified, in Chapter 11.

3.6 Adjectives

It has been suggested for a few languages that there is no distinction between Noun and Verb. And for a larger number it has been said that there is no Adjective class. Indeed, some grammars include no mention at all of ‘adjective’. Close examination of the example sentences and vocabulary may reveal that concepts coded by adjectives in other languages are in this language all realized as verbs, or all as nouns. In point of fact, I know of no language which has been thoroughly and insightfully described for which an Adjective class cannot be recognized.

The function and properties of an Adjective class vary widely from language to language. There are four basic types:

- (a) *Adjectives have similar properties to verbs*. That is, an adjective can occur as head of a phrase filling predicate slot in clause structure, just as an intransitive verb may. For each such language, some criteria can be discerned to distinguish Adjective and Verb. The actual criterial properties vary from language to language; they typically include slightly different

possibilities between verb and adjective for being modified when functioning as predicate, for functioning as modifier within an NP, for occurring in comparative constructions, and for forming adverbs.

- (b) *Adjectives have similar properties to nouns.* That is, an adjective may occur as modifier in an NP and it may also make up a complete NP (a decision then has to be made between saying that the adjective is NP head, and saying that a noun head has been ellipsed). An adjective may undergo the same morphological processes as a noun; for example, taking number and/or case marking. However, there always are some criteria which enable the linguist to distinguish two word classes. They may relate to the internal structure of NPs (if an adjective is head there may be fewer possibilities for modification than if a noun is head) or to the fact that only an adjective may occur in a comparative construction, or may form adverbs.
- (c) *Adjectives share grammatical properties with both verbs and nouns.* An adjective can function similarly to an intransitive verb is being head of a predicate, and it may inflect like a noun when occurring in an NP.
- (d) *Adjectives have grammatical properties different from those of verbs and of nouns.* English is of this type—an adjective may neither function as predicate head nor as NP head; it does not share any inflection with verb or with noun. Unlike nouns and verbs, an adjective occurs in a comparative construction (marked by either *-er* or *more*), and adverbs may be formed from many—but not all—adjectives.

There are some languages with two subclasses of adjective, one with similar morphological and syntactic possibilities to verbs, and the other to nouns. They are linked together through some common properties, such as being the only items to occur in a comparative construction.

There are two basic semantic tasks for an adjective to perform:

- (I) Make a statement that something has a certain property. In languages of type (a), this is achieved through the adjective functioning as intransitive predicate (literally, ‘The man happies’). In languages of types (b) and (d), the adjective will make up a copula complement or verbless clause complement argument (‘The man (is) happy’). Languages of type (c) are likely to show both possibilities.
- (II) Provide a specification that helps focus on the referent of the head noun in an NP. This is shown by an adjective acting as noun modifier in a language of types (b), (c), and (d), as in ‘The happy man’. Languages of type (a) differ in how adjectives fulfil this task; in some, an adjective must occur within a relative clause construction (literally, ‘The man who happies’).

Every language has a large open Noun class and almost all have a Verb class of significant size. In some languages the Adjective class is fairly large and open (that is, new words may be added, either through language-internal derivation or as loans). However, a fair number of languages have a small, closed adjective class, with anything from three or four to a few score members.

There is a semantic basis to the make-up of a small Adjective class. As mentioned in §1.11, if it has only a dozen or so members, these are likely to belong to four semantic types:

DIMENSION—‘big’, ‘little’, ‘long’, ‘short’

AGE—‘old’, ‘young’, ‘new’

COLOUR—‘black’, ‘white’, ‘red’

VALUE—‘good’, ‘bad’

An Adjective class with twenty or so members is likely to include some from the PHYSICAL PROPERTY type, such as ‘raw’, ‘hard’, ‘heavy’, ‘wet’. Larger classes will include some HUMAN PROPENSITY items, such as ‘clever’, ‘greedy’, ‘rude’, ‘jealous’.

Where have all the other adjectival concepts gone, in a language with a small class? Study of a selection of reliable grammars and dictionaries enables an inductive generalization to be made. PHYSICAL PROPERTY ideas are most likely to be coded as verbs (one says, literally ‘It heavies’) and HUMAN PROPENSITY concepts either as nouns (saying, literally, ‘She has cleverness’) or as verbs.

There is further discussion of adjective classes in Chapter 12.

3.7 Pronouns, demonstratives, and interrogatives

A speech act involves participants (speaker and addressee) in a place, at a time. All languages have sets of ‘shifters’, whose reference shifts when the role of the participants change, when the place changes, or when the time changes:

- (a) Participant shifters—pronouns. As turn-taking progresses in a conversation, and I am succeeded by you, so ‘you’ becomes ‘I’.
- (b) Spatial shifters—demonstratives. After this man moves from here, he becomes ‘that man there’.
- (c) Temporal shifters—words referring to time intervals with respect to the present. What is ‘today’ today becomes ‘yesterday’ tomorrow. See (c) in §3.8, and the discussion of tense in §3.15.

In traditional grammar, the word ‘pronoun’ is used with a wide range, to cover what are here called pronouns (otherwise personal and possessive pronouns), demonstratives, interrogatives, reflexive and reciprocal pronouns (see

§3.22), relative pronouns (Chapter 17), and indefinite terms such as ‘someone’ and ‘anything’ (see below and Chapter 15).

(a) **Pronouns.** All languages have 1st and 2nd person pronouns, referring to speaker and addressee respectively. There is almost always a number distinction in a pronoun system, at the least singular (sg) and plural (pl); for sample larger systems, see Table 1.1 in §1.4. However, it is not the case that every language has distinct pronouns for 1sg, 2sg, 1pl and 2pl. In Chipewyan, an Algonquian language of Canada, there is one plural form covering both persons:

1sg	si	}	1pl/2pl	nuhni
2sg	nën			

And standard English has one 2nd person form covering both singular and plural:

1sg	I	1pl	we
}			
2sg/2pl		you	

(However, in many spoken dialects a new 2pl form has been innovated. The forms vary; they include *yous* and *y'all*.)

Some languages have in non-singular numbers a distinction of 1st person between inclusive (including addressee) and exclusive (excluding addressee), as in the Australian language Kayardild:

	<i>singular</i>		<i>dual</i>	<i>plural</i>
1	ɲada	{	ɲaku-rra	ɲaku-lda
			ɲa-rra	ɲa-lda
2	nɲjka		ki-rra	ki-lda

Note that it would be inappropriate to place the 1sg pronoun in either the inclusive or the exclusive row; the inclusive/exclusive distinction does not apply in the singular.

Some languages have ‘me and you’ as a minimal term, on a par with 1sg and 2sg. For example in Hdi, a Chadic language from Cameroon, we find:

	<i>minimal</i>	<i>augmented</i>
1	íí	áɲní
1 + 2	úú	ámú
2	kághá	kághúní

Forms in the minimal column refer to one person (the 1 and 2 rows) or to two people (the 1+2 row). The augmented column indicates one or more persons in addition to those in the minimal set.

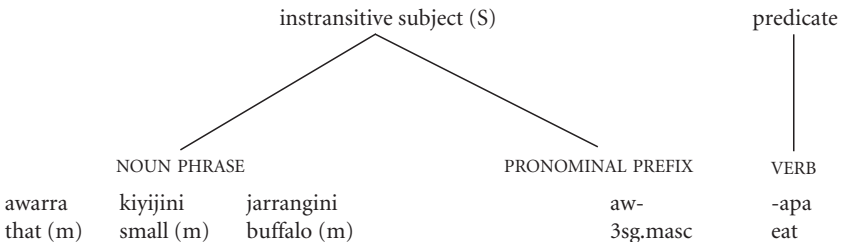
Many languages have a 3rd person pronoun in the paradigm, referring to someone other than speaker or addressee. In some cases these have the same morphological make-up and inflectional categories as 1st and 2nd person forms; in others they differ markedly (see §3.9). There are languages with no 3rd person pronoun as such; the function which this would fulfil may be dealt with in part by demonstratives.

Gender or noun class may be marked on some pronouns. Most typically on 3sg (as with English *he* and *she*) but sometimes also on 2nd and/or 1st person, in sg and/or in other numbers.

A pronoun may be used as head of an NP, in place of a noun. Pronouns are rather like proper nouns, in having unique reference; there are generally rather limited possibilities for modification of a pronoun, as of a proper noun.

Some languages have a single array of pronouns. Others have two sets, generally called ‘free’ and ‘bound’. A free pronoun will be a single (phonological and grammatical) word and can function as head of an NP (replaceable by a noun). A bound pronoun is likely to be a clitic or affix, typically attached to the verb (a common alternative is for it to be added to the end of the first constituent of the clause). Bound pronouns are often obligatory; a clause may consist just of verb plus bound pronoun, something like ‘go-I’ or ‘go-you’ or ‘go-she’. This was illustrated under (c) in §1.10 for the Australian language Tiwi. Pronominal information, as expressed in a bound pronoun, may be supplemented by the appropriate free pronoun as NP—something like ‘I_{NP} I-go_{VERB}’ places emphasis on ‘I’ as the person going. And a third person argument may be shown by a full NP in addition to the bound pronoun, as in ‘[That small buffalo]_{NP} he-eats_{VERB}’.

As explained in §1.10, what we have here is a core argument with discontinuous realization. Repeating (21) from §1.10, ‘That small buffalo eats’ in Tiwi can be shown as:



Under this approach, it is neither necessary nor appropriate to enquire whether the main marker of intransitive subject is *aw-* or *awarra kiyijini jarrangini*. The underlying S argument in this clause simply has double realization in surface structure—by a bound pronoun which (in terms of the

morphological structure of the language) is an obligatory component of the verbal word, and by a full NP.

The status of bound pronouns varies between languages. Most often a bound pronoun is obligatory for the S argument in an intransitive and for A and O in a transitive clause. Some languages (including Latin) only have bound pronouns for S and A, others only for S and O. In a few instances there may be three bound pronouns, including reference to a third core argument (E). When bound pronouns are obligatory, the free forms are used sparingly, typically for emphasis or to mark the introduction of a new participant into a discourse. And there are some languages where one may include either a free or a bound pronoun, but not both. Fuller discussion of all kinds of pronouns is in Chapter 15.

(b) **Demonstratives.** As mentioned in the discussion of noun phrase structure under (a) in §3.4, a demonstrative is a grammatical element which can be used—generally, accompanied by a gesture—to point to an object in the situation of discourse. Many languages have two nominal demonstratives, ‘this’ and ‘that’, which may modify a head noun in an NP or may make up a full NP. It is not uncommon for there to be a three-term contrast, either ‘near’, ‘mid-distance’, and ‘far’, or ‘near speaker’, ‘near addressee’, and ‘near neither’. And there can be further kinds of specification; for instance, a language spoken in hilly country may have demonstratives referring to distance ‘up’ and ‘down’, as well as ‘near’ and ‘far’.

A few languages have just one nominal demonstrative ‘this’. However, all languages appear to have at least two adverbial demonstratives, ‘here’ and ‘there’; and there may be larger systems, similar to nominal demonstratives. Some languages also have a set of verbal demonstratives, ‘do it like this’ (again accompanied by an appropriate gesture).

Besides their deictic (that is, pointing) use, demonstratives may introduce new information, and make anaphoric reference back in discourse (as in *Stalin was dictator of the USSR and that man had millions of innocent people killed*). See the discussion in §§15.2–3.

(c) **Interrogatives.** As mentioned under (b) in §3.2, a content interrogative clause will include a content interrogative word. Typically, the possibilities include ‘who’, ‘what’, ‘which’, ‘how many’, ‘why’, ‘how’, ‘where’, ‘when’, and sometimes also an interrogative verb ‘do what/how’.

In many languages, each of these items may be assigned to a word class; ‘what’ is classified as a noun, ‘who’ as either a pronoun or a noun (depending on the language), ‘which’ as an adjective, and so on. But, in addition, the content interrogatives form a natural class of their own, with certain shared properties. For example, Jarawara has a system of mood suffixes to a verb; the

content interrogative suffix *-ri* must be used if a clause includes *himata* ‘what’ or *hibaka* ‘who’ (which both belong to the noun class), *hika* ‘where’ (which is a spatial adverb), or *ee -na-* ‘what about, how many’ (which belongs to the verb class).

In quite a number of languages, one set of words combines interrogative and indefinite senses. In Jacaltec, a Mayan language of Guatemala, *maca* has meanings ‘who’ and ‘someone’, *tzet* is ‘what’ and ‘something’, while *b’ay* is ‘where’ and ‘somewhere’.

Indeed, an indefinite/interrogative form may carry both meanings at once. Dick Moses once gave me the following sentence in the Australian language Yidiñ:

wañju _A	walba _O	yanggi:n̄
INDEFINITE/INTERROGATIVE.ERGATIVE	rock.ABSOLUTIVE	split.PAST

He translated it as ‘Someone must have cut that rock—who did it?’ *Wañju* indicates indefiniteness ‘someone did it’ and at the same time enquires concerning the identity of that someone.

3.8 Syntactic specification of space and time

(a) Function

Space and time may be specified by peripheral arguments within a clause. These are prototypically phrases, which typically occur on the periphery of the clause. In English their most common position is clause-final—as in *Mary_A received [her prize]_O [in the town hall]_{PERIPHERAL}* and *Mary_A received [her prize]_O [on Friday afternoon]_{PERIPHERAL}*. Alternatively, they may be placed initially—[*In the town hall*]_{PERIPHERAL}, *Mary_A received [her prize]_O* and [*On Friday afternoon*]_{PERIPHERAL}, *Mary_A received [her prize]_O*.

As mentioned under (a) in §3.4, an NP may include—after the head—an embedded NP which specifies place or time; for example [*The photo [on the verandah]_{SPATIAL}S looks good*]. Since a peripheral argument generally occurs clause-finally and an O NP follows the verb in English, a sentence such as *John will examine the photo on the verandah* has two possible parsings:

- (1) John_A will examine [the photo [on the verandah]_{SPATIAL}]_O
- (2) John_A will examine [the photo]_O [on the verandah]_{PERIPHERAL}

In (1) *on the verandah* is a modifier within the O NP. The sentence could describe John examining a framed photo which is hanging on the wall of the verandah. In (2) *on the verandah* is a peripheral constituent of the clause. This sentence could describe John taking a photo out onto the verandah to examine it, since the light is better there.

Similar examples can be provided with NPs referring to time:

- (3) Mary_A will watch [the movie [at five o'clock]_{TEMPORAL}]_O
 (4) Mary_A will watch [the movie]_O [at five o'clock]_{PERIPHERAL}

Sentence (3) relates to a movie scheduled to be shown on TV at five o'clock; it talks about *the movie at five o'clock*. In contrast, (4) could be used when Mary has a DVD of the movie and can choose the time at which she watches it.

It is interesting to note that when the head of the O NP is pronoun *it*, there is no ambiguity. In *John will examine it on the verandah* and *Mary will watch it at five o'clock*, the phrases *on the verandah* and *at five o'clock* can only be peripheral constituents of the clause; these sentences are parsed like (2) and (4), not like (1) and (3). This is because a singular pronoun as NP head cannot take temporal or spatial modification.

The semantic types of MOTION and REST verbs expect—but do not require—a spatial argument, as in:

- (5) Mary_A planted [the rose tree]_O [in the garden]_{INNER.LOCATIVE}

This is sometimes called an 'inner locative' since it relates to the meaning of the verb.

Any concrete activity must take place somewhere, and it is always possible to show this by including an 'outer locative', as in:

- (6) John_A cut [his hand]_O [in the garden]_{OUTER.LOCATIVE}

There is here no association between the locative NP and the meaning of the verb.

Similarly, some verbs expect—but do not require—a temporal argument, as in:

- (7) Mary_A set [the alarm clock]_O [for five o'clock]_{INNER.TEMPORAL}

This indicates that Mary set the alarm to go off at five o'clock. A statement about setting an alarm clock is likely to be accompanied by specification of what time it was set for.

With a different sense of *set*, the only temporal argument possible is one of the outer variety:

- (8) Mary_A set [the table]_O [at five o'clock]_{OUTER.TEMPORAL}

A statement about setting a table does not carry any expectation concerning specification of time.

Note that the sentence

- (9) Mary_A set [the alarm clock]_O [at five o'clock]

is ambiguous. It could mean, like (7), that she set the alarm to go off at five o'clock (inner temporal) or that it was at five o'clock (outer temporal) that she set it to go off, at some time later than five o'clock.

One difference between outer and inner locatives and temporals in English is that the inner variety is restricted to occurrence at the end of its clause, while the outer type may be either final or initial.

(b) Nature

In English, both spatial and temporal settings may be shown, syntactically, in one of three ways:

	SPATIAL	TEMPORAL
BY A WORD	here, outside	tomorrow, soon
BY A PHRASE	in the garden, up a tree	in the afternoon, on Saturday
BY A CLAUSE	where he sat	after he came, when she left

Instead of *on the verandah* in (1–2) we could use *outside*, or *outside on the verandah*. And in place of *at five o'clock* in (3–4), either *tomorrow* or *at five o'clock tomorrow* could be employed. Similar possibilities—involving both words and phrases—apply for a peripheral argument in a clause, and for a modifier within an NP.

There is one set of temporal words—*yesterday*, *today*, and *tomorrow*—which have an additional role, as head of an NP. For example, *This is yesterday's bread*, and *Tomorrow should be a fine day*.

The head of an NP may be modified by a relative clause which can refer to space or to time; for example, *Mary_A saw [the place [where he sat]_{RELATIVE.CLAUSE}]_O* and *John_A knows [the time [when she came home]_{RELATIVE.CLAUSE}]_O*.

There are many types of temporal clauses which may link up with a main clause; for example *After she goes, I'll tell you* or *When he comes, I'll tell you*. On superficial examination, there appear to be clauses which provide spatial information within a main clause; for example *John will put the cushion where the old man sits*. However, this could be analysed as a reduced form of *John_A will put [the cushion]_O [in the place [where the old man sits]_{RELATIVE.CLAUSE}]_{PERIPHERAL}*, with *where the old man sits* being a relative clause to *the place* within the peripheral NP.

(c) Semantics

The following types of temporal and spatial specifications recur in many languages. We can illustrate with forms from English:

[i] *Temporal*

- Shifters—*yesterday, today, tomorrow, later, earlier*
- Definite time specification—*in the morning, at night, on Sunday, in the winter, in March, at year's end, in 1939*
- Frequency, either general—*often, generally, usually*—or specific—*monthly*
- Duration—*for a long time, still*
- With respect to expectation—*already, too soon, not yet*

A given temporal concept may be realized by a word in one language and by a phrase in another. Temporal specification is also frequently coded through morphological processes applying to the verb—tense, aspect, and the like. These are discussed in §3.15.

Temporal linkage between clauses may include

- Sequence—(*and*) *then, after, before*
- Temporal extent to, or from, a time—*since, till/until*
- Time of an event—*when*
- Temporal inclusion—*while*

Some languages have more specialized marking; for example, two events may overlap, without either being fully included within the other.

[ii] *Spatial*

- Shifters—*here, there*
- Definite locational specification. In English, this can be shown by a preposition before an NP, such as *at, to, from, in, on, into, above*. Or by a spatial adverb (some of which have developed from preposition-plus-word with spatial overtones), such as *outside, overboard, downstairs, upwards*.

We sometimes find that a number of prepositions have special properties. For example, *between* relates to the position of something in relation to two other things—*Palo Alto is between San Francisco and San Jose*. And *among* must be followed by an NP with plural reference, as in *He put the cat [among the pigeons]*.

Both spatial and temporal shifters can be used in what appear to be contradictory ways. One may work late, look at the clock, see that it is after midnight, and remark:

- (10) Goodness me, it's tomorrow already

Of course, it can't be tomorrow, it must be today. What the speaker means is 'it's the day after the pending sleep, which I haven't yet embarked on.'

Similarly with spatial adverbs. The place where I am is here. But one day, when wondering whether a coach which was supposed to pick me up had already gone, I phoned the coach company. The clerk said: *You're supposed to wait on the corner of Central Avenue and Myrtle Street.* I replied:

(11) I'm there

What I was doing was casting the sentence in terms of the clerk's orientation. If I had wanted to employ my orientation, I would have felt the need to say something more than *I'm here*, perhaps *I am here, on the corner of Central Avenue and Myrtle Street.*

3.9 Marking of core and peripheral arguments

(a) Core arguments: system

If a speaker uses a transitive clause whose predicate is 'watched', whose core arguments are 'the crafty poacher' and 'the cunning gamekeeper', the speaker must have some means for communicating to an addressee which argument is in A and which in O function—is the poacher watching the gamekeeper or is the gamekeeper keeping an eye on the poacher? That is, there should be some marking of A or O or both.

An intransitive clause involves one core argument, in S function. There is always a tendency to economize on grammatical marking. As demonstrated several times above, one grammatical element is often used for several purposes. Since S occurs in a different clause type from A and O, there is no need to use different markings (one of which could be zero) for all of S, A, and O. In fact, we find two recurrent patterns. The first involves S and A being marked in the same way (as nominative) and O in a different way (accusative), as shown in Figure 3.2.

The second is for S and O to be marked in the same way (called absolutive) and A in a different way (ergative), as in Figure 3.3.

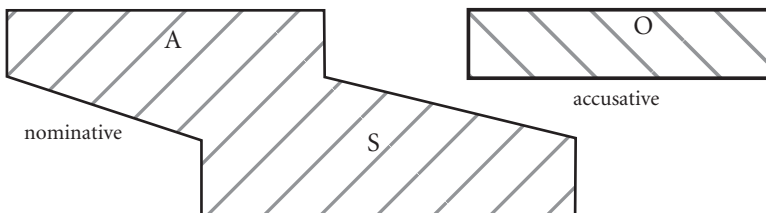


FIGURE 3.2. Nominative-accusative system

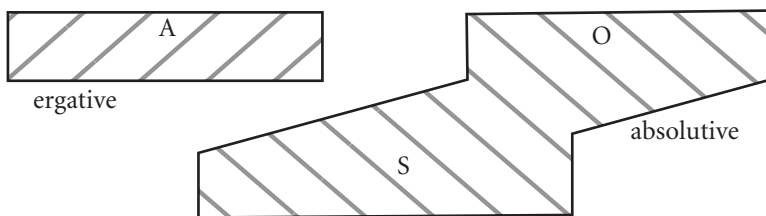


FIGURE 3.3. Absolutive-ergative system

There is a third alternative, where S, A, and O are all treated differently; this is called ‘tripartite marking’. It is rather rare and only occurs for some types of NP head, and in conjunction with (nominative-)accusative or (absolutive-)ergative marking, or both.

Many languages have a combination of accusative and ergative marking. This split can be motivated in one of three ways. First, it may relate to the nature of the head of the argument NP, in terms of the ‘nominal hierarchy’, in Figure 3.4. Basically, the further to the left an item is on the hierarchy, the more likely it is to be in A rather than in O function (for example, ‘you’ are more likely to do something to an animal than vice versa). Typically, items to the left have accusative marking for O function, with A and S often being left unmarked (nominative), while items to the right have ergative marking for A function, with S and O being unmarked (absolutive).

Accusative marking, extending in from the left, and ergative marking, from the right, must at least meet in the middle (I know of no language where this does not happen). Or they may overlap, creating a small area with tripartite marking. This can be illustrated for Cashinawa, a Panoan language from Peru. As shown in Table 3.2, 1st and 2nd person pronouns have an accusative system of case marking, proper and common nouns have an ergative system, while 3rd person pronouns show tripartite marking, with *-a* for O function (accusative), nasalization for A (ergative), and zero for S function.

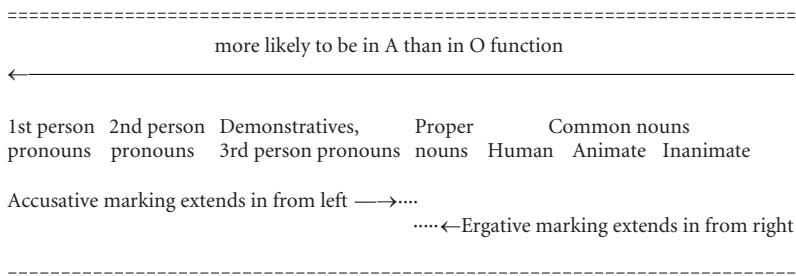


FIGURE 3.4. The nominal hierarchy

TABLE 3.2. Marking of core arguments in Cashinawa

A	∅	habū	nasalization
S	∅	habu	∅
O	-a	haa	∅
	1st and 2nd person pronouns	3rd person pronoun	proper nouns and common nouns

As is quite typical, absolutive is the formally unmarked term, shown by zero, in an ergative system, and similarly for nominative in an accusative system. (There are just a few languages with unmarked accusative; none is known with unmarked ergative.)

In other languages with split marking, the division between accusative and ergative systems may occur at a different place on the hierarchy (and there is often no ‘tripartite system’ zone of overlap).

Further kinds of split can be conditioned by:

- Tense and aspect. An ergative system may occur in past tense or perfective aspect and an accusative one elsewhere.
- Clause type. One sometimes finds an ergative/accusative split conditioned by clause type. Relative clauses tend to be ergative (and a main clause may, in contrast, be accusative). Purposive clauses tend to be accusative (and a main clause may, in contrast, be ergative). Different kinds of split are found in different languages; see Dixon (1994: 101–4).

Under (b) of §3.3, the principle for role–argument association was stated: that role which can control or initiate an activity is placed in A function. Now intransitive verbs vary in their meanings. For some—such as ‘run’, ‘stand’, ‘talk’—the referent of the S argument is likely to control the activity, as an A argument does. For others—including ‘cough’, ‘trip’, ‘die’—the referent of the S argument is unlikely to exercise control, and is similar to an O argument.

In some languages, intransitive verbs divide into two classes; for one class, S is marked like A (this is called Sa) and for the other it is marked like O (So). This is shown in Figure 3.5.

There are also languages in which S is not sharply divided. Some verbs have S marked like A, some have it marked like O, while a further set allow both possibilities. For example, a verb may mean ‘fall’ (an involuntary action) when it takes So marking and ‘deliberately throw oneself to the ground’, when marked as Sa. This is shown in Figure 3.6.

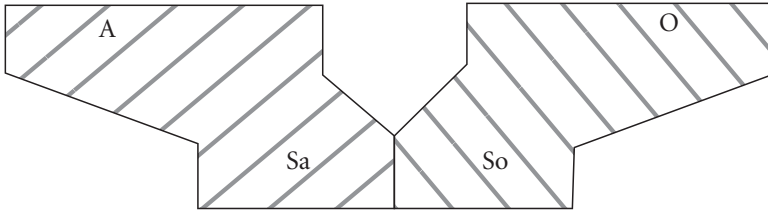


FIGURE 3.5. Split-S system

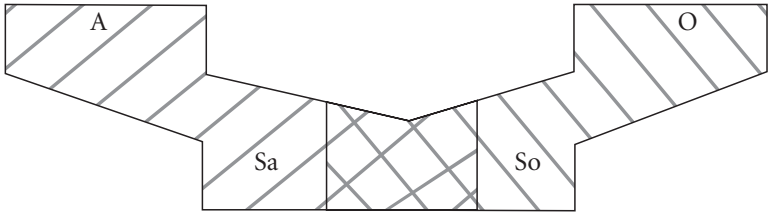


FIGURE 3.6. Fluid-S system

Sa verbs have been called ‘active’ and So ‘stative’, with ‘active-stative’ employed as the label for a split-S system. Unfortunately, the same label has also been used for fluid-S systems. As emphasized under (d) in §2.5, this terminology can be confusing. Split-S and Fluid-S are different systems, and must be accorded distinctive labels.

Discussion of all the points briefly summarized above is in §13.5.4 and, in more detail, in Dixon (1994: 70–110).

(b) Core arguments: types of marking

There are three basic ways of marking the function of a core argument.

- (i) *Marking on an NP* realizing (or partly realizing) the argument. By choice from a system of case affixes or clitics, or by an adposition—a cover term for prepositions and postpositions—which may be a separate word or a clitic (see §5.4). A clitic or adpositional word will precede or follow the NP. A case affix may go on (1) just the last word, or (2) just the first word; or (3) just the head; or (4) all words in the NP; or (5) all words of a certain grammatical type in the NP.

Sometimes, case will go on just the last word if all the words of the NP occur together; if they are separated into more than one position in the clause, it will go onto the last word of each part.

- (ii) *Marking by a bound pronominal* which realizes (or partly realizes) the argument—see (17a), (18a), and (19a) from Tiwi, in §1.10. This is generally attached to the verb or a verbal auxiliary, but in some languages

may cliticize onto the first constituent of the clause. Such marking will provide full information about a 1st or 2nd person argument but for 3rd person will only provide such information as is coded by bound pronouns (for example, gender or noun class, and number).

If both A and O have the same number and belong to the same gender or noun class, then some other mechanism needs to be brought in to distinguish them. This may be achieved by constituent order. Or, in some languages, an ergative or accusative case is optional, being used just to supplement bound pronouns when ambiguity would otherwise result.

- (iii) *Constituent order*. In some languages there is ordering of the phrasal constituents in a clause, e.g. AOV and SV, or OVA and SV. English has fairly fixed ordering AVO and SV (pronouns also have case forms, providing redundant information about syntactic function). Many languages have a degree of freedom of ordering, but may invoke a canonical order when ambiguity might otherwise result.

There are languages (such as Thai) which do not manifest any of these three mechanisms. Which argument is A and which O may be inferable from common knowledge and/or the discourse context (in life, one often encounters ‘crocodile attacks dog’ but seldom ‘dog attacks crocodile’) or perhaps made clear through the way a discourse is organized (for example, ‘there was a dog, the dog got hurt, dog crocodile attacked’).

(c) Peripheral arguments: types of marking

Bound pronouns generally mark no more than two core arguments (S, or A and O); if they do code a third argument, then it is probably also to be regarded as core for that language. Relative order of clause constituents may provide recognition of A and O but it is never available for distinguishing between types of peripheral argument. The upshot is that—even for a language which does not employ cases or adpositions for core arguments—one of these marking mechanisms is likely to be used for peripheral arguments.

Peripheral arguments may refer to space and time, as described in §3.8, and also to a variety of other roles, which vary from language to language. These include (the first four are illustrated for English):

- instrument; for example (*slice it*) *with a knife* and (*make the damper*) *with flour and water*
- accompaniment; for example (*he settled down*) *with his wife*
- recipient; for example (*he showed the weapon*) *to the policeman*
- beneficiary; for example (*she wrote the letter*) *for her illiterate cousin*

- aversive, referring to something for fear of which the action described by the verb takes place or should take place; something like '(shelter in the cave) for fear of the thunderstorm!'

Cases form a closed grammatical system; each NP of a certain type must make one and only one choice from the system. Some languages have a small number of cases and others as many as twenty, with meanings such as 'to the outside of', 'to the inside of', 'along', 'on'. The number of adpositions in a language varies, from just a handful to a hundred or more. (For example, the *OED* identifies more than 100 prepositions in English.) If there are no cases for marking peripheral arguments (or just a few of them), then adpositions will take over this task.

As mentioned under (a), absolutive or nominative (or, rarely, accusative) may receive zero marking. In contrast, peripheral NPs with a common noun as head generally require some explicit marking. However, there may be an exception in the case of proper names of places; these nouns may be used alone with either a locative or allative meaning, depending on the nature of the accompanying verb (for example, 'stay [at] Split Rock' or 'go [to] Split Rock'). In English there is just one noun, *home*, which can be used alone with allative or locative sense; for example. *She's going home* (rather than **She's going to home*) and—just in some dialects—*He's staying home* (*He's staying at home* is an alternative).

As explained in §1.11, a case affix or an adposition is an indicator of the function of an NP; it is not a lexical component of it. Some linguists use the label 'prepositional phrase' for an NP marked by a preposition, such as *to the fat man* (and they then say that the preposition is 'head' of the phrase). But prepositions have similar function to cases. If *to the fat man* is a 'prepositional phrase', then—as pointed out in §1.11—the equivalent phrase in Latin, *vir-ō obēs-ō* ('man-SINGULAR.DATIVE fat-MASCULINE.SINGULAR.DATIVE'), should be called a 'case phrase' (and should the case be head of the phrase?). Core and peripheral arguments may be realized—in whole or part—by NPs, with their functions being marked by cases and/or prepositions and/or constituent order. It is the lexemes which make up the argument—and the NP—not whatever marking convention is adopted.

Marking of core and peripheral arguments is not always as neat and tidy as set out above. In some languages, nominative is used in most circumstances to code subject (S or A), but for some verbs in certain contexts, dative or locative—whose primary use is to mark peripheral arguments—may be used for subject. Sometimes, a case other than accusative may be used to mark an argument in O function. Such non-canonical marking of core arguments is particularly found when the S or A does not exercise control over the activity,

or when the O is not physically affected. And the non-canonically marked S, A, or O may then not show all the grammatical properties associated with that syntactic function in the language. There is more information on this in §13.6. A full discussion, plus a number of case studies in individual languages, will be found in Aikhenvald, Dixon, and Onishi (2001).

The reader will by now be familiar with the idea that one grammatical element may be used for several semantic and functional purposes. There are many examples of a single affix used to mark two quite different arguments, sometimes one core and one peripheral, other times two peripheral. Typical instances involve one case form for:

- ergative (A function) and instrumental ('with' an implement)
- instrumental and locative ('at')
- dative ('to' or 'for' a recipient or beneficiary) and allative ('to' a place)
- locative and allative

We also encounter one nominal affix used both for dative case (marking function within a clause) and for genitive (marking possessive function within an NP).

In each instance, there will be some criterion in the grammar enabling the linguist to recognize two grammatical elements, which just happen to be coded in the same way. For example, in Dyirbal ergative and instrumental fall together in form, but have quite different syntactic behaviour (for example, the antipassive derivation affects an NP with ergative marking, but leaves unchanged an instrumental NP); see §4.3. Also in Dyirbal, allative and dative affixes have the same form on nouns, but are quite different on determiners which occur with a noun in an NP.

3.10 Complement clauses

We saw in §3.5 that each argument slot in a clause may be filled by a phrase with a noun (or sometimes a verb, or a pronoun or demonstrative) as head. There is an alternative—in many (but not all) languages, a type of clause can fill an argument slot. As briefly described in §3.2, this is called a 'complement clause' (CoCl).

(Some linguists describe this phenomenon in a most misleading manner, saying that what we call complement clause is 'simultaneously NP and sentence' (see Stenson 1981: 63). First, it is not a sentence but a clause; see (a) in §2.5. And although a complement clause functions in a similar way to an NP by filling an argument slot, it is—as shown in the discussion of (1–2) below—quite different from an NP.)

Typically, a complement clause functions as O argument for verbs such as ‘see’, ‘hear’, ‘know’, ‘believe’, and ‘like’, as in *Mary_A believes [that John_A did [the burglary]_O]_{CoCl:O}*. Some languages allow a complement clause in other functions. In English it may fill the A slot for verbs from the ANNOYING type—see (b) in §3.7—as in [*That John_A did [the burglary]_O]_{CoCl:A} annoyed Mary_O*. And in English just a few verbs may accept a complement clause in S function; for example [*That John_A returned [the stolen goods]_O]_{CoCl:S} *didn’t matter*.*

The main criteria for recognizing a complement clause are:

- (I) It has the internal structure of a clause, at least as far as core arguments are concerned. Whether peripheral arguments may be included—and if so, which ones—varies from language to language.
- (II) A complement clause functions as a core argument of another clause. If a complement clause is in O function, for instance, it should show at least some of the syntactic properties of O in that language. For example, in English a THAT complement clause in O function may generally be passivized, as in [*That John_A did [the burglary]_O]_{CoCl:S} *is believed by everyone*.*
- (III) A complement clause will always describe a proposition; this can be a fact, an activity, or a state (it may not be a place or a time).

The importance of criterion (I) can be illustrated by comparing the English sentences (1), with a complement clause as A argument, and (2), with an NP as A:

- (1) [John’s playing the national anthem]_{CoCl:A} pleased Mary_O
- (2) [John’s playing of the national anthem]_{NP:A} pleased Mary_O

The complement clause in (1), *John’s playing the national anthem*, has similar structure to a main clause, with an A NP, *John* (with possessive ‘s, one marker of this variety of complement clause in English) and an O NP, *the national anthem*, which immediately follows the verb with no preposition intervening. The verb of the complement clause is *play*, with suffix *-ing*, the other marker of this kind of complement clause in English.

In contrast, *John’s playing of the national anthem*, in (2), is an NP where the nominalization *playing* is head noun, *John’s* is possessive modifier, and *of the national anthem* is a post-head modifying prepositional phrase (similar to *of the table* in *the legs of the table*).

There are five criteria for distinguishing between a complement clause, as in (1) and an NP, as in (2):

- (i) In the complement clause, the O NP, *the national anthem*, immediately follows the verb, as in a main clause. In the NP, the underlying O must be marked by *of*.
- (ii) In the NP of (2), the possessor *John's* is a modifier of the head noun and can be replaced by another modifier such as the article *the*, giving *The playing of the national anthem pleased Mary*. In (1), the subject, *John*, bears 's, which is a marker of this variety of complement clause; *John's* cannot here be replaced by *the*; that is, we cannot have **The playing the national anthem pleased Mary*.
- (iii) The verb of the complement clause, *playing*, may be modified by an adverb. As in a main clause, this typically follows the object, as in [*John's playing the national anthem competently*]_{CoCl:A} *pleased Mary*_O. The head noun of the NP, the nominalization *playing*, can be modified by an adjective, which must precede it, as in [*John's competent playing of the national anthem*]_{NP:A} *pleased Mary*_O.
- (iv) A complement clause is, like every other clause type, negated by *not*; we get *John's not playing the national anthem annoyed Mary*. In contrast, an NP—as in (2)—may only be negated by prefixing *non-* to the head noun, giving *John's non-playing of the national anthem annoyed Mary*.
- (v) An ING complement clause may include auxiliaries *have (-en)* and *be (-ing)*; one can say *John's having been playing the national anthem pleased Mary*. An auxiliary may not be used with a nominalization, such as we have in (2).

These two structurally different—although superficially similar—sentences have different meanings. (1) states that Mary was pleased that John (who might in the past have often appeared to be rather unpatriotic) played the national anthem, whereas (2) describes how she was pleased by the manner of his playing it, so that it was performed mellifluously. Each sentence could be reduced to *John's playing pleased Mary*, which would be ambiguous between the two syntactic structures and meanings.

What makes (1) and (2) such an intriguing pair of sentences is that the verb *play* adds *-ing* in being nominalized, the same suffix that marks the verb of one variety of complement clause. Many verbs in English have different forms in the two circumstances. Compare:

COMPLEMENT CLAUSE	NP WITH NOMINALIZATION AS HEAD NOUN
John's refusing the offer	John's/the refusal of the offer
John's demolishing the building	John's/the demolition of the building
John's discovering the money	John's/the discovery of the money
John's knowing the secret	John's/the knowledge of the secret

Each pair shows a semantic difference similar to that between (1) and (2).

Some languages have a single complement clause construction while others distinguish several. The three most common types, cross-linguistically, all occur in English, as THAT, ('s) -ING, and (FOR) TO clauses. These are exemplified in:

- (3) John_A remembered [that he had locked the house]_{CoCl:O} (just the fact of doing so)
- (4) John_A remembered [locking the house]_{CoCl:O} (every detail of it)
- (5) John_A remembered [to lock the house]_{CoCl:O} (but he'd mislaid the keys and was unable to do so)

Whereas the THAT complement clause, in (3), states that John just remembers *the fact* that he locked the house, the ING clause, in (4), implies that he remembers *details of the activity*—locking the back door first, then the french window at the side (and noticing a crack in the glass there), and finally the front door, where he had to wiggle the key to get it to turn. A Modal to complement clause relates to *the potentiality of the subject initiating the activity*. If nothing else is stated, one would infer that he did lock the house, but there could be a parenthetical addition, as in (5), providing a reason why—despite proper intentions—he was unable to do so.

Towards the end of §1.11, a distinction was drawn between Primary-A semantic types (including *take*, *chew*, *burn*, etc.) whose arguments can only be realized as NPs, and Primary-B types, for which at least one core argument may be either an NP or a complement clause. In §1.9 there was discussion of the complement clause properties for verbs in English from two Primary-B types, LIKING and THINKING.

Secondary concepts may be realized as verbal affixes (or through other morphological processes) or as lexemes, depending on the language. It is useful to distinguish three varieties of secondary verb. Secondary-A do not add any arguments to those of the verb in the complement clause; for example, *John_A started/tried [to eat [the cake]_O]_{CoCl:O}*. Secondary-B may add an argument, which is often identical to the complement clause subject, the latter then being omitted, as in *I_A wish/hope [to eat [the cake]_O]_{CoCl:O}* (one could say *I_A wish/hope [for Mary to eat [the cake]_O]_{CoCl:O}*). Secondary-C verbs require an additional argument which is unlikely to have the same reference as complement clause subject (and when it does have, neither can be omitted); for example, *Mary_A forced [John_A to eat [the cake]_O]_{CoCl:O}*.

Not all languages have complement clauses. But there must be some means of stating that someone sees or believes or remembers a fact or an activity or a

potential action. This may be achieved by employing some other clause type, as a ‘complementation strategy’. All sorts of strategies are found. The simplest is just apposition of clauses, as in ‘John did the burglary; that annoyed Mary’. Or a Serial Verb Construction may be employed, as in ‘John remembered-locked the door’—see (b) in §3.4.

There is detailed discussion of complement clause constructions, and of complementation strategies, in Chapter 18.

3.11 The sentence

A clause consists of a predicate plus appropriate arguments. It may include either of the two kinds of embedded clause, relative clause and complement clause. A simple sentence has just a main clause. A complex sentence has one or more clauses linked together.

‘Clause’ and ‘word’ (and also ‘phrase’) are relatively easy to define (see §3.1, §3.4, and Chapter 10). Not so ‘sentence’. In written language it can be defined as what comes between two full stops (or periods); but the use of punctuation marks is variable from writer to writer and from style to style. Schools teach that one should not begin a sentence with a conjunction. But people do, for good communicative effect. Already in this paragraph there have been two clauses commencing with *but*, one preceded by a semicolon and the other by a full stop. Either punctuation mark would be acceptable in each case. Or just a comma could have been used. Notice *or* at the beginning of the last clause; this could have been preceded by comma or semicolon or full stop.

Despite this stylistic variation, it would not be hard to devise an algorithm for recognizing sentences in written English. Examine each instance of a full stop, and consider whether it could be felicitously replaced by comma or semicolon or colon. If it can be, then so replace it. After this replacement is complete, what now comes between full stops is a sentence. In the last paragraph, the full stop would be replaced before ‘not so sentence’, before ‘but people do ...’, before ‘either punctuation mark ...’, and before ‘or just a comma ...’, leaving the paragraph with five sentences.

Linguists transcribe colloquial speech in English (or in any other language) and bemoan the lack of grammatical neatness, making it difficult to recognize sentences (and sometimes also clauses and phrases). In such cases, one should request the speaker to edit the transcription of what they have said so that it becomes, in their opinion, acceptable written English. Then apply the procedure outlined above, to recognize sentence boundaries.

For languages without a strong tradition of writing, or where only a small proportion of the population is literate, one must deal just with the spoken mode. There is wide variation between languages in how easy it is to recognize

‘sentence’. For one language on which I have done extensive fieldwork, there is a clear prosodic signal. In Jarawara, the final syllable of a sentence is nasalized and bears rising intonation (Dixon 2004a: 530). In some languages, there is a grammatical marker which signals the end of a sentence. There are languages where the verb must come sentence-finally, with this being a clear indicator of the boundary.

Working with Fijian, I found the recognition of sentences far from easy, and wrote as follows. ‘Dividing a discourse up into sentences is not an easy task in Fijian (as it is not in most other languages). The various criteria may not coincide—for instance, intonation may suggest two sentences where the syntax of linkers points to one, or vice versa. A given string of clauses may be grouped into sentences in several different ways, depending on the pace of the narrative or conversation, the speaker’s attitude to the material, and what [they] want to focus on, among other factors. Similarly, the listener or linguist hearing the clauses may group them into sentences in different ways’ (Dixon 1988: 257–8).

One should distinguish between full sentences and those which are reduced as a result of ellipsis of elements inferable from discourse context. In English, an NP on its own may constitute a reduced sentence; for example, *An apple* as reply to the question *What would you like?* Indeed, I have heard an utterance being just a prefix, as when the reply to *Would you prefer smoking or non-smoking?* was *Non-*.

Languages vary as to how much reduction is permitted. In one Australian language, Yidiñ, a reply to a question must be a full clause, with predicate and appropriate arguments. A reply to ‘What are you going out for?’ could be ‘I’m going to hunt wallabies’. Dyirbal is Yidiñ’s southerly neighbour (but there is no close genetic relationship between them). It pursues an entirely different strategy, preferring the most economical reply. A question ‘What are you going out for?’ would be likely to bring forth a one-word response, something like *barrgan-gu* (wallaby-DATIVE), that is ‘for wallabies’. As will be inferred, Dyirbal has a far higher proportion of reduced sentences than does Yidiñ.

A complex sentence must involve two (non-embedded) clauses, linked together. In a two-clause linkage, it is useful to recognize:

the Supporting clause	S
the Focal clause	F (this determines the mood of the whole sentence)

There is likely to be a grammatical marker attached to one of the clauses (in most cases, not to both) indicating the type of linkage involved:

marker attached to Supporting clause	Ms
marker attached to Focal clause	Mf

For example:

- (1a) Ms S Mf F
Because the Duke has married — he does not have to sell the
an heiress, castle
- (1b) — The Duke has married so he does not have to sell the
an heiress, castle

The order of clauses is invariable in (1b) but may be reversed in (1a):

- (1c) F Ms S
The Duke does not have to sell the because he married an heiress
castle,

In each of (1a–c), the NP *the Duke* occurs in the first clause—which is Supporting in (1a–b) and Focal in (1c)—and is referred to by anaphoric *he* in the second. Interestingly, these could be reversed just in (1a), with the Duke stated in the second clause, and referred to by cataphoric (‘referring forwards’) *he* in the first:

- (1a’) Because he has married an heiress, the Duke does not have to sell the
castle

That is, in English, one can get cataphoric replacement of an NP by a pronoun such as *he* just in a Supporting clause which has a marker and which precedes the Focal clause.

A number of types of clause linkage can be recognized. Four of the main mechanisms are briefly illustrated, from English, in Table 3.3 (note that there are many other markers for each box; just a sample is included here).

These can be illustrated in:

TEMPORAL:

- (2) (a) After the Duke married an heiress, he did not have to sell the castle
(b) The Duke married an heiress, and then he did not have to sell the
castle

TABLE 3.3. Linkage types, illustrated with sample markers in English

LINKAGE TYPE	MARKERS WITH SUPPORTING CLAUSE	MARKERS WITH FOCAL CLAUSE	
		CONJUNCTIONS	HALF CONJUNCTIONS
Temporal	after, when	(and) then	
Consequence	because, in case, lest	(and) so. in order that, (in order) to	therefore
Addition		and	moreover
Contrast	although	but	however

CONSEQUENCE: (1a–c) and:

- (1) (d) The Duke has married an heiress, therefore he does not have to sell the castle
 (e) The Duke has married an heiress, in order that he will not have to sell the castle

ADDITION:

- (3) (a) The Duke has married an heiress, and he came into a substantial inheritance
 (b) The Duke has married an heiress; moreover he came into a substantial inheritance

CONTRAST:

- (4) (a) Although the Duke has married an heiress, he does have to sell the castle
 (b) The Duke has married an heiress, but he does have to sell the castle
 (c) The Duke has married an heiress, however he does have to sell the castle

Markers in the ‘half conjunctions’ column (this term is from Henry Sweet 1891: 143) may occur initially or—when used contrastively—later in a sentence; for example *The Duke has married an heiress; he does, however, have to sell the castle*. The conjunctions and the markers of a Supporting clause must be clause-initial.

Not every language has each of the four types of clause linkage:

- *Temporal*. It is likely that all languages have some means for marking temporal connection between clause, including a ‘when’ linkage and generally also ‘(and) then’.

‘When’ refers to a temporal connection between events; for example, [*When*]_{Ms} [*Stephen ate pork*]_S [*he got stomach-ache*]_F. There may be temporal linkage between generalized events, and then *when* can be replaced by *if*, with similar meaning:

- (5) *When/if*_{Ms} [*John eats pork*]_S [*he gets stomach ache*]_F

Conditional *if* is thus an extension from temporal *when*. In many—but not all—languages, there are also conditional constructions which do not involve a temporal linkage, such as:

- (6) *If*_{Ms} [*Shanghai is bigger than London*]_S, (*then*_{Mf}) [*Shanghai must be a huge conurbation*]_F

and also counterfactuals, such as:

- (7) If_{M_s} [America had adhered to the Monroe doctrine]_S [the world would have been a safer place]_F

(‘Conditional’ has sometimes been referred to as a ‘mood’. This is misconceived; it is a marker of clause linking.)

- *Consequence* (subsuming ‘cause/reason’, ‘result’ and ‘purpose’). It is likely that all languages have one or more grammatical techniques for indicating a linking of Consequence between two clauses. Most often there is an M_f marker (typically purposive ‘so that’ or ‘in order to/that’) rather than an M_s one (such as ‘because’).

A subtype is ‘possible consequence’; for example, [*I’ll lift up that mat*]_F [*in case*]_{M_s} [*my lost pen is under it*]_S. There may sometimes be an overtone of apprehension, as in [*Don’t cross the border*]_F [*lest*]_{M_s} [*you get arrested*]_S!

- *Addition*. This is a universal type of linkage, shown in many languages simply by apposition of clauses.
- *Contrast*. Only found in some languages.

In many languages, it is possible to have linkage within linkage. In English one could say:

- (8) Since_{M_s} [although_{M_s} [John is an asthmatic]_S [he is very strong]_F]_S, [we will employ him]_F

However, speakers would normally prefer to avoid such involuted grammar with a sequence of two Supporting markers (and would prefer something like [*Although*]_{M_s} [*John is an asthmatic*]_S [*he is very strong*]_F]_S, so_{M_f} [*we will employ him*]_F).

There is a further type of clause linkage, disjunction, shown by (*either*) or in English. As utilized by logicians, this relates to a closed set of possibilities—‘(either) X or Y, there being no other possibility’. An example sentence is:

- (9) Either John will drive Matilda to school or she’ll walk there

In real life, no set of choices is ever closed; there are always further possibilities however unlikely they may be (Matilda’s aunt may happen to be passing and give her a lift, or Matilda may be sick and stay at home all day). Over the 3,000 or 4,000 languages spoken across the world, only a small minority have a marker of closed disjunction. It is more usual to work in terms of open disjunction, including something like ‘may’ or ‘might’ in each clause. If we hear something which translates as ‘John may drive Matilda to school or she

may walk there', this will have a similar import to (9) in English—it is highly likely that one of the stated possibilities will eventuate, but not one hundred per cent certain.

(This discussion of clause linking is a drastically truncated version of Dixon 2009.)

3.12 Negation

Languages vary considerably in how they treat negation. The only universal mechanism is negation of a main clause—in some languages this is the sole way of showing negation. Others have a multiplicity of techniques: for negating sentence, various types of subordinate clause, predicate, and also argument. Many languages have a negative interjection 'no', which can be the full reply to a polar question. However, quite a number of languages lack this. In Jarawara, for example, the only mark of negation is suffix *-ra* to a verb which is predicate head. A negative reply to a question such as 'Are you going?' could only be, literally, 'I go-not' (together with various tense/modality and mood specifications).

Clausal negation may be shown by a separate word (such as *not* in English), or by a verbal affix or clitic (as just mentioned for Jarawara), or by a negative verb. For example, in Quileute—a Chimakuan language from Washington State, USA—the negative morphemes *wa* or *é·* or the two in succession (*é· wa*) function as the main verb, and the action negated is expressed by a subordinate verb'.

A comprehensive discussion of negation across a variety of languages is planned for Volume 3. Meanwhile, some example of the types of negation found in English can be offered. We can first compare:

- NEGATION WITH SCOPE OVER SENTENCE

- (1) I don't beat my dog because I love her (I do beat her, but for some other reason)

This is [*I beat my dog because I love her*]_{NEG}.

- NEGATION WITH SCOPE OVER MAIN CLAUSE

- (2) I don't beat my dog, because (I don't beat her, and the reason I love her I don't is that I love her)

This is [*I beat my dog*]_{NEG}, *because I love her*. Note that sentences (1) and (2) have different intonation patterns, partially indicated by commas in the written form.

Now compare:

- NEGATION WITH SCOPE OVER PREDICATE

- (3) He could not write (he could decline the invitation to review the book
the review if he is afraid of offending the author)

- NEGATION WITH SCOPE OVER MAIN CLAUSE

- (4) He could not/couldn't write the review (say, he is illiterate, or too busy)

Indeed, these two types of negation can co-occur:

- (5) He couldn't not write (he would like to decline the task but can't since
the review it is part of his job contract)

There is also the possibility of negating an NP, as in *No honest man would tell a lie* (with similar meaning to *An honest man would not tell a lie*). And English has a number of inherently negative lexemes, including *dissuade (from)*—with similar meaning to *persuade not (to)*—and *forget*—with similar meaning to *not remember*.

The contrast between 'positive' and 'negative' is referred to as the 'polarity' system.

3.13 Morphology

Syntax—study of the organization and interrelation of grammatical elements—is relatively straightforward when compared with morphology—study of the composition of words. Indeed, the term 'syntax' was introduced into English about 1600. 'Morphology' was first used around 1820 in biology, for study of the form and structure of organisms, being taken into linguistics about 1860, to describe the structure of words.

The basis of a word is a lexical root to which various morphological processes may apply, to produce the finished item. Edward Sapir outlines the range available in his classic book *Language* (1921: 61–81). We can recognize six major types of morphological process. The first, compounding, involves the linking together of roots; the other five deal with processes applying to a root or a compound.

1. Compounding. Two roots may be joined to form one stem; for example, *tapeworm*. A compound (plus any further morphological processes which may have applied) makes up one grammatical word. In English, it bears a single stress. In addition, a compound generally has a lexical meaning which is not just the sum of the meanings of its parts. The classic example is 'blackbird, with a single stress, referring to a particular species of bird (the European common

thrush, *Turdus merula*); this contrasts with the phrase 'black 'bird, where the two words are in syntactic construction, each having its own stress and each retaining its own meaning (the phrase refers to any bird which is black).

In English, a compound can involve a combination of noun, verb, and adjective roots. Generally, the items being combined have underlying syntactic connection, for example:

gunman	man who uses a gun to kill people
gunpowder	powder used in a gun
razor-sharp	sharp as a razor
sky-blue	a blue colour like the blue of the sky

In these compounds, the first word effectively modifies the second—they describe a type of man, powder, sharp, and blue respectively. Other compounds may describe a person with certain attributes, such as:

cry-baby	a person who cries like a baby
cut-throat	a person who cuts other people's throats (now extended to ruthlessness in business)

Other words are more idiomatic, as *sweethearts* for a pair of lovers (although this could conceivably be interpreted as saying that each has a heart which feels sweet towards the other).

The meaning of a word can vary, according to who uses it and to whom. When a man refers to a woman as his *girlfriend*, this implies a romantic (and, potentially, sexual) relationship. But when a woman refers to another woman as her *girlfriend*, then in most circumstances it simply describes a boon companion (with no sexual overtones).

Serial verb constructions—mentioned under (b) in §3.4—may involve a type of verb compounding. There is also 'noun incorporation', whereby what is in underlying structure an argument or a part of an argument becomes incorporated into the verbal word. Sometimes an O argument is fully incorporated, producing an intransitive clause; from 'I drank a beer' we get, literally, 'I beer-drank'. Othertimes the head of an O argument is incorporated with what was the possessor now becoming the O; from 'I am making Subih's house' we get, literally, 'I am house-making Subih'.

Different languages employ different types of compounds, and to varying degrees. (Linguists do vary in the criteria they employ for recognizing compounds, and in their analytic approaches to them.) Sapir (1921: 65) mentions that 'Eskimo and Nootka and, aside from paltry exceptions, the Semitic languages' cannot compound lexical roots.

2. Reduplication. This describes the repetition of all or part of a root either before or after it. (Particularly sophisticated reduplication involves repetition

of a syllable in the middle of a root.) The meanings carried vary enormously between languages. Reduplication of a noun can indicate plural (*yuri.yuri* ‘many kangaroos’ in Dyirbal) or diminutive (*sqa’yux* ‘man’, *sqe’qeyux* ‘boy’ in Salish) or even 1st person possession (*at^yu* ‘aunt/uncle’, *at^yut^yu* ‘my aunt/uncle’ in Parecis). In Amharic, full reduplication of an adjective indicates selection; thus, *rajjim* ‘long’, *rajjim.rajjim* ‘those that are long’.

Reduplication of a verb may indicate ‘do repeatedly’ (for example, *amá-n* ‘arise’, *manamá-n* ‘go up and down’ in Yuma), or ‘do intensively’ (*walapa* ‘boil’, *wawalapa* ‘boil vigorously’ in Tonkawa), or ‘happen continuously’ (*erasi* ‘he is sick’, *erasirasi* ‘he continues to be sick’ in Siriono), and so on.

A verb in Jarawara may take any of three kinds of reduplication; these can be illustrated with *joko* ‘push’:

- initial syllable, ‘do a little bit’ *jo.joko* ‘push a bit’
- first two syllables, ‘do with force’ *joko.joko* ‘give a tremendous shove to’
- last syllable, ‘many participants’ *joko.ko* ‘push lots of things’

Either of the first two processes may be combined with the third—*jo.joko.ko* ‘push lots of things a bit’, *joko.joko.ko* ‘give lots of things a tremendous shove’.

Most languages have reduplication as a productive process, often with different forms and meanings for different word classes. This process is absent from the major languages of Europe, including English (except for minor idiomatic forms such as *yum-yum* and *goody-goody*).

3. Shift of stress or change of tone. In languages with tones, a morphological process can involve change of tone. For example, in Anywa, a Western Nilotic language spoken in the Sudan and Ethiopia, *mɛŋ* (with low tone) is the verb ‘be deaf’ while *mɛ̃ŋ* (with a rising tone) is the noun ‘deafness’.

Similarly, in languages with variable placement of stress (or accent), change of position can indicate a morphological process. English examples include nouns *‘import* and *‘project* (with stress on the first syllable) and verbs *im‘port* and *pro‘ject* (with stress on the second syllable).

4. Internal change. English has many examples of a morphological process involving vowel change in the middle of a root. This applies for almost all the ‘strong verbs’; for example:

- | | | | |
|------|--------|------|--------|
| take | /teɪk/ | took | /tu:k/ |
| give | /gɪv/ | gave | /geɪv/ |
| sing | /sɪŋ/ | sang | /sæŋ/ |
| | | sung | /sʌŋ/ |

Internal change may also apply to consonants. There are a few examples in English, including noun *house* /haus/, and verb *to house* /hauz/.

5. Subtraction. A morphological process may add something to a verb—this is affixation, dealt with under (6)—or it may delete something. For instance,

in Samoan the optimal analysis involves recognizing each verb root to end in a consonant. For past tense *-ia* is added to the root, and for imperative the final consonant (whatever it may be) is omitted. For example:

	‘see’	‘break’	‘weigh’	‘cheat’
POSTULATED ROOT	silaf	gaum	fuat	?oleg
PAST TENSE FORM	silaf-ia	gaum-ia	fuat-ia	?oleg-ia
IMPERATIVE FORM	sila	gau	fua	?ole

The alternative analysis would be to take the imperative form as root. But past tense would then be *-fia* or *-mia* or *-tia* or *-gia* (and there are further alternatives with other verbs). Which past tense form to use with a given verb would not be predictable, but would have to be specified on an individual basis. It is plainly simpler to take each root as ending in a consonant (even though in Samoan no word may end in a consonant) and to state a subtraction process for imperative.

6. Affixation. Not all languages show all of the five morphological processes just discussed: compounding, reduplication, shift of stress or change of tone, internal change, subtraction. But they all—save perhaps for a totally isolating language, such as Vietnamese is reputed to be—have affixation, making this the most common method for modifying the meaning, changing the word class, or indicating the function of a root.

Suffixes, added at the end of a word, are the most common types of affix—in English and in most other languages. For example *mann-ish-ness*, *dark-en-ing*, *equal-iz-ed*. Some languages only show suffixes; others also employ prefixes, added at the beginning of a word, such as in *dis-own* and *super-un-fit*. (No languages are known which have prefixes but no suffixes.)

Less common are circumfixes, where a prefix and a suffix must occur together. For example, in Fijian an intransitive verb may be made causative by adding prefix *va’a-* and a transitive suffix; *ravi* ‘lean’ becomes *va’a-ravi-ta* ‘put leaning’. We also find infixes, which are affixes inserted within a root. The paradigm for verb ‘walk’ in Dakota is:

I walk	ma-wá-ni
You walk	ma-yá-ni
He/she walks	mani

That is, 1sg and 2sg subjects involve bound pronouns *-wá-* and *-yá-* respectively being inserted into the root as infixes. For 3sg there is no infix (or, one could say, a zero infix).

Normally, a root has a certain extent, with affixes being added before and/or after or at one place in the middle. Words in Semitic languages show an unusual structure, with the root consisting of three consonants and

morphological processes inserting vowels between and around them. Consider Egyptian Colloquial Arabic, which has root *k-t-b* for ‘write’, giving rise to words:

káatib ‘clerk’, kátaba ‘clerks’ kitáab ‘book’, kútub ‘books’

There may also be prefix and/or suffix, as in:

kátab he wrote yí-ktib he writes/will write
 kátab-it she wrote tí-ktib she writes/will write

Other unusual morphological processes may involve change of value within a vowel harmony system, or could involve application of something like nasalization or aspiration or glottalization or rhoticization as a prosody over all or part of a word.

In many languages it is useful to distinguish between ‘derivational’ and ‘inflectional’ processes, in connection with the use of labels ‘root’, ‘stem’, and ‘word’. A root is a basic lexical element. When two roots are compounded they form a stem:

ROOT.ROOT=STEM for example, *baby.sit* (verb)

A derivational process applies to a root and forms a stem, or applies to a stem and forms a further stem. It may change word class, or simply add semantic modification. For example:

	ROOT <i>central</i>	(adjective)
add SUFFIX <i>-ize</i> , forming	STEM <i>central-ize</i>	(verb)
add PREFIX <i>de-</i> , forming	STEM <i>de-central-ize</i>	(verb)
add SUFFIX <i>-ation</i> , forming	STEM <i>de-central-iz-ation</i>	(noun)

and

	STEM <i>baby-sit</i>	(verb)
add SUFFIX <i>-er</i> , forming	STEM <i>baby-sitt-er</i>	(noun)

Derivations are optional, inflections obligatory. After all derivational processes have applied, then a choice must be made from the appropriate inflectional system, relating to the word class of the final stem. Stem *baby-sit* is a verb and thus takes a verbal inflection, giving *baby-sits* or *baby-sat* or *baby-sitting*. Stem *baby-sitter* is a count noun and must be inflected for number, thus we can get plural *baby-sitter-s*. *Central-ize* and *de-central-ize* are verbs, and must take verbal inflection; *de-central-iz-ation* is a noun and takes nominal inflection.

In the description of many languages, recognition of a distinction between inflectional and derivational processes is a useful analytic tool. But for others

it is not. Like every other feature of basic linguistic theory, this is something which should be invoked when it can play a useful role in description and explanation, but not forced into operation when it is not appropriate. There is fuller discussion of inflection and derivation in §5.3.

An inflectional system is a closed set of items, one of which must be chosen. Each term in the system has meaning in contrast with the others. This was illustrated in Table 1.1 of §1.4 for number systems, with the meaning of ‘plural’ being complementary to the meanings of the other terms in the system—‘more than one’, or ‘more than two’, or ‘more than a few’. Quite often, a term in a system will involve a zero process. If all the other terms in the system are realized by suffixes then it is convenient to say that the term involving a zero process is realized by a zero suffix ($-\emptyset$). Consider the regular inflection for number of an English noun, such as *dog*:

SINGULAR *dog- \emptyset* referring to just one dog
 PLURAL *dog-s* referring to two or more dogs (or to no dogs)

Singular, marked by zero (a blank in the slot available for a number suffix in the template of noun structure in English), has the specific meaning of referring to a single individual, in contrast to plural, which has the specific meaning of referring to more than one individual. Consider:

SINGULAR SUBJECT *The dog- \emptyset stand-s in the yard*
 PLURAL SUBJECT *The dog-s stand- \emptyset in the yard*

The noun *dog- \emptyset* , as subject, selects the 3rd person singular present/generic ending, orthographic *-s*, on the verb, whereas noun *dog-s* selects the non-3rd-singular ending $-\emptyset$.

There are two ways in which zero plays a role within grammatical systems. Zero may be the sole realization of a term, as for singular on nouns in English. Or zero may be one of a number of alternative realizations of a term. Plural is shown by orthographic *-s* on most count nouns in English, but by zero on some, including *sheep* (also *deer*, *fish*, and just a few others). We get:

SINGULAR SUBJECT *The sheep- \emptyset stand-s in the yard*
 PLURAL SUBJECT *The sheep- \emptyset stand- \emptyset in the yard*

Number agreement with the 3rd singular ending *-s* on the verb shows that the \emptyset on *sheep* in the first sentence above is the invariant realization of singular number, while agreement with the non-3rd-singular ending $-\emptyset$ on the verb in the second sentence shows that the \emptyset on *sheep* is here the zero allomorph of plural.

We have seen that there are two uses of zero within a grammatical system:

- (a) The sole realization of a term—for example, singular in the number system for nouns in English.
- (b) One of the alternative realizations of one (or more) terms in a system—for example, plural in the number system.

It would not be possible for there to be two terms in a system both having entirely zero realization, type (a). However, there could be two terms of type (b), each having a zero allomorph (in different, or in overlapping, circumstances). And, as illustrated for *sheep*, a system can have one term of type (a) and one (or more) of type (b). This may lead to potential ambiguity, which is likely to be resolved from the discourse context.

It is often instructive to set out the forms of each word in a paradigm (called the Word-and-Paradigm model). Consider the following partial paradigm for ‘good’ in Latin:

SINGULAR	GENDER		
CASE	masculine	feminine	neuter
nominative	<i>bonus</i>	<i>bona</i>	<i>bonum</i>
accusative	<i>bonum</i>	<i>bonam</i>	<i>bonum</i>
dative	<i>bonō</i>	<i>bonae</i>	<i>bonō</i>
PLURAL	GENDER		
CASE	masculine	feminine	neuter
nominative	<i>bonī</i>	<i>bonae</i>	<i>bona</i>
accusative	<i>bonōs</i>	<i>bonās</i>	<i>bona</i>
dative	<i>bonīs</i>	<i>bonīs</i>	<i>bonīs</i>

One set of inflections applies to an adjective in Latin, with information concerning three grammatical systems (gender, number, and case) fused into a single suffix. One could segment off the root as *bon-*, but it is not possible to segment suffix *-us*, for instance, into singular, nominative, and masculine elements. We have here one morphological process, of affixation, providing portmanteau realization of three quite independent grammatical systems. As discussed under (e) in §1.10, these are systems of quite different character: gender is an inherent feature of the noun that heads the NP in which the adjective occurs, as modifier; number is a referential feature of the NP; and case marks the function of the NP in its clause. What the Word-and-Paradigm method does is show, in paradigmatic form, the oppositions involved.

Now look at a partial paradigm for a noun where the realizations of grammatical elements are not fused together—*mularri* ‘initiated man’ in Dyirbal.

CASE INFLECTION	NO DERIVATIONAL	COMITATIVE
	PROCESS HAS	DERIVATIONAL PROCESS
	APPLIED	(‘with’) HAS APPLIED
absolutive (S, O functions)	<i>mularri</i>	<i>mularribila</i>
ergative (A function)	<i>mularrigu</i>	<i>mularribilagu</i>
ablative (‘from’)	<i>mularrijunu</i>	<i>mularribilajunu</i>

The comitative process derives an adjectival stem, ‘with an initiated man’. Now, unlike in the Latin example, it is a straightforward matter to separate out root and affixes, with a hyphen between each:

mularri *mularri-bila*
mularri-gu *mularri-bila-gu*
mularri-ηunu *mularri-bila-ηunu*

We see that the comitative derivational process forms a stem by adding suffix *-bila*, and the system of case inflections shown here has suffix *-ηunu* for ablative, *-gu* for ergative (on these forms), and zero for absolutive (we should really write *mularri-∅*). Such a segmentation into root(s) and affix(es) is called an Item-and-Arrangement method of showing morphological structure.

A useful distinction which is often made is that between ‘free’ and ‘bound’ forms. A free form can make up a complete word on its own, a bound form cannot. All affixes are, by their nature, bound. In Latin, all roots are bound since they must take some affix to form a word (one cannot just say *bon-*). A noun in Dyirbal is a free form, since it can stand alone. Derivational processes are optional, but case inflection is obligatory; however, absolutive case has zero realization, so *mularri* with no affix—that is, with an inflectional process which has zero realization—is a full word. It will be seen that the free/bound distinction is really only useful where the sole (or at least the major) morphological process is affixation. With internal change or stress shift or tone change, the free/bound contrast is not applicable.

For every word, in every language, a number of segmental phonetic units can be recognized, whose substitution results in a change of meaning; these are the letters of an ideal alphabet. The term ‘phoneme’ (based on Greek *phōnēma* ‘sound’) was introduced about 1880 for this unit. By analogy, the term ‘morpheme’ (based on Greek *morphē* ‘image, form’) was introduced at about the same time to describe a segmentable unit within word structure. (Note that the term ‘morpheme’ came into use about twenty years after ‘morphology’.) Just as a sentence consists of a number of clauses, a clause of a number of phrases, a phrase of a number of words, so—it was supposed—a word should consist of a number of morphemes.

This works fine for agglutinating languages (see §5.5 for this term), such as Dyirbal, where a word can be clearly segmented into root(s) and affix(es). *Mularri-bila-gu* consists of three morphemes, root *mularri* ‘initiated man’, comitative derivational suffix *-bila* ‘with’, and ergative inflectional suffix *-gu*. Sometimes morphemes have alternative forms, and about 1948 (around seventy years after ‘morpheme’ was first used) the term ‘allomorph’ was introduced to describe these. Ergative case in Dyirbal has allomorphs *-ŋgu* after a disyllabic stem ending in a vowel and *-gu* after a root or stem of more than two syllables ending in a vowel (there are further allomorphs used with roots and stems ending in a consonant).

Analysis of a word into morphemes is not so appropriate for fusional languages like Latin, or when processes other than affixation or compounding have applied, or when phonological rules have applied after affixation. For example, the English verb forms *take* and *took* involve a process of internal change; there is no affix that can be hyphenated off.

In 1954, Charles Hockett discussed and contrasted the ‘Item-and-Arrangement’ method with ‘Item-and-Process’ (the labels were his), the latter describing the methodology followed here, and emanating from Sapir. He preferred Item-and-Arrangement ‘because nowadays we like to be as formal as possible’ and because it is ‘clearly more homogenous’ than Item-and-Process. But why should a theoretical model be made homogeneous when the facts of language are heterogeneous?

How then to deal with *take* and *took*? Hockett suggested five alternative ways of tackling this problem (the fact that he was led to give five suggests that none was considered satisfactory). Let me just quote two: ‘(3) *took* is an allomorph of the morpheme which appears elsewhere as *take*, plus a zero morpheme of /ed/; (4) *took* is a discontinuous allomorph /t...k/ of *take*, and an infix allomorph /u/ of /ed/.’ This was ingenious but misguided. Why try to make internal change into a non-canonical variety of affixation (and similarly, presumably, for stress shift and tone change)? Internal change is simply internal change.

Let us now consider a paradigm which is less easily segmentable than that for Dyirbal but lacks the fusion of Latin. This comes from the Australian language Yidiñ (and follows the same parameters as the paradigm for Dyirbal), illustrating nouns *buña* ‘woman’ and *wagu:ja* ‘man’:

CASE INFLECTION	NO DERIVATIONAL PROCESS HAS APPLIED	COMITATIVE DERIVATIONAL PROCESS (‘with’) HAS APPLIED
absolutive (S, O functions)	<i>buña</i>	<i>buña:y</i>
ergative (A function)	<i>buña:ŋ</i>	<i>buñayŋgu</i>
ablative (‘from’)	<i>buñam</i>	<i>buñayimu</i>

CASE INFLECTION	NO DERIVATIONAL PROCESS HAS APPLIED	COMITATIVE DERIVATIONAL PROCESS (‘with’) HAS APPLIED
absolutive (S, O functions)	<i>wagu:ja</i>	<i>wagujayi</i>
ergative (A function)	<i>wagujajgu</i>	<i>wagujayi:ŋ</i>
ablative (‘from’)	<i>wagujamu</i>	<i>wagujayim</i>

These could be segmented into root and affixes but there would be a number of indeterminacies. For instance, should the form *buña:ŋ* be segmented into *buña:-ŋ*, where the long vowel is assigned to the root, or into *buña-:ŋ*, where the long vowel is assigned to the suffix (and similarly for *buña:y*, and others)? What about the long vowel in the middle of *wagu:ja*, which does not appear in any of the other forms in this paradigm? Is it an infix indicating absolutive case when no derivational process has applied?

It will be seen that an Item-and-Arrangement approach to this data is misconceived. What we have is straightforward processes of affixation, followed by two phonological rules.

ROOTS	<i>buña</i> ‘woman’, <i>waguja</i> ‘man’
DERIVATIONAL PROCESS	add comitative suffix <i>-yi</i> ‘with’
INFLECTIONAL PROCESS	add case suffixes
	absolutive \emptyset
	ergative <i>-ŋgu</i>
	ablative <i>-mu</i> after an odd number of syllables
	<i>-m</i> after an even number of syllables

This generates paradigms:

CASE INFLECTION	NO DERIVATIONAL PROCESS HAS APPLIED	COMITATIVE DERIVATIONAL PROCESS (‘with’) HAS APPLIED
absolutive (S, O functions)	<i>buña-\emptyset</i>	<i>buña-yi</i>
ergative (A function)	<i>buña-ŋgu</i>	<i>buña-yi-ŋgu</i>
ablative (‘from’)	<i>buña-m</i>	<i>buña-yi-mu</i>

CASE INFLECTION	NO DERIVATIONAL PROCESS HAS APPLIED	COMITATIVE DERIVATIONAL PROCESS (‘with’) HAS APPLIED
absolutive (S, O functions)	<i>waguja-\emptyset</i>	<i>waguja-yi</i>
ergative (A function)	<i>waguja-ŋgu</i>	<i>waguja-yi-ŋgu</i>
ablative (‘from’)	<i>waguja-mu</i>	<i>waguja-yi-m</i>

The following rules now apply:

1. In every word with an odd number of syllables, the penultimate vowel is lengthened. This applies to the following forms from the paradigms:

buña-ŋgu → *buña:-ŋgu* *waguja* → *wagu:ja*
buña-yi → *buña:-yi* *waguja-yi-ŋgu* → *waguja-yi:-ŋgu*

Stress is assigned to the first syllable involving a long vowel; if there is no long vowel, it goes onto the first syllable of a word. Further stresses are then assigned, recursively, to the syllables next but one before or after a stressed syllable. Rule (1) ensures that if a word has an odd number of syllables, there will be an even number of feet, each consisting of one stressed and one unstressed syllable, plus one left-over unstressed syllable.

2. If a word has an odd number of syllables, ending in $V_1:-C_1(C_2)V_2$ (with a morpheme break before C_1) and if C_1 is an allowable word-final consonant (a nasal, liquid or y), then $(C_2)V_2$ is omitted. This applies to the following forms from the paradigms:

buña:-ŋgu → *buña:ŋ* *buña:-yi* → *buña:y* *waguja-yi:-ŋgu* → *waguja-yi:-ŋ*

The form *wagu:ja* does not reduce for two reasons—there is no morpheme boundary before the *j*, and *j* is not an allowable word-final segment.

These rules now generate the occurring forms. It can be seen that the rules are such as to ensure that a maximum number of words in Yidiñ have an even number of syllables; that is, a word should consist of a whole number of disyllabic feet, either trochaic or iambic. Of the twelve forms in the paradigms, only one—*wagu:ja*—is left with an odd number of syllables.

In summary, a word is created by taking a root (or two roots, in a compound) and applying morphological processes, which can be of various types. These may be followed by phonological processes, such as those just illustrated for Yidiñ. Without doubt, the most common process is affixation. But it is misguided to assume that affixation is the canonical process and that all others should be taken to be peculiar variants of affixation (can there be an ‘affix of tone change’?). To do this obscures much and explains nothing.

A certain type of information may be shown by an inflectional system in one language but by derivational processes in another. For example, in English number is an inflectional system on count nouns—*dog-ø* refers to one animal, *dog-s* to more than one. In other languages, number affixes are derivational and optional. For example, in Dyirbal *guda* refers to any number of dogs, either one or more than one; the derivational suffix *-jarran*, with dual reference, may optionally be added—*guda-jarran* ‘two dogs’.

Grammatical systems which are often realized through morphological processes include case (§3.9), negation (§3.12), tense, aspect, and evidentiality (discussed in §3.15), gender and noun class (§3.16), number (§3.17), and definiteness (§3.18). In §3.19 we examine dependencies between these systems.

3.14 Derivations

There are two kinds of derivational processes: (a) those which change word class; and (b) those which don't, but simply add a semantic modification. They are illustrated in Table 3.4 for English, which has a wide variety of derivational affixes. There are no derivational processes forming an adverb directly from a verb, and none forming a stem of another word class from an adverb.

We can now discuss the types in turn.

(a) Derivations which change word class

- (i) *Verb from noun*. These typically apply to a noun referring to a state, the derived verb then meaning 'make the state come about', as with *glori-fy* 'make have glory' and *glamor-ize* 'make have glamour'. The suffix *-ate* may be used with nouns that have concrete reference, as *hyphen-ate* 'insert hyphens into' and *oxygen-ate* 'mix with oxygen'.
- (ii) *Adjective from noun*. Meaning here may indicate 'made of', as in *wood-en*, 'behave like a', as in *mann-ish*, and 'be characterized by', as in *passion-ate*, *beauti-ful*.
- (iii) *Adverb from noun*. There are just a few of these derivations in English, including *home-wards* 'towards home', *clock-wise* 'circular motion in the same direction that the hands of a clock move', and *side-ways*

TABLE 3.4. Samples of derivational processes in English

DERIVATION APPLIES TO ROOT OF WORD CLASS:	DERIVATION FORMS A STEM OF WORD CLASS:			
	noun	verb	adjective	adverb
noun	<i>librari-an</i> <i>consul-ate</i> <i>aristocra-cy</i>	<i>glori-fy</i> <i>glamor-ize</i> <i>hyphen-ate</i>	<i>passion-ate</i> <i>wood-en</i> <i>mann-ish</i>	<i>home-wards</i> <i>clock-wise</i> <i>side-ways</i>
verb	<i>employ-er</i> <i>arrange-ment</i> <i>loath-ing</i>	<i>dis-inherit</i> <i>circum-navigate</i> <i>pre-judge</i>	<i>forget-ful</i> <i>turn-able</i> <i>attract-ive</i>	—
adjective	<i>nice-ness</i> <i>delica-cy</i> <i>hard-ship</i>	<i>dark-en</i> <i>simpl-ify</i> <i>stabil-ize</i>	<i>redd-ish</i> <i>yellow-y</i> <i>extra-clever</i>	<i>clever-ly</i> <i>ten-fold</i>

‘with the side of an object facing forwards instead of, as would be expected, the front’.

- (iv) *Noun from verb*. This type of derivation spans a range of meanings, including:
- Describing a volitional agent (as A argument); for example, *employ-er*, *kill-er*, *organiz-er*.
 - Describing an instrument used in an action; for example, *heat-er*, *transport-er*.
 - Describing an O argument of the verb; for example, *employ-ee*, *pay-ment*.
 - Describing the locus of an activity; for example, *entr-y*, *resid-ence*.
 - Describing an activity (*shoot-ing*), state (*admir-ation*, *loath-ing*), result (*arrange-ment*), or property (*resembl-ance*).
- Often, one derivational process may have different semantic effect with verbs of different semantic types; for example *-ment*, when added to a verb, can derive a noun referring to a state (*bewilder-ment*) or a result (*arrange-ment*) or a place (*settle-ment*), among others. (A full account of deverbal nominalization in English is in Dixon 2005a: 317–52.)
- (v) *Adjective from verb*. Describes a potential property of an argument of the verb. It may relate to the subject (*forget-ful*, *turn-able*, *attract-ive*) or to the object (*forgett-able*).
- (vi) *Noun from adjective*. Forms an abstract noun referring to the property; for example, *nice-ness*, *delica-cy*. Or to a state associated with the property word, as *hard-ship* ‘having to undergo circumstances that are hard’.
- (vii) *Verb from adjective*. ‘Becoming’ and/or ‘making’ something get into the state described by the adjective; for example, *dark-en* ‘get dark, make dark’, *simpli-fy* ‘make simple’, *stabil-ize* ‘become stable, make stable’.
- (viii) *Adverb from adjective*. Many adjectives form an adverb by adding *-ly*, meaning ‘do it in that way’; for example, *clever-ly*. (There is discussion of which adjectives from which semantic types form which type of adverb, in English, in Dixon 2005a: 381–5.) Just a few other suffixes belong in this set; for example *ten-fold* as in *It increased tenfold*, meaning that it became ten times as big as before.

(b) Derivations which do not change word class

- (ix) *On adjectives*. These typically specify the degree of a property, or the extent to which it applies. We get *redd-ish* ‘a tinge of red’, *yellow-y*

‘not quite yellow, but similar to it’. There are also many prefixes, including those indicating negation (*un-happy*, *dis-loyal*) and degree (*extra-clever*, *semi-conscious*, *ultra-bright*).

- (x) *On verbs*. There appear to be no suffixes which simply add semantic modification to a verb in English (other languages have plenty). There are, however, a fair number of prefixes, mostly of Latin or Greek origin. Examples include *dis-inherit* ‘deprive of inheritance’, *circum-navigate* ‘navigate around’, *pre-judge* ‘judge before all the evidence which should be taken into account has been assembled’, and *hyper-correct* ‘change according to a mistaken idea of what is correct’.
- (xi) *On nouns*. English has a wealth of suffixes, with a range of meanings. These include:

- Person associated with a place, as *librari-an*, *Africa-n*, *London-er*.
- Place associated with an important person or with a group of people, as *king-dom*, *consul-ate*, *orphan-age*.
- General names for a social system (*aristocra-cy*) or social relationship (*friend-ship*).

There are many others; for example *fish-monger* ‘person who sells fish’ and *bedd-ing* ‘things which go on a bed, such as sheets and blankets’.

Having broadly surveyed derivations in English, we can now essay a few general remarks. As mentioned before, each portion of a grammar intersects with every other part. As described in §3.5, in some languages only nouns may function as arguments and only verbs as predicates; such languages are likely to have a number of processes for deriving a noun stem from a verb root and vice versa. In contrast, those languages which allow both noun and verb to function either as argument or as predicate will have less need of such derivations (although they may well have some—to indicate different semantic types of nominalization, for example).

Languages vary greatly in the number of word-class-changing derivations they take; some have all the types illustrated in Table 3.4, others only a few. The most common derivation is probably noun from verb; the nominalization may refer to agent, patient, instrument, activity, locus, etc. Also fairly frequent is the derivation of verb from adjective (and sometimes also from noun), indicating ‘become’ or ‘make’.

There is also variation concerning derivations which do not change word class. Adjectives typically show morphological processes which indicate degree—‘very’ or ‘just a bit’. Nouns may form augmentatives and diminutives (‘a big one’ and ‘a little one’). Another common derivation applies to the name of a place or geographical feature, referring to someone or something

associated with it (similar to *Chicago-an*, but also relating to such things as ‘mountain’ and ‘grass’).

The richest set of derivations, cross-linguistically, concerns verbs. When a language has a strict division between intransitive and transitive verbs—see (b) in §3.3—there are likely to be processes deriving a transitive stem from an intransitive root, and vice versa. These are discussed in §3.20. Some of the verbal processes which do not affect transitivity are outlined in the next section, on non-spatial setting.

As discussed in §3.8, the spatial setting of a clause is in the majority of cases shown through a peripheral argument, typically an NP marked by an adposition or a case with local meaning. Most often, if a language has a case system, there will be just a handful of local cases; the canonical ones are locative (‘at’), allative (‘to’), and ablative (‘from’) (in some languages, one case covers both locative and allative meanings). However, some languages have many more. *Lezgian*, from the North-East Caucasian language family, shows more than a dozen local cases, specifying ‘at’, ‘to’, and ‘from’ an object, or ‘at’, ‘to’, or ‘from’ with respect to behind, or over, or on, or in it.

It is possible for spatial setting to be shown not through nominal morphology (a case system) but by means of verbal morphology. There may be derivational processes applying to the verb indicating ‘associated motion’, whether the action is done while or after ‘going’ or ‘coming’ (exemplified for *Yidiñ* in §1.11).

A particularly rich set of spatial prefixes to the verb is found in *Koasati*, a Muskogean language spoken in Louisiana. There is a nine-term system, whose terms are chosen according to whether the action is: (1) on the ground or in fire; (2) in water; (3) on a raised, artificial, or non-ground surface; (4) on or in a vertical plane; (5) in the middle of; (6) on the face (for example, ‘A spot is hanging on the horse’s forehead’); (7) on the mouth (‘I will shave my beard’); (8) on the throat (‘I have dry skin on my throat’); (9) in the throat (‘I am throat-dry’, meaning ‘I am thirsty’). Interestingly, the *Jarawara* language, spoken down in Amazonia, has verbal suffixes with very similar meanings to the first three in *Koasati*: ‘on the ground’, ‘in water’, and ‘on a raised surface’.

A further technique for indicating spatial placement involves body-part nouns being employed to indicate general orientation. For instance, in many Oceanic languages, *mata* means ‘face’ and also ‘front of’; for ‘in front of him’ one says, literally, ‘at his face’. These are discussed further in §16.5.

3.15 Non-spatial setting

It is a common misconception that all languages have a grammatical system of tense marking, and that it prototypically has three terms, ‘past’, ‘present’, and ‘future’. In fact, there are a fair number of languages which do not have

anything which could be called a tense system in their grammar. And when there is one, it typically covers ‘past’ (sometimes indicating different gradations of past time) and ‘non-past’, much less often ‘future’.

The predicate and core arguments of a clause describe an event. It will have a spatial setting—as discussed in the previous section—and a non-spatial setting. This includes a number of parameters most of which are briefly presented here. All will be discussed in more detail in Volume 3.

1. *Evidentiality*. There may be a grammatical system providing obligatory information about the **evidence** on which a statement is based; for example, whether seen, heard, inferred, assumed, or reported. This was illustrated in §1.5 and §1.6.
2. *Reality*. A contrast between **realis**, referring to something that has happened or is happening, and **irrealis**, referring to something that didn’t happen in the past (but could have) and to all or most of the post-present domain. Within irrealis we can get a number of **modality** choices, all referring to some aspect of the future. The modal verbs in English include reference to prediction (*will*), possibility (*might*), potential (*can*), necessity (*must*), obligation (*should*). Modalities in other languages include optative (that which is wished for).

It was shown under (b) in §3.2 how modality is quite different from a mood system (typically covering declarative, imperative, and interrogative).

3. *Degree of certainty*. Indicating whether a given activity or state is probable, possible, improbable, etc.
4. *Phase of activity*. The activity may be beginning or continuing or resuming or finishing.
5. *Completion*. The term **perfect** has a variety of uses but it is most commonly employed to describe a past action which is completed but still has present relevance. It is opposed to **imperfect**, referring to something which began in the past and is still continuing.
6. *Boundedness*. An event which has a definite end-point is called **telic**; for example, *John sang the national anthem*. One which is unbounded is **atelic**; for example, *John sang at the party*.
7. *Extent*. An event may be **punctual**; that is, it happens more or less instantaneously (for example, a bomb explodes or a goal is scored). Or it may be **durative**—alternatively called ‘continuous’ or ‘progressive’—unfolding over a period of time (for example, writing a poem or playing a game).
8. *Composition*. In Slavic languages (and in some others) each event must be accorded one of two aspectual values. It may be **perfective**, where the

event is regarded as a whole, without regard for its temporal constituency (even though it may be extended in time). If, in contrast, it is marked as **imperfective**, this indicates focus on the temporal make-up of the event. For example, ‘John baked the cake (perfective) while Mary was sleeping (imperfective)’; this indicates that Mary started sleeping before John baked the cake, continued during this activity, and then slept some more afterwards. (Care must be taken to distinguish perfective/imperfective from perfect/imperfect—(5) above—which, despite their similar names, refer to quite different contrasts.)

The term **aspect** is employed in varying ways by different linguists. In the narrowest sense it is used just for perfective/imperfective. But the scope is often extended to refer to extent, or boundedness, or completion (and more besides).

9. *Tense*. In contrast with the categories listed above, tense is a shifter (see §3.7)—what is ‘present’ at the time of speaking will soon become ‘past’.

All languages allow some assertions which are not time-related, such as ‘Dogs bark’. But most statements refer to an action, state, or property situated within a time frame with respect to the moment of speaking, and this may be shown by a system of tense choices.

As mentioned before, since ‘future’ is basically an irrealis area, the most common tense systems just distinguish ‘past’ from ‘non-past’. Some languages do include ‘future’ in their tense system, generally referring to something that can confidently be predicted to happen. (For example, on a Wednesday one could say, without fear of misleading, ‘Tomorrow will be Thursday’.) As illustrated in §1.7, there may be several divisions within past tense, and also sometimes several within future (but never more in future than in past).

Tense may also indicate one time with respect to another. In English one can say *When John arrived* (past₁), *Mary had departed* (past₂ before past₁), utilizing the ‘previous’ aspect *have...-en* (Dixon 2005a: 209–29); other languages can indicate such time inclusion by means of their tense system.

Each language includes some of the parameters which describe non-spatial setting. Those which lack a tense system are likely to include in their grammars a selection from systems (5–8).

Tense is typically realized through an inflectional system on the verb. The other systems may be shown by morphological processes that are either derivational or inflectional. There is, however, considerable variation. Any of the parameters may be realized through an optional auxiliary verb or by particles, which may be separate words or clitics. These forms most often occur in juxtaposition with the verb, but are in some

languages placed elsewhere; for instance, after the first constituent of the clause.

There is one final point, which has worried many linguists. Is tense a category of the predicate (generally filled by a verb) or of the clause as a whole? That is, in *John kicked Fred* is it the event ‘John kick Fred’ which relates to past time, or is it just the action ‘kick’? And similarly for evidentiality, reality, degree of certainty, phase, completion, boundedness, extent, and composition. I can perceive no basis for this question. The predicate refers to an action, state, or property which is located within the domain of time; the predicate selects appropriate arguments which, together with it, make up a clause that has its time location determined by the predicate. Tense, aspect, and so on are properties of the predicate *and* of the clause—of both, not of one or the other.

In some languages a secondary use of tense is to apply to an individual NP. One might say something like ‘[My wife]_{FUTURE} is eating [the pie]_{YESTERDAY.PAST}’, which would be rendered in English by *My wife-to-be is eating yesterday’s pie*. This use of tense simply applies to one argument. But the canonical employment of a tense system or of any other system showing non-spatial setting—whether realized through a morphological system applying to the verb, or in some other way within the clause—applies equally to predicate and to clause.

3.16 Noun classes and genders, and classifiers

Gender was one of the first grammatical categories to be described. In the fifth century BCE, Protagoras had recognized three genders in Greek, by correlating the orthographic endings of words with the ‘masculine’, ‘feminine’, and ‘inanimate’ natures of the things they stood for. Well-known Indo-European languages such as Greek, Latin, German, and French have a gender system which consists of three terms—‘masculine’, ‘feminine’, and ‘neuter’—or just two—‘masculine’ and ‘feminine’. There is always some semantic basis to the allocation of nouns to gender classes (hence the names of the classes), and always also some exceptions. That is, most designations for people of the female sex will be nouns with feminine gender. (Recently, the grammatical label ‘gender’ has had its meaning extended to describe the reproductive type of a person, replacing ‘sex’; this makes statements such as those of the previous sentence difficult to formulate.)

When African languages came to be investigated by linguists it was found that their nouns fall into a number of classes similar to the gender classes

of Indo-European languages, except that there were usually a fair number of them and they often did not include a distinction between ‘masculine’ and ‘feminine’. For example, one class might cover people (of both sexes), another long or extended objects, a third fruits and non-extended objects, a fourth artefacts, and so on. ‘Gender’—carrying with it the expectation of a ‘masculine’/‘feminine’ distinction—seemed inappropriate and a new term ‘noun classes’ was introduced. This is now the label in most general employment. ‘Gender’ is used for a small system of noun classes, which includes a sex-based contrast.

There are two main criteria for recognizing a system of noun classes.

1. It codes a grouping of all the nouns of the language into a smallish number of classes. The number of noun classes generally varies from two to about ten (only rarely are there more). Generally, each noun belongs to just one class. However, there may be a small number of exceptions, nouns which can be in one of two (or more) classes depending on their reference. For example *vračh* ‘doctor’ in Russian may be either masculine or feminine depending on its referent, and similarly for *jaja* ‘baby’ in Dyirbal.
2. There must be some overt indication of the class of a noun (when in certain functional positions) within the clause in which it occurs, and this must *not* be entirely within the noun-word itself.

The class of a noun can be shown on modifiers to it within the NP. For example, in French one says *un homme intelligent* ‘a clever man’ and *une femme intelligente* ‘a clever woman’. One cannot tell the gender of *homme* and *femme* from the forms of the words, but one can by noting that *homme* takes masculine forms of the indefinite article and the adjective ‘clever’, *un* and *intelligent*, while *femme* takes the feminine forms, *une* and *intelligente*.

Alternatively, gender can be marked on the predicate, as in Jarawara:

- (1) *bati_S noho-ka*
father be.hurt-DECLARATIVE.MASCULINE
(His) father is hurt
- (2) *mati_S noho-ke*
mother be.hurt-DECLARATIVE.FEMININE
(His) mother is hurt

Here the gender of the noun filling the intransitive subject (S) slot is shown by the gender-marked form of the declarative suffix which is attached to the verb in predicate slot. It is *-ka* for masculine and *-ke* for feminine.

In Swahili, noun class is shown in three ways: by a prefix to the noun itself, by a prefix to modifying words within the NP, and by a cross-referencing prefix within the predicate. In the following sentence, *-kombe* ‘cup’ belongs to the general inanimate noun class, marked by prefix *ki-* in the singular (as here) and *vi-* in the plural.

- (3) [ki-kombe ki-dogo ki-wili]_s ki-mevunjika
 CLASS-cup CLASS-small CLASS-two CLASS.OF.S-be.broken
 The two small cups are broken

In summary, noun class may or may not be marked on the noun itself (it is in Swahili, but not in French or Jarawara) but, for it to be recognized as a grammatical category, there must be some marking of the class of a noun outside that noun.

There are just a few examples of languages with two different systems of noun classes, at different places in the grammar—typically, a small gender system for pronouns (sometimes extended to verbal agreement) and a larger noun class system for adjectival and numeral modifiers.

Noun classes are generally realized through affixation or other morphological processes. In languages which lack a system of noun classes (and in some that have one), there is often a set of classifiers. These are generally free forms, there is typically a large number of them (sometimes into the hundreds), and they occur only accompanying a noun within its NP. For example, in the Australian language Yidiñ many (but not all) nouns may be accompanied by a classifier:

miña	gangu:l
EDIBLE.ANIMAL.CLASSIFIER	grey.wallaby
grey wallaby	

mayi	gubu:m
EDIBLE.VEGETABLE.CLASSIFIER	black.pine
black pine nut	

There are always some nouns which do not take any classifier (in Yidiñ these include ‘dog’, ‘sun’, and ‘boomerang’). And a fair number of nouns may occur with more than one; for example, besides *mayi gubu:m* ‘black pine nut’ (with the ‘edible vegetable’ classifier *mayi*), one finds *jugi gubu:m* ‘black pine tree’ (with the ‘tree’ classifier *jugi*).

There are a number of different varieties of classifier. These include numeral classifiers (for example, ‘three HUMAN.CLASSIFIER woman’ for ‘three women’), possessed classifiers (for example ‘chicken my-PET.CLASSIFIER’ for ‘my chicken’), verbal classifiers (for example, ‘I ROUND.OBJECT-picked coconut’),

and locational classifiers (for example, ‘ON-BRANCH.LIKE.CLASSIFIER tree’ for ‘on a tree’). Some languages feature classifier sets of several types.

Aikhenvald (2000) provides an informed and comprehensive treatment of noun categorization devices—noun classes (including genders) and classifiers.

3.17 Number systems

It was pointed out under (a) in §3.7 that virtually every language has a number distinction in its pronoun system. However, linguists do argue about the status of ‘number’ here. For example ‘2nd person plural’ may be used for addressing a group of people (and referring to all of them), but also when addressing one person, then meaning ‘you (singular) and others’. Is the latter instance really to be called a plural (since it does not involve several you’s)?

Number systems of the following sizes were illustrated in Table 1.1 of §1.4:

- A two-term system, {singular, plural}, is the most common.
- A three-term system, {singular, dual, plural}, is well attested.
- A four-term system is not too uncommon. Most often the extra term is ‘paucal’, referring to ‘a few’—{singular, dual, paucal, plural}. Sometimes it does refer to three—{singular, dual, trial, plural}.

As noted in §1.4, ‘plural’ has a different meaning in each type of system, complementary to the other terms—‘more than one’, ‘more than two’, ‘more than a few’, and ‘more than three’ respectively.

Some number systems include additional terms, such as ‘collective’, referring to a group of similar objects, and ‘associative’, indicating a set linked together by some characteristic; for example ‘SMITH-ASSOCIATIVE.PLURAL’ (‘the Smiths’) could refer to the whole Smith family (including in-laws, whose name is not Smith, and perhaps also close friends, who may have a different name).

Many—but not all—languages have a number system applying to word classes other than pronouns.

1. **Nouns.** There may be an inflectional system—as in English—whereby a noun can only be quoted in number-specified form (to avoid number specification one has to employ disjunction of the number-specified forms, as when a policeman reports *The house was entered by person or persons unknown*). For languages without an inflectional system of number, the citation form of a noun will mean ‘one or more’, interpreted according to context.

Often, number marking may apply just to certain nouns. In English, only to count nouns; one cannot as a rule add the plural suffix to non-count nouns such as *mud* or *sincerity*. It is not uncommon to find number marking

confined to human nouns, or to animates. In Jarawara, for example, all and only animates—including sun, moon and stars, which are here accorded animate status—are obligatorily marked for number (by a non-singular 3rd person pronoun within the predicate).

Some languages have special plural forms just for a few nouns, typically those referring to types of people ('man', 'woman', 'boy', 'girl'). Or some of these terms may have a suppletive plural; in Jarawara, for instance, 'child' is *inamatewe* and 'children' *matehe*.

2. Adjectives. In languages where adjectives have similar grammatical properties to nouns, they may share the same inflections. There are languages where number is not shown on nouns but is marked on adjectives (often, being fused with noun class). Some languages have just a few nouns (such as recent loans) which do not mark number, but this is shown on a modifying adjective.

3. Demonstratives. These typically show number, as *this/these* and *that/those* in English; however, this is far from universal.

4. Content interrogatives (often also showing an indefinite sense). Only show number in a minority of languages (including Finnish).

5. Verbs. A predicate may include some indication of the number of one or more core arguments in one of three ways:

- Through a bound pronoun associated with the verb, within the predicate.
- Through a morphological process to the verb which marks the number of a core participant. For example, in Axininca Campa, an Arawak language spoken in Peru, there is a verbal suffix which indicates that at least one core participant is plural, without specifying which (thus, in a transitive clause, one can't tell whether A or O or both have plural reference).
- Some verbs have suppletive stems relating to the number of a core argument; this is S in an intransitive and almost always O (scarcely ever A) in a transitive clause. For instance, in Comanche, an Uto-Aztecan language, 'hold, carry' is *yaa* for a singular and *himi* for a plural O. In Sumerian, an ancient language from Mesopotamia, 'live' is *til* for a singular and *sig* for a plural S.

Most—but not all—languages have numeral lexemes. There may be just two or three of these (say, 'one', 'two', 'many') or there may be a full set, with an algorithm allowing the creation of a number specification of any magnitude. 'Numerals' are sometimes a subclass of adjectives, sometimes a subclass of verbs, and sometimes a separate word class.

There will be a fuller discussion of number systems in grammar, and of numeral lexemes, in Volume 3.

3.18 Definiteness

A definite article appears in some well-known languages—for example, Greek, German, French, English—but is in fact found in only a smallish proportion of languages worldwide. It seems to be an important feature of languages in which it does occur, but many languages get by without it. How is this?

In fact, the definite article should probably be regarded as an optional extra in those languages which do have it. There are styles of English in which the definite article is routinely omitted, with little apparent loss of communicative efficiency. For example, newspaper headlines typically omit *the* and all forms of *be*, as in (where an underlined blank indicates an omission):

- (1) drug squad hit by leader's departure

And definite articles are not employed in the style used by some (but by no means all) cookery books, as in:

- (2) Clean cauliflower ... Transfer mixture to a soufflé dish ... Bring water to boil again

(Note that the indefinite article, *a*, is retained here.)

Languages differ a little in how and when they use their definite article. For example, in Portuguese one would say *a minha mãe*, including the feminine form, *a*, of the definite article, where in English *my mother* could not be preceded by *the*.

The definite article can be realized as a morphological process to a noun (or some other constituent of an NP). In English it is generally a proclitic, /ðə=/, attached to the first word of an NP. Other languages show other possibilities.

In the Polynesian language Tongan there are two mechanisms for showing definiteness. Stress is generally on the penultimate vowel of a word, but a morphological process of stress shift can apply, moving it to the last syllable and indicating definiteness. In addition, there are two 'articles' which come at the beginning of an NP, *e* and *ha*. These can be combined in three ways:

ARTICLE <i>e</i> AND STRESS SHIFT	<i>e afó</i>	'the fishing line (an already identified referent)'
ARTICLE <i>e</i> AND NO STRESS SHIFT	<i>e áfo</i>	'a fishing line (as opposed to other types of thing, such as a spear or a rope)'
ARTICLE <i>ha</i> AND NO STRESS SHIFT	<i>ha áfo</i>	'a fishing line (as opposed to other fishing lines, other tokens of the same type)'

The definite article—in English and other European languages—indicates that the referent of the NP in which it occurs should be identifiable to the

addressee. This identification could be (i) from what was said just before in the discourse, or (ii) inferred from the situation in which the discourse occurs. Sentence (3a) illustrates (i), with alternative ways of conveying the same information being shown in (3b)—with a demonstrative as NP head or NP modifier—and in (3c)—with a 3rd person pronoun.

- (3) (a) We heard a screechy noise, and the noise frightened my brother
 (b) We heard a screechy noise, and that (noise) frightened my brother
 (c) We heard a screechy noise, and it frightened my brother

If English were to lose its definite article, then (1b) or (1c) would do just as well as (1a).

As an illustration of (ii), if a screechy noise sounded out, I might say *Listen to the noise, what is it?* If English lacked articles, and a noun such as *noise* would make up a full NP, then *Listen to noise!* would be fully comprehensible in these circumstances; alternatively, a demonstrative could be employed. Definite articles typically develop out of demonstratives, and these two kinds of grammatical element are often substitutable one for the other.

When a language has no definite article, there is usually some way of indicating that an NP has an identifiable referent. This may be achieved by using a classifier or noun class marker. For example, in Gola, a West Atlantic language from Liberia and Sierra Leone, the inclusion of a noun class prefix indicates definiteness, as in *kul* ‘a tree’, *ke-kul* ‘the tree’.

In its realization, definiteness may be linked with some other grammatical category. In both Turkish and Bengali, for instance, accusative case marking is obligatory on an NP in O function only if it has definite reference.

Some languages have what could be called a ‘definiteness strategy’. Ngiyambaa, from Australia, has bound pronominal clitics. A 1st or 2nd person clitic cannot co-occur with the corresponding free form (which has identical semantic content), but the 3rd person clitic *-na* (marking S, O, or indirect object function) can be used with a free form NP and then indicates definiteness.

Compare Ngiyambaa sentences where a 3rd person argument is shown just by a clitic, in (4a), or just by an NP, in (4b), or by both, as in (4c):

- (4) (a) $\eta\text{aa-nhi-ju-na}$
 see-PAST-1sgA-3sgO
 I saw him
- (b) $\eta\text{aa-nhi-ju}$ burraay_O
 see-PAST-1sgA child+ABSOLUTIVE
 I saw a child/children

- (c) η aa-nhi-ju-na burraay_O
 see-PAST-1sgA-3sgO child+ABSOLUTIVE
 I saw the child (lit. I saw him, child/children)

In (4b) the NP *burraay*, without any reinforcing clitic, has an indefinite meaning ‘a child/children’. The addition of *-na*, in (4c), indicates definiteness, and also singular number, ‘the child’.

3.19 Dependencies between grammatical systems

Two grammatical categories can be independent of each other in a given language. There may be, say, a polarity system of {positive, negative}, and a four-term tense system in both positive and negative clauses. But sometimes one system is dependent upon another. That is, the number of choices available in one system may be dependent on the choice made from another system. As mentioned in §1.11, in Amele (from the Gum family, spoken in Papua New Guinea) in clauses with positive polarity there is a distinction between today’s past, yesterday’s past, and remote past, and also between future and relative future; but negative clauses show a single past and a single future tense. We say that the number of choices available in the tense system depends on which term is chosen from the polarity system. That is: Polarity > Tense (tense depends on polarity).

It is instructive to examine dependencies—where these exist—between a number of types of grammatical system. (For most dependencies, just one example is mentioned here; there are in most cases quite a few more instances.)

- I. Polarity of a clause (the only polarity contrast found in all languages). In some languages this contrasts with polarity of the predicate; see (3) and (4) in §3.12.
- II. Tense, aspect (in the narrow sense of imperfective/perfective), and evidentiality. As discussed in §3.15, each of these applies equally to predicate and to clause. The following dependencies between these systems are attested:
 - Tense > Aspect. For example, in Yimas, from New Guinea, an aspectual distinction is made only in present tense.
 - Aspect > Tense. Russian distinguishes three tenses—past, present, and future—in imperfective aspect, but only two—past and future—in perfective.
 - Tense > Evidentiality. Tuyuca, a Tucanoan language from Colombia, has five evidentiality choices—visual, non-visual, apparent, second-hand, and assumed—in past tense, four (omitting second-hand) in present, with no evidentiality specification possible in future.

(There are yet no reported instances of Aspect > Evidentiality, Evidentiality > Tense, or Evidentiality > Aspect.)

III. There are a number of grammatical systems associated with an argument/NP. The person system—1st, 2nd, or 3rd—applies to arguments. Noun class applies to a noun which is head of an NP. Number is a referential category applying to an NP as a whole. Again we get a number of dependencies in each direction between these systems.

- Person > Number. More number distinctions are made in some persons than in others.
- Number > Person. More person distinctions are made in some numbers than in others.

These two dependencies were illustrated under (a) in §3.7 for English and Chipewyan respectively.

- Noun class > Number. The Australian language Ngandi has seven noun classes; a number distinction is made just for the two with animate reference (masculine and feminine).
- Number > Noun class. In Amharic a noun class (gender) distinction is made for 2nd and 3rd person in the singular, not in the plural.
- Person > Noun class. Many languages, including French and German, have a noun class or gender specification just in 3rd person.
- Noun class > Person. The contrast between 2nd and 3rd persons is neutralized in feminine gender and plural number for pronouns in Hebrew.

We have seen dependencies in each direction (in different languages) between the categories of set II—relating to predicate and clause—and between those of set III—relating to argument/NP. But when we examine pairs of systems from different sets, only unidirectional dependencies are found: clausal polarity in set I > system(s) from set II > system(s) from set III. These will now be briefly exemplified.

I, Polarity > II, Tense, Aspect, Evidentiality

There can be fewer choices in negative polarity than in positive for each of the systems from set II.

- Polarity > Tense. Illustrated for Amele at the beginning of this section.
- Polarity > Aspect. Aspectual distinctions are made only for positive clauses in Kresh, a Nilo-Saharan language spoken in the Sudan.
- Polarity > Evidentiality. In the Yuman language Maricopa an evidentiality specification cannot be made within the scope of the negative marker.

I, Polarity > III, Person, Noun Class, Number

There can be fewer choices in negative polarity than in positive for the systems from set III. For example, in both Tariana, from Amazonia, and Manambu, from New Guinea, all specifications of person, noun class (gender), and number within the verb are neutralized in a negative clause. The same applies for person and number in Estonian (a language with no noun classes).

II, Tense, Aspect, Evidentiality > III, Person, Noun Class, Number

- [Tense and Aspect] > Person. In the Balto-Finnic language Veps, verbs distinguish three persons and two numbers in present and simple past tense; in past perfect only number is distinguished, the three persons falling together.
- Evidentiality > [Person and Number]. Person and number are distinguished on the verb in Estonian only in a clause with non-reported evidentiality, not in one marked as reported.
- Tense > [Person and Noun Class (gender)]. For the Hebrew verb, person is only distinguished in non-present tenses, and gender is restricted to 2nd and 3rd person in non-present (it is marked for all types of subject in the present).
- [Tense and evidentiality] > Person. In Udmurt, a Uralic language, verbs show neutralization of 2nd and 3rd persons in past tense, eyewitness evidentiality, and positive polarity.

These provide just a sample of the unidirectional dependencies: set II systems > set III systems.

A clause includes a predicate; the choice of predicate determines the number and types of arguments. It is surely to be expected that grammatical systems associated with an argument—set III, Person, Noun Class, and Number—may be dependent on those associated with either predicate or clause—set II, Tense, Aspect, and Evidentiality; and that systems from both these sets may be dependent on the set I system, Polarity, which relates just to the clause. That is, the unidirectional dependencies we have noted correspond to the structural profile of a clause and its constituent elements, as set out in §3.5. (The system of mood—described under (b) in §3.2—relates to the clause. Preliminary study suggests that, as would be expected, systems from sets II and III may depend on mood, and not vice versa.)

Case is not a referential category, but rather a system which marks the function of an argument in its clause. We would expect to encounter instances where case choices may depend on those in any of the systems of sets I, II, and III. This is what is found. For example:

- Polarity > Case. In some Balto-Finnic languages (including Estonian) an NP in O function may in positive clauses be marked with either objective or partitive case—indicating ‘all the O’ and ‘some of the O’ respectively—but in negative clauses only partitive is available.
- Tense > Case. The Australian language Pitta-pitta has in a non-future clause distinct case marking for each of intransitive subject (S), transitive subject (A), transitive object (O), and indirect object. However, in a clause with future tense, there is one case suffix covering both S and A, and another for O and indirect object.
- Aspect > Case. Hindi has absolutive-ergative case inflection in perfective aspect, nominative-accusative elsewhere.
- Person > Case. A number of Australian languages have an absolutive-ergative system for 3rd person pronouns, nominative-accusative for 1st and 2nd person.
- Noun class (gender) > Case. Latin shows more case distinctions for masculine and feminine than for neuter nouns.
- Number > Case. This is also found in Latin, where there are more case distinctions for nouns in singular than in plural number.

Just one dependency has been noted in the opposite direction: Case > Number. In some languages, the number system only applies in certain cases; for example, only in the absolutive in Chukchi, and only in oblique cases (not in the nominative) in Kurdish.

3.20 Changing valency

As set out under (c) in §3.2, as a rule every clause—leaving aside copula constructions—is either intransitive, with one core argument (in S function), or transitive, with two core arguments (in A and O functions). Some verbs may be ambitransitive, occurring in each clause type (and are then either S = A or S = O ambitransitive). Other verbs—in some languages, all verbs—will be either strictly intransitive or strictly transitive.

Several of the most common kinds of derivation change the transitivity value of a verb, from intransitive root (or stem) to transitive stem, or vice versa. Since a transitive clause has two core arguments, there are two ways in which the sole argument of an intransitive clause can relate to one of the transitive core arguments. Set out most succinctly, we can have:

DETRANSITIVIZING either $O \rightarrow S$ (passive) or $A \rightarrow S$ (antipassive)
 TRANSITIVIZING either $S \rightarrow O$ (causative) or $S \rightarrow A$ (applicative)

These are basically syntactic processes; they are generally marked by a word-level derivation—which is typically a morphological process—applying to the

verb which is predicate head. Thus, for example, a derived passive construction will feature the derived passive form of a verb.

I. Detransitivizing derivations

There are four basic characteristics for prototypical passive and antipassive derivations.

Prototypical Passive Derivation (applying to a transitive clause)

- (a) Applies to an underlying transitive clause and forms a derived intransitive.
- (b) The underlying O becomes S of the passive.
- (c) The underlying A goes into a peripheral function, being marked by a non-core case, adposition, etc.; this argument can be omitted, although there is always the option of including it.
- (d) There is some explicit (that is, non-zero) formal marking of a passive construction; this can be a morphological process applying to the verb, or a periphrastic verbal construction (as in English, where it involves auxiliary verb *be*, plus suffix *-en* or *-ed* on the verb).

Prototypical Antipassive Derivation (applying to a transitive clause)

- (a) Applies to an underlying transitive clause and forms a derived intransitive.
- (b) The underlying A becomes S of the antipassive.
- (c) The underlying O goes into a peripheral function, being marked by a non-core case, adposition, etc.; this argument can be omitted, although there is always the option of including it.
- (d) There is some explicit formal marking of an antipassive construction (similar possibilities as for passive).

At first glance, the two derivations appear to be identical, save that A and O are interchanged. However, human language is never as glib as a series of algebraic-type formulas. In fact, passive and antipassive have quite different meanings and functions. (The name ‘antipassive’ is unfortunate, implying that this derivation is the opposite of passive. The term is now too well entrenched for it to be sensible to suggest a change; and it does have the advantage of being unambiguous.)

In a passive, the original A argument is downgraded in importance and, in consequence, the original O is brought into greater focus. Passive is often used when the underlying O is 1st or 2nd person, this then being taken into subject function; the passive sentence *I was arrested by a policeman* will often be preferred to the active one *A policeman arrested me*. Similarly if the O

argument has definite reference but the A is indefinite—*My pet spaniel was shot by a youth*, rather than *A youth shot my pet spaniel*.

Quite often the underlying A argument is obvious and need not be mentioned. This can be achieved by using a passive. For instance, *My wife was promoted to professor* (everybody knows who did it, the university promotions committee). Sometimes the speaker prefers not to mention the identity of the A argument. If a headmaster announces *I have been told that there has been marijuana-smoking in class four*, the pupils will say to each other, *Oh dear, he knows* (perhaps without wondering how he knows). Suppose that the headmaster had used an active clause, which requires something in A slot, and said *Someone has told me that there has been marijuana-smoking in class four*. The reaction to this might be *What fink told him?*

In the great majority of passive clauses the underlying A argument is not stated. But it must be possible to include it, with peripheral marking, for the construction to be a prototypical passive. If the underlying A cannot be stated, then we have an ‘agentless passive’.

Passive typically applies to a transitive clause. In some languages a passive derivation may also apply to an intransitive. For example, in English an NP which had been marked by a preposition may become passive subject in certain circumstances, as in *This glass has been drunk out of (by someone)*.

An underived clause is called ‘active’, with the active/passive contrast constituting a system of ‘voice’. The same label can, by extension, be applied to an active/antipassive contrast. (It can be confusing to also use ‘voice’ for causative, applicative, etc., as is occasionally done.) We can now illustrate with active and passive in English, and active and antipassive in Dyirbal (note that words can occur in any order in a clause in Dyirbal).

- (1) ACTIVE Johnny_A ate [the mango]_O
- (2) PASSIVE [The mango]_S was eat-en ([by Johnny])
- (3) ACTIVE Mayŋgu-ø_O Jani-ŋgu_A jaŋga-ñu
 mango-ABSOLUTIVE Johnny-ERGATIVE eat-PAST
 Johnny ate the mango
- (4) ANTIPASSIVE Jani-ø_S jaŋga-na-ñu
 Johnny-ABSOLUTIVE eat-ANTIPASSIVE-PAST
 (mayŋgu-gu)
 mango-DATIVE
 Johnny ate (the mango)

Just as the underlying A occurs as a peripheral argument of a passive, and can be omitted, so the underlying O occurs as a peripheral argument for an antipassive, and can be omitted. But note that the underlying O must

be statable in a prototypical antipassive. If it cannot be included, this is a ‘patientless antipassive’ (something which is rather rare).

In each language the active is the unmarked construction type, used in neutral circumstances. Pivots will be discussed in the next section. Basically, a language with an accusative-type pivot (involving S and A)—such as English—may need a passive derivation to put an O argument into pivot function. Similarly, a language with an ergative-type pivot (S and O)—such as Dyirbal—may need an antipassive to put an A argument into pivot function.

Some languages have just a passive derivation, some just an antipassive, with others showing neither. There may be a number of passives or antipassives (each with a different semantic profile). And some languages have a mixture of the two. For example, Mam, a Mayan language from Guatemala, has one antipassive derivation and four passives. One passive indicates that the underlying O suffered the action accidentally—‘José was hit (PASSIVE₁, accidentally) by Miguel’—and another that it was done on purpose—‘José was hit (PASSIVE₂, on purpose) by Miguel’.

Passive typically focuses on the resulting state; that is, the effect of the action on the patient. On hearing *John was wounded* one knows that he is in a state of being wounded. In contrast, antipassive focuses on the activity itself—that is, on the agent’s performing of the activity—effectively downgrading the underlying O. It is understood that the activity involves a patient, but little or no attention is paid to the identity of the patient.

In English, an intransitive clause must include statement of an S and a transitive clause of an A argument; but the O argument need not be stated with a number of verbs. One can say *Johnny has eaten lunch* (or *the mango*) or just *Johnny has eaten*. In an ergative language like Dyirbal, S and O arguments are obligatory and A optional. Sentence (3) could be shortened by omitting *Jani-ŋgu*, giving *Mayŋgu jaŋga-ŋu* ‘The mango has been eaten’. But the only way to say ‘Johnny has eaten’ is through an antipassive like (4), with demoted O omitted, *Jani jaŋga-na-ŋu*. That is, a passive focuses on the O argument and what happens to it (some A must be implied, but its identity may be unimportant). An antipassive focuses on the action and the A argument controlling it (some O must be implied, but its identity may not be of any consequence).

There will be a fuller discussion of passive and antipassive in Volume 3.

II. Transitivity derivations

A causative derivation is rather common, occurring in very many—but not in all—languages. The basic characteristics of a prototypical causative are:

Prototypical Causative Derivation (applying to an intransitive clause)

- (a) Applies to an underlying intransitive clause and forms a derived transitive.
- (b) The argument in underlying S function goes into O function in the causative.
- (c) A new argument (the causer) is introduced in A function.
- (d) There is some explicit formal marking of the causative construction. This may be a morphological process applying to the verb which is predicate head; or the language may employ periphrastic means (for example, the *make* construction in English, as in *She made him run*).

Causative applies to intransitive verbs in every language in which it is found, sometimes only to intransitives. In a number of languages it is used with just a few transitive verbs—typically ‘eat’ and ‘drink’—in other languages with many. There is considerable variation concerning what happens to underlying A and O in the causative of a transitive. The causer is always A. The original A may be the new O with the original O becoming a peripheral argument; or the original O may stay as is, with the original A becoming a peripheral argument. There are several other possibilities, all set out in Dixon (2000).

Some languages have several causative derivations, each with its own meaning. The contrasts may be direct or indirect causation, accidental or intentional, natural or with effort. The referent of the underlying A argument may be got to perform the action willingly or unwillingly. These and other semantic parameters are described and exemplified in Dixon (2000).

The basic characteristics of a prototypical applicative derivation are:

Prototypical Applicative Derivation (applying to an intransitive clause)

- (a) Applies to an underlying intransitive clause and forms a derived transitive.
- (b) The argument in underlying S function goes into A function in the applicative.
- (c) A peripheral argument (which could be explicitly stated in the underlying intransitive) is taken into the core, in O function.
- (d) There is some explicit formal marking of the applicative construction, generally by a morphological process applying to the verb which is predicate head.

There are significant differences from causative. Under (d), applicative is normally marked by a morphological process to the verb, whereas causative is often shown periphrastically. And in a causative the introduced argument (the

In some languages a single morphological process may mark causative with verbs of certain semantic types and applicative with verbs of other types. Where a language has distinct markings for passive, antipassive, causative, and applicative, it is sometimes the case that several of these derivations may apply, in sequence.

There are other processes that can change transitivity. When a noun is incorporated into a verb, the valency of the verb is sometimes reduced; see (1) in §3.13. And common ways of indicating that two core arguments have the same reference—reflexive—or are interchangeable—reciprocal—involve reducing valency; this is discussed in §3.22.

3.21 Topic and pivot

A unit of language activity may consist just of a sentence (which can be one clause or more). Or it may be a sequence of sentences, making up an utterance. Or a conglomerate of utterances (which may overlap) by a number of people taking part in a discourse. Such as the following, with speakers A, B, and C (here square brackets indicate constituents):

- (1) A Have you_A heard [the news [about John]]_O?
- (2) B ____A resigned [his job]_O
- (3) C ____S got sacked [by the old devil], if you ask me
- (4) B No, he_S really did resign
- (5) A ____{CS} always was a bit daft
- (6) C ____{CS} like his father

These utterances are linked by the fact that they are all talking about *John*, which is the ‘topic’ of this stretch of discourse. *John* is stated just in 1, and then in an NP marked by a preposition, *about John*, within the O NP, *the news about John*, of the verb *heard*. The topic is understood to be subject for each of (2–6), shown by pronoun *he* in (4) and left unstated (indicated by ___) in (2–3) and (5–6). The topic is in A function in (2), it is S of a derived *get*-passive construction in (3), and is in S in (4) and in CS (copula subject) function in (5–6). (In (6) the copula *is* is also omitted.) ((1–6) represent a typical discourse in colloquial spoken English. In the written mode either a pronoun (*he*) or *John* would have to be included in each of the ___ blanks.)

An argument of a clause is topic if it is coreferential with an argument of a clause which is immediately (or almost immediately) preceding or following. In some languages, any core argument may be topic in each clause. But a fair

number of languages have grammatical conditions on what may be topic. The term ‘pivot’ is used for a grammatically defined topic.

There are two common types of pivot:

I. S/A pivot, as in English. The topic should be in A or S (or CS) function in each clause in which it occurs. It may then be omitted from the second clause. Note that the underlying structure of (3) is [*The old devil*]_A *sacked* *John*_O. Here *John* is in O function, which does not satisfy the pivot constraint. One could not omit *John* from this; that is, speaker C could not have said [*The old devil*]_A *sacked* ___O, *if you ask me*. (*John* could, of course, be replaced with *him*, giving *The old devil sacked him, if you ask me*.) In order to be able to omit *John* from (3) the clause must be rephrased as a passive. We say that the passive ‘feeds’ the S/A pivot constraint in English, taking an argument which is in O (not a pivot function) and putting it in S (a pivot function).

II. S/O pivot. This can be illustrated for Dyrbal. Consider the following three simple sentences:

- (7) *Jani-∅_S* *bani-ñu*
 Johnny-ABSOLUTIVE come-PAST
 Johnny came
- (8) *Bili-∅_S* *bani-ñu*
 Billy-ABSOLUTIVE come-PAST
 Billy came
- (9) *Bili-∅_O* *Jani-ŋgu_A* *ŋarñja-ñu*
 Billy-ABSOLUTIVE Johnny-ERGATIVE watch-PAST
 Johnny watched Billy

Since the language has an S/O pivot, two clauses can be linked together if they have a common argument (the pivot) and this is in S or O function in each. Since *Bili* is in S function in (8) and in O function in (9), these two clauses may be coordinated, with the second occurrence of *Bili* omitted (note that in Dyrbal coordination is shown just by apposition of clauses, there being no linker ‘and’):

- (10) *Bili-∅* *bani-ñu*, *Jani-ŋgu_A* *ŋarñja-ñu*
 Billy-ABSOLUTIVE come-PAST Johnny-ERGATIVE watch-PAST
 Billy came and Johnny watched him

However, it is not possible to directly link (7) and (9). The shared argument, *Jani*, is in S (a pivot function) in (7) but in A (not a pivot function) in (9). Just as (3) had to be recast as a passive to satisfy the S/A pivot for English, so (9) has to be recast as an antipassive to meet the S/O pivot constraint in Dyrbal:

- (9ap) Jani- \emptyset_S η ar η ja-na- η nu (Bili-gu)
 Johnny-ABSOLUTIVE watch-ANTIPASSIVE-PAST Bili-DATIVE
 Johnny watched (Billy)

Now (7) and (9ap) can be joined together since they share an argument, *Jani*, which is in a pivot function, S, in both:

- (11) Jani- \emptyset bani- η nu η ar η ja-na- η nu
 Johnny-ABSOLUTIVE come-PAST watch-ANTIPASSIVE-PAST
 (Bili-gu)
 Bili-DATIVE
 Johnny came and watched (Billy)

Just as a passive may ‘feed’ an S/A pivot, so an antipassive is used to ‘feed’ an S/O pivot. Many languages with an S/A pivot have a passive derivation, and it is likely that every language with an S/O pivot has an antipassive. But note that passive and/or antipassive are found in languages which lack pivots, being then used solely for their semantic effects.

Many languages do not have a syntactic pivot, of either type. They have to infer the identity of an omitted argument from the semantics of that part of the discourse. For a sentence such as

- (12) Johnny_A hit Billy_O and ___S cried

a speaker of a pivotless language would infer that the unstated S argument for ‘cried’ is likely to be Billy, since someone who is hit is likely to be hurt, and people who are hurt do tend to cry. And for a sentence such as

- (13) Johnny_A hit Billy_O and ___S laughed

they would think that, since someone who has been hit is unlikely to laugh then, by elimination, Johnny is most probably the understood S argument for ‘laughed’.

Now in a language with an S/A pivot, such as English, the omitted S in both (12) and (13) must be Johnny, since an S argument may only be omitted if it is coreferential with S or A of the preceding clause. It may be unexpected for Johnny to cry after he has hit Billy, but this is what the English sentence *Johnny hit Billy and cried* states.

And in a language with an S/O pivot, the unstated S in (12) and (13) must be taken to be coreferential with the argument in pivot function, O, in the preceding clause. That is, for the translation of (12) and (13) into Dyirbal, it must be Billy who cried and laughed (however unexpected the laugh interpretation may be).

There are a number of different means which languages employ for marking an argument as topic. They include:

- Constituent order. When syntactic function is shown by something like case inflections, then the order in which constituents are placed may play a discourse role. Often, the argument which is topic comes first in the clause (sometimes it may appear last).
- There may be a special particle or clitic marking the topic.
- In some languages, the use of a bound pronoun is not obligatory; it may only occur when the argument in question is topic.
- The inclusion of a classifier (or of a noun class marker) in an NP may be a marker of topicality.
- Topic marking may be related to definiteness and/or to case. For example, an argument in S or A function might be marked with nominative case only when it is topic of its clause.
- An argument brought into S function by a passive or antipassive derivation, and one brought into O function by an applicative derivation, are likely to be topic.

Only when there is a grammatical constraint on the function of a topic NP do we use the label 'pivot'. Some languages have an S/A pivot, some an S/O pivot, and some combine the two varieties of pivot in different areas of grammar. As mentioned before, a language may have clear marking of what is a topic, without there being any pivot constraints.

Some languages have what is called 'switch-reference'. This involves a choice between two markings, typically on the verb. 'Same subject' indicates that this clause has identical S or A argument to the preceding clause, and 'different subject' states that their subjects differ. Whereas pivots can be of S/A or S/O variety, all instances of switch-reference marking relate only to S and A. No language is known which has both a pivot constraint on clause combining and also switch-reference marking.

There is also the concept of 'contrastive focus', sometimes confused with 'topic'. Whereas topic is a discourse strategy, serving to link together successive clauses, focus involves one argument (or the predicate) being accorded prominence within a single clause. For example:

- (14) (a) John_S got to the office late,
 (b) ____S did [his work]_O sloppily,
 (c) ____{CS} was rude [to the office boy],
 (d) then he_A insulted [*the boss*]_O,
 (e) and ____S went home early.

John is the topic of this sequence of five clauses. *The boss* is accorded stress and is thus accorded focus status in (14d). Note that the NP *the office boy* is not placed in focus; it is, perhaps, unexceptional to be rude to the office boy. But

one should be careful how one treats the boss, and the significance of the NP *the boss* as O for *insult* is brought out by putting it in focus.

Some of the methods summarized above for marking an argument as topic (constituent order, use of bound pronoun, inclusion of a classifier, etc.) may alternatively be used to indicate an NP in focus. They must be assessed for each grammar, on an individual basis.

There will be fuller discussion of pivots, switch-reference, and focus in Volume 3.

3.22 Argument identity: reflexives and reciprocals

A transitive clause has two core arguments, generally referring to different people or things, as *John hid the money*, and *Mary burnt Jane*. However, underlying A and O may have the same referent—this is a ‘reflexive’ construction. A language generally uses one of three techniques for dealing with this.

- I. Maintain the transitivity of the clause, and place a reflexive pronoun in O slot; this may be a bound form, for a language which has bound pronouns, or a free pronoun, as in English *John hid himself*, *Mary burnt herself*. In English the reflexive pronoun reflects the person, number, and gender of the subject. (Indeed, it shows number even when the subject pronoun doesn’t, for 2nd person, as in *I see you cut yourself* and *I see you cut yourselves*.) Some languages have a single reflexive pronoun for all persons, numbers, and genders—‘I cut self’, ‘John cut self’, etc.
- II. Derive an intransitive stem. Since A and O coincide in reference, there is in effect one argument. A morphological process is applied to the verb that derives a reflexive stem which is intransitive, and takes a core argument in S function (this is the underlying A = O). For example, in Dyrbal we get simple transitive clause (1) and derived intransitive reflexive (2).

(1) Mani- \emptyset _O Jani-ŋgu_A buyba-n
 money-ABSOLUTIVE Johnny-ERGATIVE hide-PAST
 Johnny hid the money

(2) Jani- \emptyset _S buyba-yirri-ñu
 Johnny-ABSOLUTIVE hide-REFLEXIVE-PAST
 Johnny hid himself

The suffix *-yirri-* derives an intransitive stem which has reflexive meaning. (Note that *-n* and *-ñu* are allomorphs of the past tense suffix.)

- III. Do nothing at all. Just state the A and O arguments in the normal way, saying *I hid me*, *You cut you*, *John burnt him*, and so on. This works well

for 1st and 2nd person pronouns, but may lead to ambiguity in the case of 3rd person. Thus, in *John burnt him*, the *him* may refer back to John or to someone else. Some languages have a reflexive pronoun just for 3rd person. Others have to bring in something like a relative clause to resolve the ambiguity (if, indeed, this is not achieved by the discourse context). For example, in Fijian one can say, literally ‘John, who relates to himself, cuts him’, which provides a reflexive meaning.

If we have two clauses with the same two predicates interchanging A and O functions—*John hit Bill* and *Bill hit John*—a ‘reciprocal’ construction can be formed from them. Languages employ techniques which are similar to the first two for reflexives.

- I. Maintain transitivity, and place a (bound or free) reciprocal pronoun in O slot; for example *John and Bill hit each other/one another* in English.
- II. Derive an intransitive stem, with reciprocal meaning, by applying a morphological process to the verb. It takes an S core argument, which is the coordination of the underlying A and O. For example, in Dyirbal we find:

- (3) Jani- \emptyset _O Bili- η gu_A bara-n
 Johnny-ABSOLUTIVE Billy-ERGATIVE punch-PAST
 Billy punched Johnny
- (4) Bili- \emptyset _O Jani- η gu_A bara-n
 Billy-ABSOLUTIVE Johnny-ERGATIVE punch-PAST
 Johnny punched Billy
- (5) [Jani-garra- \emptyset Bili-garra- \emptyset]_S
 Johnny-ONE.OF.A.PAIR-ABSOLUTIVE Billy-ONE.OF.A.PAIR-ABSOLUTIVE
 baral-baral-nbarri- η nu
 REDUP-punch-RECIPROCAL-PAST
 Johnny and Billy punched each other

Suffix *-nbarri-*, plus reduplication of the root (a double-barrelled morphological process), derives an intransitive stem, *baral-baral-nbarri*, which takes an S argument, the NP *Jani-garra Bili-garra*—literally ‘Johnny, being one of a pair, and Billy, being one of a pair’ (that is ‘Johnny and Billy’).

A reciprocal construction is not confined to a situation of just two participants. One can say *The boys punched each other*, where there is an indeterminate number of boys. This does not imply that every boy punched and was punched by every other boy. One would infer that most of the boys punched

some other boys, and that most of them were punched by some other boys, or something along these lines.

The majority of languages use similar techniques for reflexive and reciprocal—maintaining transitivity with reflexive and reciprocal pronouns, as in English, or deriving intransitive stems, as in Dyirbal. Sometimes reflexive and reciprocal are marked identically, the construction being taken to be reflexive when the subject is singular—as in (2)—and reciprocal when the subject is plural—as in (5). In a number of languages, a reciprocal marker is based on the reflexive (never the other way round). For example, in Korafe, from New Guinea, the reflexive pronoun is *tofo* and the reciprocal one *tofotofo*.

There are, however, some languages which employ different techniques. For example, Swahili has a reflexive bound pronoun in the prefix slot reserved for an O argument, but an intransitivizing derivational suffix *-na-* for reciprocal.

Volume 3 will deal in more detail with reflexive and reciprocal constructions, including discussion of further grammatical mechanisms. And it will describe some of the secondary senses of reflexive and reciprocal pronouns and processes (such as that of *myself* in *I myself hid the money* in English).

3.23 Comparative constructions

A comparative construction is often a rather unusual feature of the grammar, for languages in which it occurs.

There are three basic elements in a prototypical comparative scheme: the two participants being compared, and the property in terms of which they are compared. Consider the English sentence:

- (1) John_{CS} is [more handsome]_{CC} [than Felix]
 COMPAREE INDEX PARAMETER MARK STANDARD

The participants are:

- COMPAREE—that which is being compared, here *John*.
- STANDARD of comparison—what the comparee is being compared against, here *Felix*.

The property is:

- PARAMETER of comparison—here *handsome*. The parameter is typically an adjective, but in some languages it may also be a verb and/or an adverb and/or a noun and/or a time word.

A prototypical comparative scheme will generally (but not invariably) also include:

- INDEX of comparison—here *more* (with a different choice of English adjective, it could have been *-er*, for example *tall-er*; see §4.8).

Within any clause, there must be some marking of the function of each core and peripheral argument. The comparee is always some kind of subject; in (1) it is copula subject. The standard has a wider range of functions; it may be in O function or it may be a peripheral argument, as in (1). We then have:

- MARK of the grammatical function of the standard—here *than*.

There are two main varieties of comparative construction.

- I. In a language where an adjective may be head of an intransitive predicate, the parameter has this function, and is modified by the index of comparison. For example, in Indonesian:

(2)	COMPAREE	INDEX	PARAMETER
	Dias	[lebih	tinggi]INTRANSITIVE.PREDICATE
	3sg	MORE	be.tall

MARK STANDARD

[dari saya]

FROM 1sg

He is taller than me

- II. In a language where an adjective may not function as intransitive predicate, it is likely to be employed as copula complement, and in this function will act as parameter (modified by the index) in a comparative construction, as in (1) from English.

There are various possibilities for the index and mark:

- The index of comparison may be a special form used only in this function (as *more* and *-er* in English), or it may be a general modifier such as ‘very’. Or there may be no index stated, the other features of the construction indicating that it describes comparison of the ‘more’ variety.
- The mark of the standard may be a special form, such as *than* in English. But it is most often a marker which has some other function in the grammar. Many languages are like Indonesian, in (2), using ablative (whose main meaning is to indicate ‘from’ a place); others employ dative, locative, ‘on’, ‘upon’, genitive, and ‘in front of’.

Other types of comparative construction are found. These include

- A verb like ‘surpass’ (as index) combined with an adjective or verb (as parameter) in a serial verb construction. For example ‘John_A

[be.handsome surpass]_{SERIAL.VERB.PREDICATE} Bill_O' for 'John is more handsome than Bill'.

- Having the parameter nominalized, saying '[John's strength]_A exceeds [Tom's strength]_O' for 'John is stronger than Tom'.

The prototypical comparative construction compares two participants in terms of the degree of some gradable property relating to them, as in (1) and (2). A non-prototypical construction involves the comparison of two properties in relation to one participant; for example, *This box is longer than it is wide*. Or one can, in a few languages, compare clauses; an extreme example in English might be *John loves his grandmother more than Mary hates football*.

Whereas the great majority of languages have a relative clause construction, a causative construction, and so on, prototypical comparative constructions are lacking from a sizeable number of languages. Non-prototypical constructions, such as those illustrated in the last paragraph, are rarer still.

Those languages which lack a comparative construction per se generally have some strategy for making comparative judgements, involving two clauses in apposition. To express the sentiment 'John is more handsome than Felix', one may have to say something like 'John is handsome; Felix is ugly' or 'Felix is handsome; John is very handsome.'

Volume 3 will provide a fuller treatment of comparative constructions—those mentioned here and others besides—plus discussion of superlatives; and see Dixon (2008).

Sources and notes

3.2. Tongan from Churchward (1953) and Malenaite Taumoefolau (personal communication). For Trumai see Guirardello (1992, 1999). Ainu from Tamura (2000).

3.5. Mandarin Chinese from Zhu Deni (personal communication). Nootka from Swadesh (1938: 78) and Jacobsen (1979: 85).

3.7. Chipewyan from Cook (2004: 94). Kayardild from Evans (1995: 202). Hdi from Frajzyngier (2002: 83). Jacaltec from Day (1973: 93). (Note that Craig 1977 provides different forms.) Yidiñ from Dixon (1977a: 182.)

3.9. The 'nominal hierarchy' is an extension of work by Silverstein (1976).

3.11. 'Sentence' in Jarawara, see Dixon (2004a: 530), in Fijian see Dixon (1988: 257–8).

3.12. Quileute from Andrade (1933: 268).

3.13. The first *OED* citation for 'syntax' in its linguistic sense is Cawdrey (1613), and for 'allomorph' is Nida (1948). (This is probably by analogy with

‘allophone’, which the *OED* has as introduced by Whorf in 1938.) The *OED* has ‘morphology’ used in its linguistic sense in English from 1870, doubtless based on ‘Morphologie’ in German which the *OED* says was introduced in 1859 by August Schleicher, a linguist who advocated a ‘biological approach’ to language. Stankiewicz (1972: 24) has ‘phoneme’ introduced in 1881 by Kruszewski (a student of Baudouin de Courtenay) and ‘morpheme’ introduced in the same year by Baudouin de Courtenay himself.

Most of the examples of reduplication are from Harold Key (1965). Amharic from Amberber (1995). Further examples are in Hurch (2005). Tone shift in Anywa from Reh (1996: 45). Infixes in Dakota from Rankin et al. (2002). Egyptian Colloquial Arabic from Mitchell (1962: 36, 72–3).

Discussion of ‘Item-and-Arrangement’ and ‘Item-and-Process’ is in Hockett (1954). Full details on phonological rules in Yidiñ are in Dixon (1977a, 1977b).

In recent years some linguists have utilized ‘parsing programs’ to analyse texts on a computer. These operate on an ‘Item-and-Arrangement’ principle and may encounter severe difficulties when faced with morphological processes other than compounding and affixation.

3.14. Lezgian from Haspelmath (1993: 74). For ‘associated motion’ in Australian languages, see Dixon (2002: 201–2). Koasati from Kimball (1991: 116–26). The fullest and most informative discussion of derivation is Aikhenvald (2007).

3.15. Aikhenvald (2004) provides an inclusive cross-linguistic account of evidentiality.

3.16. Corbett (1991) unhelpfully uses the label ‘gender’ to refer to all kinds of noun classes. However, Aikhenvald (2000)—the seminal text on this topic—follows the established naming practice, as outlined here.

3.17. Axininca Campa from Payne (1981). Comanche from Charney (1991: 114). Sumerian from Thomsen (1984: 135).

3.18. Tongan from Churchward (1953: 6–7, 25). Taumoefolau (2002) suggests that this ‘stress shift’ actually involves repetition of the final vowel—*afɔ* becomes *afɔɔ* and it is the long final vowel which attracts stress; see also Anderson and Otsuka (2006). Gola from Westermann (1947: 17) and Heine (1982: 193). Ngiyambaa from Dixon (1980: 365–6), based on Donaldson (1980: 128–9, and personal communication).

In an examination of English texts, I found that about half of the instances of demonstratives as determiners may be replaced by *the*; however, only a rather small number—fewer than 5 per cent—of instances of *the* could be replaced with a demonstrative.

3.19. The information in this section is, essentially, a condensation of Aikhenvald and Dixon (1998). That paper has fuller discussion, further exemplification, and greater detail for the examples briefly quoted above. (It also investigates possible dependencies which relate to definiteness.)

Amele from Roberts (1987: 110, 223–6). Yimas from Foley (1991: 241 ff.). Tuyuca from Barnes (1984). Tariana from Aikhenvald (2003: 400–1). Manambu from Aikhenvald (2008: Chapter 14). Ngandi from Heath (1978: 35). Amharic from Leslau (1995: 466 ff.). Kresh from Brown (1994: 165–6). Maricopa from Gordon (1986a: 85). Veps from Laanest (1975: 91). Udmurt from Tepljashina and Lytkin (1976: 179). Pitta-pitta from Roth (1897: 7 ff.). Chukchi from Skorik (1961) and Dunn (1999). Kurdish from Bakaev (1966: 263).

3.20. Mam from England (1983).

3.23. Indonesian from Sneddon (1996: 179).

Analysis, Argumentation, and Explanation

4.1 What is linguistic analysis?

Basic linguistic theory provides a flexible, analytic framework in terms of which the grammar of each individual language can be formulated. It furnishes an array of grammatical categories and construction types—together with varieties of interrelations between them—from which appropriate choices are made.

There are prototypical plans of organization for a natural language, as surveyed in the previous chapter. Some kinds of variation from the prototypical schemas occur in relatively few languages; for example, a fairly small number of languages lack a tense system, or a complement clause construction. Other variations are more common—only around one-quarter of the world's languages include an evidentiality system in their grammar. No variation from the prototypical plans can be excluded, as a possibility.

Although it has been suggested that some languages lack a distinction between noun and verb, it will be shown in Chapter 11 that—for every language which has been thoroughly analysed—distinct noun and verb classes can be recognized. However, we cannot on a priori grounds deny the possibility of there being a language for which this distinction could not be made. In similar fashion, all languages we know of have transitive and intransitive clause types. It is unlikely that any language should lack this distinction, but this (or any other possibility) cannot be absolutely excluded.

That is, an individual language may differ in any particular way from the recurrent schemas outlined in Chapter 3. It may differ in several ways. But no natural language will differ from these templates in many ways (let alone, in all possible ways).

In its fundamental approach, basic linguistic theory—the paradigm of linguistics as a branch of natural science—differs from every formal theory. These each put forward a fixed framework, so that their practitioners simply have to match up an individual language with the elements of the theory. (It is a little like completing a rather involved questionnaire.) Every language,

one formal theory will aver, operates with a unit ‘verb phrase’ (this essentially consists of verb plus direct object). Look, and it will be found. Complement clauses may be regarded as a universal feature of language structure. All the practitioner has to do is decide which construction type should be called ‘complement clause’ in a particular language.

A scientific linguist, working within basic linguistic theory, examines their textual corpus to see if there is anything which could be analysed as a complement clause construction; if so, they provide argumentation in support of this analysis. As set out in Chapter 18, there are certain general characteristics of complement clauses. Not all languages show all of them, but a selection must be satisfied for a construction to be appropriately recognized.

The term ‘analysis’ is thus used in two quite different manners. Within a formal theory, ‘analysis’ means fitting a language into a pre-existing formal matrix. All languages have X, Y, and Z. Thus, what are X, Y, and Z in your language of study? This is what they term ‘analysis’.

Basic linguistic theory does not assert that any particular feature is present in every language. Instead, it provides a range of linguistic elements and parameters, which are available to be drawn on, as appropriate, in the formulation of the grammar of a language. A scientific linguist analyses the data they have collected. They do of course look for examples of familiar categories and construction types, but do not feel bound to recognize these. The construction types of a newly described language will be studied and their properties examined. There may—or may not—be one which satisfies the criterion to be called a ‘complement clause’; that is, it should show significant similarities to what have been called complement clauses in other languages. This recognition, and the detailed argumentation associated with it, are what ‘analysis’ involves for a linguist who works within basic linguistic theory.

A formal linguist may *assume* that a new language has a complement clause construction, and *seek* it out. A scientific linguist must *justify*, by *argumentation* in terms of the internal grammatical structure of the language in question, the recognition of a clause type which can fill a core argument slot in clause structure.

Method of investigation tends to correlate with academic stance. Formalists prefer elicitation. They say this is the only way they can get what they want (and they do know what they want.) Ask the informant to translate from the lingua franca (for example, English or Spanish) a few sentences which include complement clauses. What do you get? Well, from my observations, one would be unlikely in this way to obtain complement clause constructions in the language of study. (Although the language may well have this construction type, which will occur in texts and in conversation.)

A scientific linguist will assemble a wide database, spanning texts in a variety of genres plus information gathered from participant observation. They analyse the language as it actually is, with a minimum of preconceived ideas about what it should be. Working in terms of basic linguistic theory, one analyses a language as a system in its own right, every part relating to the whole. There will be familiar features but they may be linked together in unusual ways. And there may be some unusual aspect of an established category, or else some entirely new feature.

In contrast to the deductive stance of formal theories, basic linguistic theory is an inductive pursuit, providing a set of parameters which are suggestive but malleable. Every description of a language in terms of basic linguistic theory provides feedback to the theory, enabling it to be refined and extended. Basic linguistic theory provides an integrated scheme of what is found across all manner of natural languages. Nothing is absolutely required to be included, and anything is acceptable, so long as it is arrived at by a principled scientific analysis, supported by cogent argumentation.

This chapter will outline, and briefly exemplify, the nature of linguistic analysis and argumentation. How to enunciate a problem of analysis, put forward alternative solutions to it, consider the pros and cons of each, and choose between them. How different analyses may be suitable for different purposes. And then, the kinds of synchronic and diachronic explanation which can be provided for the structural features of a language.

As in Chapter 3, many of the illustrations below are from my own fieldwork on a variety of languages. In each case I have published a full grammar, so that the reader can—if they wish—examine how a particular feature fits into the interlocking grammatical scheme of the language.

4.2 Analyses should be clear and plausible

There is always a choice between several possible analyses for each part of a grammatical description. The analysis to be preferred is that which is maximally simple and sensible.

Consider plural marking on English count nouns, exemplified by:

dog-s /dɒg-z/

cat-s /kæt-s/

horse-s /hɔ:s-əz/

We can identify two ways of describing the varying forms of plural.

A

- basic form /-əz/

- after a stem which does not end with a sibilant, suffix /-əz/ omits the /ə/ and reduces to be just /-z/
- this reduced form /-z/ assimilates in voice to a preceding stem-final consonant

B

- basic form /-z/
- since English does not allow a sequence of two sibilants, insert /ə/ between a stem-final sibilant and /z/
- /z/ assimilates in voicing, becoming /s/ after a voiceless stem-final consonant

Analysis B provides an *explanation* for the varying allomorphs. Since sequences /sz/, /zz/, /ʃz/, and /ʒz/ do not occur in English, such a putative sequence is broken up by the insertion of the neutral vowel /ə/.

The statement in Analysis A, that /-əz/ reduces to /-z/ except after a sibilant, appears ad hoc and without explanation. If the basic form is /-əz/, there is no reason why this should not be retained after all consonants.

It can be seen that Analysis B is simple, perspicuous, and principled. It is plainly to be preferred over Analysis A.

We may consider one more example, also from English, concerning the derivation of adjectives from some names for places and people. Consider, as a sample of larger sets of derivations:

- (i) *Tibetan*, /ti'betən/; *Lutheran*, /'lu:θərən/
 (ii) *Kenyan*, /'kɛnjən/; *Corsican*, /'kɔ:sikən

The first analysis is taken from a standard text.

- A** In (ii) 'we have the not uncommon problem of an indeterminate boundary between base and suffix: *Kenyan* derives from *Kenya* + *-an*, with reduction of *a* + *a* to a single *a* which cannot be assigned uniquely to base or affix' (Bauer and Huddleston 2002: 1691).

This analysis is:

- the suffix has a single form, *-an*
- after a stem ending in *a*, we get *-a* + *an* → *-an*, and can't tell which *a* drops and which is retained.

Note that this analysis is based on orthographic forms, although English orthography is far from providing a one-to-one reflection of phonology. In fact, all the *a*'s mentioned here are the central vowel schwa, /ə/.

- B** Say each root has a fixed form, but the suffix has phonologically conditioned allomorphs:

- *-an, /-ən/* after a root ending in a consonant
- *-n, /-n/* after a root ending in the unstressed central vowel /ə/

It is a cross-linguistic feature of human languages that affixes typically have varying forms in different phonological and/or morphological environments. It is rather unusual (although not absolutely unheard of) to assign a root different forms in different environments; for example, saying that *Kenya* /'kenjə/ takes on reduced form /kenj/ before suffix /-ən/. Analysis B is plainly to be preferred to such an analysis.

Analysis B is also to be preferred over A, which demands invariable form for root and suffix and then a reduction rule, creating a worry for Bauer and Huddleston as to which *a* is kept and which is lost. Analysis B is simple, unequivocal, and natural.

4.3 Argumentation needed to justify an analysis

Earlier chapters have included a number of examples of linguistic analysis and the associated argumentation. In §1.8 it was stressed that one cannot assume any language has a familiar set of word classes (say, noun, verb, and adjective). Word classes must be recognized on language-internal criteria; they may then be identified between languages in terms of similar syntactic function and semantic content. In discussion of the order of presentation when writing up a grammar, in §2.1, argumentation was presented for analysing a verb with incorporated object in Fijian as intransitive, in terms of the possible placement of adverbs. Three further examples will now be given, from Fijian, Dyirbal, and English.

Relative clauses in Fijian

When I was on fieldwork on the island of Taveuni, studying the Boumaa dialect of Fijian, a particular problem which arose was whether or not a relative clause construction should be recognized. In one story, a Catholic priest offers a cross to a local chief to help win a forthcoming war. The priest says:

- (1) (a) mo_A rai-ca [a kuruse]_O
 IMPERATIVE+2sg see-TRANSITIVE+3sgO ARTICLE cross
 You look at the cross!
- (b) au_A saa tara-a
 1sg ASPECT hold-TRANSITIVE+3sgO
 I hold it

There are two ways of analysing this sequence of clauses:

- A As a coordination: ‘You look at the cross! I am holding it.’ (Note that in Fijian clausal coordination just involves juxtaposition; there is no overt clausal coordinator similar to English *and*.)
- B With (b) as a relative clause to the O NP in (a), *a kuruse*: ‘You look at the cross which I am holding.’

If there were no argumentation to support Analysis B, then A should be followed. Not all languages have a canonical relative clause construction, but no language thus far known lacks simple clause chaining (often shown just by clausal apposition).

Within the typological framework of basic linguistic theory, a relative clause shows one or more of the following properties:

- (i) some formal grammatical marking such as a relative pronoun or a special inflection on the verb of the relative clause;
- (ii) a definite position in the main clause (for example, it may follow the noun it is modifying, or it may always come at the end of the main clause);
- (iii) a special intonation pattern;
- (iv) semantic interpretation appropriate to a modifier-head relation.

Fijian lacks (i). In relation to (ii), I had assembled a corpus of several dozen constructions similar to (1) and in each instance the putative relative clause came last. This might be the positioning of a relative clause in Fijian, and in that case it would not help to distinguish between coordination and relative clause analyses. Considering property (iii), it did seem to me that these possible relative clause constructions had a special intonation pattern, but it was hard to be certain.

The critical evidence turned out to be semantic. Consider the following, from a story about life in olden times:

- (2) (a) e_S lailai [a 'e-dra 'a'ana]_S
 3sg be.little ARTICLE EDIBLE-3plPOSS food
 their food was little
- (b) e_S saga
 3sg be.boiled
 it was boiled

Under the coordination analysis, A, this would mean ‘They had little food (lit. their food was little) and it was boiled’. Under the relative clause analysis, B, it would mean ‘Little of their food was boiled (lit. their food which was boiled was little)’. In fact, the latter reading is the correct one. The storyteller’s

ancestors did have plenty of food, but most of it was roasted since there were then rather few cooking pots which could be used for boiling.

About one-third of the examples I had gathered from texts of possible relative clauses were like (2), where an analysis as relative clause was necessary to give the correct semantic meaning. Together with evidence from intonation, this confirmed that a relative clause construction should be recognized for Fijian, even though it has no formal grammatical marking. (There are some other languages which also lack such a marking.) Further discussion of this is in §17.4.

Ergative and instrumental cases in Dyirbal

In Dyirbal, an NP in A function must be marked with ergative inflection, which is *-ŋgu* after a disyllabic root ending in a vowel, as in

- (3) *yara*_O *yibi-ŋgu*_A *balga-n*
 man woman-ERGATIVE hit-PAST
 The woman hit the man

A noun may take an instrumental case ending and this has identical form to ergative. The noun *yugu* ‘stick’, with instrumental ending *-ŋgu*, can be added to (3):

- (4) *yara*_O *yibi-ŋgu*_A *yugu-ŋgu* *balga-n*
 man woman-ERGATIVE stick-INSTRUMENTAL hit-PAST
 The woman hit the man with a stick

(As mentioned in §2.4, both phrasal constituent order and word order are quite free in Dyirbal; the words in these and other sentences may occur in any sequence.)

The analytic question is: can we justify recognizing ergative and instrumental as distinct inflections? Should they not be regarded as a single case? Why not say that in (4) the NP in A function consists of two nouns, *yibi* and *yugu*, both taking ergative-instrumental suffix *-ŋgu*. *Yibi*, which has animate reference, then indicates the agent, and *yugu*, with inanimate reference, refers to the instrument used by the agent.

This would be the fallback analysis, if no evidence were available to distinguish ergative from instrumental. In fact there is such evidence, relating to the syntax of the language. First, there is an antipassive derivation which yields an intransitive clause. The underlying A argument goes into S function, the underlying O now takes dative case, and the instrument stays as is. The verb is marked with an antipassive derivational suffix, *-ŋa-*, between root and tense inflection. Thus, corresponding to (4), there is the antipassive (with essentially the same meaning as (4)):

- (5) *yibi*_S *yugu-ŋgu* *balgal-ŋa-ñu* *yara-gu*
 woman stick-INSTRUMENTAL hit-ANTIPASSIVE-PAST man-DATIVE
 The woman hit the man with a stick

Secondly, there is an applicative derivation; this maintains transitivity but rearranges syntactic arguments. The NP originally marked with instrumental case goes into O function, the original O takes dative inflection, and the original A remains as is (still bearing ergative case). The verb takes applicative derivational suffix, *-ma-*, between root and inflection. The applicative corresponding to (4) is:

- (6) *yugu*_O *yibi-ŋgu*_A *balgal-ma-n* *yara-gu*
 stick woman-ERGATIVE hit-APPLICATIVE-PAST man-DATIVE
 The woman used a stick to hit the man (lit. the woman hit a stick to the man)

We see that ergative and instrumental must be regarded as distinct cases in terms of their syntactic behaviour, although they receive the same morphological realization. In the antipassive derivation, an NP in A function (marked by ergative) goes into S function (this is absolutive case, which has zero realization) while instrumental remains unchanged. In the applicative derivation, an NP in instrumental case goes into O function (also absolutive case) while the A argument retains its ergative marking.

Embedding in English

The three basic methods of clause combining are set out in Figure 3.1 of §3.2. Two involve ‘embedding’, the inclusion of one clause within another. These are (i) where a relative clause is included within an NP that is an argument of a higher clause; and (ii) where a complement clause fills an argument slot in a higher clause.

Some linguists have suggested that temporal clauses are also embedded, effectively filling a peripheral slot in the structure of a higher clause. This suggestion can be diagrammed:

- (i)
-
- ```

graph TD
 clause --- S_argument["S argument
Mary"]
 clause --- predicate["predicate
sang"]
 clause --- peripheral_argument["peripheral argument
Δ
after she had locked the doors"]

```

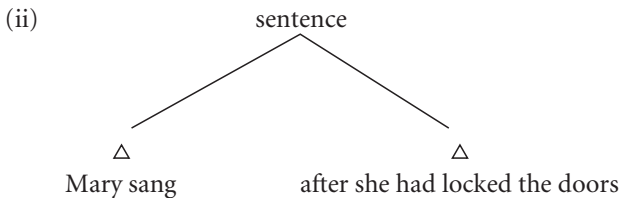
Can such an analysis be justified? In investigating this, we can consider the following clause combinations:

- (7) (a) Mary sang, after locking the doors  
 (b) Mary sang, after she had locked the doors  
 (c) Mary sang, when she had locked the doors
- (8) (a) The cat purrs, when you tickle its ears  
 (b) The cat purrs, if you tickle its ears

Note that in each of these sentences the clauses can occur in either order.

Although the temporal clauses have a different structure in (7a) and (7b–c), the relation of these clauses to *Mary sang* is essentially the same, so the same grammatical description should be provided for them. (8a) and (8b) have very similar meaning, and should certainly be dealt with in the same way within the grammar. In the conditional construction, (8b), *if you tickle its ears* could not possibly be embedded within *The cat purrs*; there is a clear logical relationship between the *if* clause and the main clause. It follows that an embedding analysis, as in (i), is also inappropriate for (8a) and then also for the *when* clause in (7c) and the *after* clauses in (7a–b). (And also for temporal clauses introduced by *before*, *since*, *until*, *while*, etc.)

That is, the analysis in (i) must be put aside in favour of:



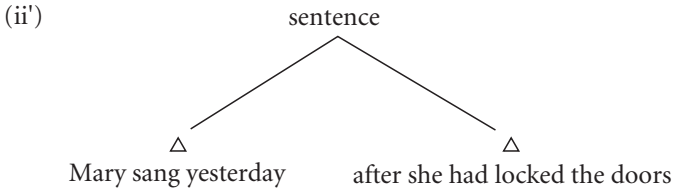
Why, one might enquire, should the analysis in (i) have been suggested? Perhaps by relating temporal clauses to temporal adverbs (which can be a word, such as *yesterday*, or a phrase, such as *in the morning* or *at ten o'clock*). For example:

- (9) Mary sang yesterday

It might be thought that *after she had locked the doors*, in (7b), fills the same slot in clause structure as *yesterday* in (9). It does not—both temporal elements may be included in a sentence:

- (10) Mary sang yesterday after she had locked the doors

This has the structure:



*Yesterday* is a peripheral argument within the main clause, whereas *after she had locked the doors* is a temporal clause linked to *Mary sang yesterday* at sentence level.

It can be seen that it is Figure (ii), not Figure (i), which provides the most appropriate analysis for sentence (7b), as for (7a), (7c), and (8a–b).

#### 4.4 Choosing between alternative analyses

It is not always easy to decide between alternative analyses. This can be illustrated for number words in Fijian.

The basic syntactic functions for verbs, numbers, adjectives, and nouns are summarized (in slightly simplified form) in Table 4.1. Almost all verbs may be head of either an intransitive or a transitive predicate. Adjectives freely function as head of an intransitive predicate, as do nouns, although they occur in this function relatively rarely. A noun may be head of an NP and an adjective may directly modify an NP head, occurring immediately after it. A verb may modify a noun only through a relative clause.

There is a small set of ‘number words’, comprising the numerals ‘one’ to ‘ten’, plus ‘many’, ‘a few’, ‘some of’, and ‘how much/how many/some (indefinite)’; their main syntactic properties are shown in Table 4.1. (Numbers ‘a hundred’ and ‘a thousand’ behave rather differently, and may be best treated as nouns.)

The number set behaves in several ways like intransitive verbs. Compare (1), which has verb *tagi* ‘cry, weep’ as predicate head, and (2), with number *rua* ‘two’ in this slot:

TABLE 4.1. Main syntactic possibilities for word classes in Fijian

|                                   | Verb | Number | Adjective | Noun |
|-----------------------------------|------|--------|-----------|------|
| 1. Head of intransitive predicate | ✓    | ✓      | ✓         | (✓)  |
| 2. Head of transitive predicate   | ✓    | –      | –         | –    |
| 3. Head of NP                     | –    | –      | –         | ✓    |
| 4. Modifier within NP             | –    | [✓]    | ✓         | –    |



- (1) [e tagi]PREDICATE [a gone yai]NP:S  
 3sgS cry ARTICLE child HERE  
 The child here is crying

- (2) [e rua]PREDICATE [a gone yai]NP:S  
 3sgS two ARTICLE child HERE  
 There are two children here (lit. The child(ren) here are two)

An adjective—such as *caa* ‘bad’—may also function as head of an intransitive predicate:

- (3) [e caa]PREDICATE [a gone yai]NP:S  
 3sgS bad ARTICLE child HERE  
 The child here is bad

Only adjectives may directly modify a noun within an NP. For example:

- (4) [e la'o]PREDICATE [a gone caa]NP:S  
 3sgS go ARTICLE child bad  
 The bad child is going

A verb may only modify a noun through a relative clause, as in

- (5) [e [la'o]PREDICATE [a gone [e  
 3sgS go ARTICLE child 3sgS  
 tagi]RELATIVE CLAUSE]NP:S  
 cry  
 The child who is crying is going

Numbers are basically like verbs, modifying nouns only through a relative clause:

- (6) [erau la'o]PREDICATE [a gone [e  
 3duS go ARTICLE child 3sgS  
 rua]RELATIVE CLAUSE]NP:S  
 two  
 Two children are going (lit. the child(ren) who are two are going)

Note that there is a dual subject pronoun, *erau*, at the beginning of the predicate in (6), agreeing in number with the NP in S function.

Although (6) is a possible and occurring sentence, an alternative is generally preferred, in which *e rua* precedes *a gone*:

- (7) [erau la'o]PREDICATE [e rua a gone]NP:S  
 Two children are going

Note that a relative clause whose predicate head is a verb could *not* be fronted within an NP; instead of (5) one could *not* say \**e la'o e tagi a gone*.

Grammarians of Fijian have wondered whether the *e* before *rua* in (7) should be regarded as a special variety of article which occurs only before numbers. On balance, this seems unnecessary. (It would also be typologically most unusual.) One can treat the *e* as the regular 3sg subject pronoun and state that in (7) the relative clause comes at the beginning of the NP; this only happens when the relative clause involves a number word.

However, there is a small set of eight nouns referring to time and distance that may be directly modified by a number word, which here behaves like an adjective. *Maacawa* is 'week' and for 'two weeks' one can use either the normal number word construction *e rua a maacawa*, or instead just say *a maacawa rua*.

It will be seen that the set of number words is like verbs in some ways—being more similar to verbs than to adjectives or nouns—but differs in other ways. Should we say that number words constitute a separate word class, or that they are a subset of verbs?

There are a number of further differences. Unlike nouns and adjectives, both verbs and number words undergo productive reduplication, but with varying semantic effect. A fully reduplicated verb marks 'do several times' or 'do over a long period'. In contrast, when a number word is reduplicated it forms an adverb 'all X', which comes at the end of a predicate. From *rua* is derived *ruarua* 'both' as in:

- (8) [erau la'o ruarua]<sub>PREDICATE</sub> [a gone]<sub>NP:S</sub>  
       3duS go both                     ARTICLE child  
       The two children are both going

Other properties are specific to number words. They can form ordinal numbers (which are derived nouns) by prefixing *i'a-*; thus *i'a-rua* 'the second one'. And they can form distributives by prefixing *yaa-*; for example, *yaa-rua* 'two each' (as in 'The children have two books each').

On balance, it seems to me most appropriate to recognize number words as a separate small word class, with some similarities to verbs. An alternative analysis would be to treat them as a subset of verbs with some rather divergent properties. So long as all grammatical properties are fully described—and, wherever possible, explained—either analysis would be acceptable. It is certainly not the case that one analysis should be labelled 'good' or 'correct' and the other 'bad' or 'wrong'.

## 4.5 Different analyses for varying purposes

As mentioned in §1.8 (and further justified in Chapter 12) it is likely that an adjective class may be recognized for every language. In some languages (such as Latin), adjectives have very similar properties to nouns, but there are still sufficient criteria to distinguish them. In other languages (such as Chinese), adjectives are grammatically similar to verbs, but again there are criteria to distinguish them. In a number of languages (including English), adjectives have properties rather different from those of nouns and those of verbs. And in a few languages (for instance, the Berber branch of Afro-Asiatic, in North Africa), they combine grammatical properties of nouns and of verbs. In some languages adjectives constitute a large open class, in others a small closed class. The major semantic types associated with every adjective class are Dimension, Age, Colour, and Value.

Now consider the nature of the adjective classes in two North Arawak languages from Brazil. Tariana has a small closed class—with about twenty members—which shares a number of grammatical properties with nouns and also some with verbs. Warekena has a rather larger class, which shows very similar properties to verbs.

In Warekena, both verbs and adjectives take bound pronominal affixes marking core arguments:

- transitive verbs—take pronominal prefixes for A argument and suffixes for O.
- intransitive verbs—take prefixes for S argument; these have the same form as A prefixes with transitive verbs.
- adjectives (which may function as head of an intransitive predicate)—take suffixes for S argument, and these have the same form as O suffixes with transitive verbs.

Adjectives may not directly modify the head of an NP; adjectives and all types of verbs may only function as modifier by means of a relative clause construction.

In Tariana we find:

- transitive verbs—take bound pronominal prefixes for A argument, nothing for O.
- ‘active’ intransitive verbs (type Sa)—take pronominal prefixes for S argument; these have the same form as A prefixes with transitive verbs.
- ‘stative’ intransitive verbs (type So)—do not take bound pronouns.

- adjectives—typically modify a noun, and then take a classifier suffix determined by the semantic nature of that noun.

In this language, a noun may function as head of an intransitive predicate, taking no bound pronominal affixes and allowing a smaller set of aspect- and mood-marking clitics than does a verb. An adjective may also function as intransitive predicate head and has the same restricted set of aspect and mood choices as a noun; in this function an adjective must take a classifier suffix.

As expected, since the two languages are genetically related, we find a fair number of lexical cognates. These include:

| WAREKENA  |                            | TARIANA                                 |
|-----------|----------------------------|-----------------------------------------|
| adjective | <i>makare</i> ‘breathless’ | So verb <i>makare</i> ‘tired’           |
| adjective | <i>amena</i> ‘sharp’       | So verb <i>pimana</i> ‘sharp’           |
| adjective | <i>ari</i> ‘white’         | adjective <i>hare</i> ‘white’           |
| adjective | <i>pure</i> ‘green’        | adjective <i>hiporite</i> ‘green, blue’ |

From these, and from correspondences between transitive verbs in the two languages and between intransitives in Warekena and Sa intransitives in Tariana, the relation between word classes and subclasses in the two languages appears to be:

| WAREKENA           | TARIANA                               |
|--------------------|---------------------------------------|
| transitive verbs   | transitive verbs                      |
| intransitive verbs | Sa intransitive verbs                 |
| adjectives         | { So intransitive verbs<br>adjectives |

But this obscures the fact that, in terms of grammatical properties, adjectives in Warekena are more similar to So verbs than to adjectives in Tariana. A more revealing tabulation is:

| WAREKENA                                                                                                                            | TARIANA                                                                                         |
|-------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| transitive verbs                                                                                                                    | transitive verbs                                                                                |
| Sa intransitive verbs                                                                                                               | Sa intransitive verbs                                                                           |
| So intransitive verbs<br>(recognised as the<br>adjective class through<br>including Dimension,<br>Age, Colour and<br>Value lexemes) | So intransitive verbs<br><br>adjectives (including Dimension,<br>Age, Colour and Value lexemes) |

Comparative analysis indicates that the proto-system, in a shared ancestor of the two languages, was similar to that in Warekena. Bound pronominal prefixes marked A with transitive and S with Sa verbs, while suffixes marked O with transitives and S with So intransitives. The So items were recognized as the adjective class since they include lexemes from the critical semantic types.

Tariana has undergone a number of changes. While bound pronominal prefixes (marking A and Sa) have been retained, pronominal suffixes (which originally marked O and So) have been lost. And the original So-intransitive/adjective class has split into two. Tariana developed an extensive set of classifiers (these are not found in Warekena) which came to be added to some members of the original So class (those referring to Dimension, Age, Colour, and Value). This newly created adjective class then aligned itself with nouns in its grammatical behaviour. As mentioned above, an adjective may directly modify a noun. All verbs—transitive, Sa intransitive, and So intransitive—may only modify through a relative clause.

Turning back to Warekena, the characterization of word classes provided at the beginning of this section is entirely appropriate when the language is considered by itself. Adjectives form a separate class but they share many properties with verbs. However, within the framework of a comparative investigation (and to explain the word class membership of lexical cognates between the languages), it is more insightful to say that Warekena has two classes of intransitive verbs—Sa and So types—just like Tariana. In Tariana, twenty or so of the most adjectival lexemes from the So class have split off to form a new class which it is appropriate to regard as the small adjective class, with many grammatical properties in common with nouns.

That is, the grammatically defined class in Warekena which includes Dimension, Age, Colour, and Value terms should be called the ‘adjective class’ within a grammar of Warekena considered outside of any genetic context. But for comparison with the related language Tariana it is most revealing to identify it as ‘So intransitive verbs’.

## 4.6 Different analyses of similar data

Many languages exhibit a contrast between short and long vowels. There may be a long vowel corresponding to every short one, or just to some of them. (It is unusual to encounter a language with more long than short vowels.)

But what is a long vowel? Is it effectively a sequence of two short vowels? If this were the case, then a long vowel could be written with the symbol for the corresponding short vowel doubled—long *a* as *aa*, etc. Or is it a distinct entity? If this were so, it would be appropriate to write long *a* as *a:* (or as *ā*). When we undertake detailed study of individual languages, it is found that

sometimes the first alternative is most appropriate, and other times the second one. A brief example can be provided for each.

Warrgamay, spoken in North Queensland, has three short vowels—*i*, *a*, *u*—and three long ones. Is a long vowel effectively a sequence of two short ones? We can resolve this question by examination of the form of the positive imperative suffix on verbs. Among its allomorphs are:

- *-ya* after a disyllabic transitive root ending in *i*; for example, *wugi-ya* ‘give!’
- *-∅* (zero) after a trisyllabic transitive root ending in *i*; for example, *gungari-∅* ‘cut!’

Now consider transitive verbs ending in *i*, with a long vowel in the first syllable: *bu:di-* ‘take’ and *ma:ni-* ‘hold’. Their imperatives are *bu:di-ya* and *ma:ni-ya*. That is, they count as disyllabic. The long vowel *u:* in *bu:di-* has the same status as the short vowel *u* in *wugi-*. This suggests that a long vowel should not be regarded as a sequence of two short vowels but as a separate entity. That is, long *u*, for example, should be written as *u:* as here (or *ū*), not *uu*. (See Dixon 1981: 17–20, 52–3.)

Fijian has five short vowels—*i*, *e*, *a*, *o*, *u*—and five long ones. Early grammarians had difficulty analysing the placement of stress in a Fijian word: usually on the penultimate but in some words on the final syllable (Hale 1846: 367).

Here is a sample of words of different shape with stress marked:

*vúla* ‘moon’    *córi* ‘be tied’    *rái* ‘look, see’    *tádra* ‘dream’  
*waqóna* ‘kava’    *coríta* ‘tie’    *ráica* ‘look at’    *tadráa* ‘dream of’

To explain these stress placements, one needs to invoke the unit ‘mora’:

- a short vowel counts as one mora
- a diphthong (such as *ai* in *rai* and *raica*) counts as two moras; that is, each component counts as one mora
- a long vowel counts as two moras

The stress rule is then entirely straightforward: primary stress goes onto the syllable containing the second mora from the end (the penultimate mora) of a phonological word.

In *rái* the first element of the diphthong, *a*, is the penultimate mora and thus the whole diphthong takes stress. In *ráica* (which is verb root *rai* plus transitive suffix *-ca*), the second element of the diphthong, *i*, is the penultimate mora, and again the whole diphthong takes stress. In *tadráa* (which is verb root *tadra* plus transitive suffix *-a*), the first *a* of this long vowel is

the penultimate mora and so the whole long vowel, *aa*, bears stress. (Fuller information is in Dixon 1988: 14–18, 24.)

It will be seen that in Fijian—a mora-counting language—a long vowel can be usefully regarded as a sequence of two short vowels, with long *a* being written as *aa* (not as *a:* or *ā*), and so on. This is in contrast to Warrgamay—a syllable-counting language—where a long vowel should *not* be regarded as a sequence of two short vowels. (There is further discussion of this in §7.6.)

As a further example of how a certain datum requires diverse analyses in different languages, consider a diphthong with [a] as the peak, followed by off-glide [i], that is [a<sup>i</sup>]. Such a diphthong occurs in Dyirbal and in Fijian but requires different treatment, largely because of the differing phonotactics of the two languages.

In Dyirbal, a word has the shape CV(C)(C)CV(C). That is, a final syllable can be open or closed; and there are no vowel sequences. Consider phonetic form [yaga<sup>i</sup>] ‘owl’, which ends in diphthong [a<sup>i</sup>]. In terms of the phonotactic template of the language, this must be accorded phonological form /yagay/, ending in semi-vowel /y/, which is a consonant, rather than /yagai/, with a final vowel sequence.

But is there independent justification for this? Perhaps the word structure just given should be amended to allow for a final /VV/ sequence. In fact there is evidence. Consider the form of ergative inflection on a selection of nouns:

|                                               |                                           |
|-----------------------------------------------|-------------------------------------------|
| <i>wadam</i> ‘snake’: <i>wadam-bu</i>         | <i>yara</i> ‘man’: <i>yara-ŋgu</i>        |
| <i>midin</i> ‘possum’: <i>midin-du</i>        | <i>yuri</i> ‘kangaroo’: <i>yuri-ŋgu</i>   |
| <i>jarruñ</i> ‘mockingbird’: <i>jarruñ-ju</i> | <i>bimu</i> ‘uncle’: <i>bimu-ŋgu</i>      |
| <i>yagay</i> ‘owl’: <i>yagay-ju</i>           | <i>yamani</i> ‘rainbow’: <i>yamani-gu</i> |

It will be seen that the ergative case has the following forms:

- -*ŋgu* after a disyllabic stem ending in any vowel (*a*, *i*, or *u*)
- -*gu* after a stem of more than two syllables, again ending in any vowel
- -*Hu* after a consonant, where *H* is a stop homorganic with the preceding consonant; that is, -*bu* after *m*, -*du* after *n*, and -*ju* after *ñ* or *y* (note that *j* and *ñ* are lamino-palatal stop and nasal respectively; *y* is a palatal semi-vowel)

That is, ‘owl’ takes allomorph -*Hu*, found after consonant-final roots, suggesting that its phonological form must be /yagay/. If it were /yagai/, ending in a vowel, then the ergative suffix would be -*ŋgu* (or possibly -*gu*).

We can now look at Fijian, which only has open syllables, of structure (C)V. Each word must be of at least two moras; that is, if it is monosyllabic, it must

include either a diphthong or a long vowel. On this principle [ra<sup>i</sup>] ‘see, look’ has to be accorded phonological form /rai/, rather than /ray/. But one must explore all possibilities. Could one perhaps say that there are some final CVC syllables, so that [a<sup>i</sup>] is /ay/?

In fact, this wouldn’t work. Consider [madrai] ‘bread’. Recall that in Fijian stress goes on the syllable which includes the penultimate mora. If we worked in terms of /madray/ then the stress would go on the first /a/, giving \*/mádray/. This is wrong; stress goes on the final syllable, indicating that the phonological form of this word must be /madrai/, with the phonetic diphthong interpreted at the phonological level as a sequence of two vowels.

The question of phonological interpretation of a phonetic diphthong is a fascinating matter, of which only the surface has been scratched here. But it should suffice to illustrate how a specific phonetic datum may require different analyses in different languages, depending on the phonological structure of words in the languages and sometimes also on grammatical structure.

## 4.7 Intertwining of levels

One of the most grievous errors which a linguistics student (or their teacher) can embrace is to believe that in describing a language one can satisfactorily complete one part before embarking on another. ‘I’m going to spend the first year working on phonology, submit that for my MA, and after that that study morphology for a PhD, and syntax for a post-doctoral fellowship.’ But you can’t! Each part depends on the others. In the last section it was shown how a decision on the phonological status of phonetic long vowels in Warrgamay requires information from verb morphology, and how a decision on the phonological status of phonetic diphthongs in Dyrirbal requires information from noun morphology.

A further example can be taken from Fijian. The adjective ‘bad’ was quoted in §4.4 with form *caa*. There is a related verb *ca-ta* ‘hate, consider bad’, which includes transitive suffix *-ta*. The root ends in a long vowel when there is no suffix but in a short vowel before a suffix. A number of verbs operate on the same pattern. Compare:

drée ‘pull’      cáá ‘bad’  
*dré-ta* ‘pull (it)’    *cá-ta* ‘hate, consider bad’

There are two constraints on phonological words in Fijian:

- (i) Each phonological word must include at least two moras
- (ii) A long vowel is never found in the penultimate syllable of a phonological word whose last syllable involves a short vowel



This suggests alternative analyses:

- A The roots are *dree* and *caa*; their vowels must be shortened when there is a following monosyllabic suffix such as *-ta*, to satisfy constraint (ii).  
 B The roots are *dre* and *ca*; their vowels are lengthened when no suffix follows, to satisfy constraint (i).

Which of these alternative forms for the two roots should be preferred? It is impossible to decide on phonological evidence alone. One needs to look to grammar for assistance.

Just a few adjectives (nine in my corpus) undergo partial reduplication, with a plural meaning. The reduplication repeats—before the root—initial C and the following vowel mora. For example:

|             |       |                |                      |
|-------------|-------|----------------|----------------------|
| <i>lévu</i> | ‘big’ | <i>le-lévu</i> | ‘lots of big things’ |
| <i>vóu</i>  | ‘new’ | <i>vo-vóu</i>  | ‘lots of new things’ |
| <i>cáa</i>  | ‘bad’ | <i>ca-cáa</i>  | ‘lots of bad things’ |

Now if the underlying form of ‘bad’ were *ca*, under Analysis B, the reduplicated form should be *cá-ca*; this has two moras and is a perfectly good phonological word in Fijian. But the actual form is *ca-cáa*, showing that the underlying form must be *caa*, from Analysis A. We infer that the underlying form of ‘pull’ should be taken to be *dree* (rather than *dre*). Thus, evidence from the morphological process of reduplication has served to resolve a dilemma of phonological analysis.

The first task of any field linguist is to formulate a preliminary statement of the phonology, so that they may begin to transcribe. If in doubt as to whether a particular phonetic distinction is phonologically significant, it is best to employ caution. Suppose that—in the initial stages of analysis—one cannot decide whether [e] and [ɛ] are contrasting phonemes or conditioned variants of one phoneme. For the time being, a narrow transcription should be adopted, writing each vowel as it is heard, as ‘e’ or ‘ɛ’. Later work may show that these are contrasting phonemes. Or it may indicate that they are allophones of a single front mid vowel. One can then abandon the use of both ‘e’ and ‘ɛ’ and employ a single letter (probably ‘e’, since this occurs on all standard keyboards). Suppose the opposite happened—the linguist began by writing all these vowels as ‘e’, but in due course—maybe as a consequence of grammatical analysis—realized that there is a contrast between /e/ and /ɛ/. Each example of ‘e’ in the data collected would then have to be gone over and checked, to see whether it was a close-mid or an open-mid front vowel.

A preliminary phonological description is plainly the first task. But a definitive statement of the phonology is one of the last things to be completed,

taking account of all alternations and conditioned variants which may have been revealed during morphological analysis.

The important point here is that work on description of a language *cannot be compartmentalized*. One has, essentially, to work at every level all the time. Suppose a problem comes up in phonological analysis which can't immediately be provided with a solution. Put it aside for the moment, and do some work on morphology. Some parts are straightforward, but one problem defies immediate resolution. Put that to one side and work for a while on syntax. After considerable study of the database and intensive thinking, a fine insight emerges which explains what had been a tricky syntactic puzzle. Furthermore, this syntactic insight feeds into the morphological problem which had been put aside, and assists in resolving it. Then this morphological solution sheds light on the phonological difficulty which had been shelved, helping to suggest a solution to it. And so it goes on.

(The very idea of trying to complete one part of a linguistic description before starting on other parts should seem bizarre to modern linguists. However, such a procedure is followed today in some nations. And it was most emphatically promulgated by the so-called 'post-Bloomfieldians' in the USA during the 1940s and beyond. This ill-advised position was countered by Kenneth L. Pike in a classic paper from 1947, 'Grammatical prerequisites to phonemic analysis'. It is included in §2.6 within the list of papers which all linguistics students are encouraged to study.)

## 4.8 Encountering the unexpected

No language presents a neat and tidy system. The entities described by formal theorists look to be compact and orderly because they are intended by the formalists to be so. The interest is in some hypothetical 'logical form' which is confirmed by judiciously controlled elicitation and/or introspection. A scientific linguist analyses texts and must account for every nuance within them. Loose ends *may not* be cast aside; they—like the central core—are to be described and explained.

Not infrequently, a certain semantic goal may be achieved by quite diverse grammatical mechanisms, X and Y. In some circumstances, X applies; in others, Y does. There is likely to be overlap between them, when either X or Y may be employed; which is chosen can be essentially a matter of style. Consider comparison in English, shown by a morphological process with suffix *-er*, or by a syntactic construction with modifier *more*. It has been suggested that *-er* and *more* are variants of a single category. Just as plural inflection on count nouns can be /-z/, /-əz/, or /-s/, depending on the phonological environment,

TABLE 4.2. Possibilities for *-er* and *more*

| FORM OF ADJECTIVE                                                | TAKES<br><i>-er</i> | <i>more</i> CAN<br>BE USED |
|------------------------------------------------------------------|---------------------|----------------------------|
| Monosyllabic, e.g. <i>big, dry, strong</i>                       | ↓                   | ↑                          |
| Disyllabic                                                       |                     |                            |
| monomorphemic                                                    |                     |                            |
| ending in /l/, e.g. <i>heavy, happy</i>                          |                     |                            |
| ending in another vowel, e.g. <i>yellow, clever, secure</i>      |                     |                            |
| ending in syllabic /l/, e.g. <i>simple, humble</i>               |                     |                            |
| dimorphemic, ending in a vowel, e.g. <i>cloudy, friendly</i>     |                     |                            |
| All others                                                       |                     |                            |
| disyllabic and longer, ending in a consonant, e.g. <i>famous</i> |                     |                            |
| trisyllabic and longer, ending in a vowel, e.g. <i>ordinary</i>  |                     |                            |

so the index of comparison is said to be *-er* /-ə(r)/ or *more* /mɔ:(r)/ depending on the environment. In fact this is a misguided approach.

Many morphological processes in English have limited applicability. For example, suffix *-en*, which derives intransitive and transitive verb stems, may only be added to adjectives ending in /p/, /t/, /k/, /f/, /s/, /ʃ/, /θ/, or /d/, and is mainly used with items belonging to the Dimension, Physical Property, and Speed types, plus three Colour terms (*black, white, and red*). For other adjectives periphrastic means have to be employed; for example, alongside *wid-en* and *soft-en*, one has to say *become proud* (not \**proud-en*) and *make jealous* (not \**jealous-en*). (See Dixon 1982: 17, 21–4.)

In similar fashion, comparative suffix *-er* is limited in the forms it can be used with—basically, all adjectives which are monosyllabic plus disyllabic forms ending in a vowel or syllabic *l*. The syntactic modifier *more* may be used with adjectives that do not take *-er*, and also with some that do. Applicability of the two mechanisms is roughly summarized in Table 4.2. It will be seen that the two mechanisms overlap in the middle rows—either may be employed with adjectives such as *yellow, clever, secure, simple, humble, cloudy, and friendly*.

There are just a few exceptions, a handful of disyllabic forms ending in a consonant for which *-er* is possible (as an alternative to *more*). Some are the antonyms of adjectives which take *-er* because they *do* have an appropriate phonological form; alongside *clever-er, rud-er, and hollow-er* we also get *stupid-er, polit-er, and solid-er*. (For fuller details see Dixon 2005a: 91–2; 2005b.)

The only function for this suffix *-er* is to be the index of comparison. In contrast, *more* has a wider range of semantic uses. It is used with nouns, then taking on a quantitative sense, as in *more sugar* or *more cows*, and may also

be used with an NP, as in *he's more a fool than she is*. And it can make up a complete NP, as in *Oliver Twist asked for more*.

It will be seen that syntactic modifier *more* and the morphological process shown by suffix *-er* are not mutually exclusive and cannot be regarded as variants of a single grammatical morpheme. They operate in different ways and while between them they do cover all adjectives available for comparison, there is a degree of overlap. (How much overlap there is varies—to some extent—from speaker to speaker.)

As stated before, it is reasonable for a linguist studying a new language to have certain expectations of what may be encountered, based on what is known of related and/or neighbouring languages. But expectations are not always fulfilled. When beginning work on Warrgamay, I anticipated that it would be like neighbouring languages and have a tense system. There is in fact a system of four verbal inflections in declarative main clauses, each with different allomorphs in transitive and intransitive clauses; just the intransitive forms are quoted here:

- ‘Purposive’, *-lagu*, indicates ‘have to’, ‘want to’, or ‘will (as a definite prediction)’
- ‘Irrealis’, *-ma*, is used for possible future (‘might happen’), negative future (‘won’t do it’), and apprehension (‘lest it happen’).
- Suffix *-gi* refers to events in the past, but only if they are irretrievably finished (for example ‘the camp was all burnt down’); the appropriate term for this is ‘perfect’.
- The fourth verbal suffix, *-y*, is used in all other circumstances—for referring to the present, or something in the past which is perhaps not totally finished. It can, in fact, be used in *all* circumstances, as an alternative to purposive, irrealis, and perfect, if the speaker does not need to or does not want to provide specification on the verb. For example:

- (1) n̄irwara    ŋayba<sub>s</sub>    gaga-y  
       tomorrow 1sg        go-UNMARKED  
       I’ll go tomorrow

In this sentence, the temporal word ‘tomorrow’ indicates future time, so the verb just takes suffix *-y*.

I refer to *-y* as the ‘unmarked inflection’; see §5.7. It can be used instead of a suffix with more specific meaning, and it must be employed when none of the other three verb inflections would be appropriate. In summary, Warrgamay lacks a tense system, but it does have a set of aspect-like markers. This was an unexpected—and rather exciting—analytic result.

A further point is that it must not be assumed that because certain categories have similar realization they should be expected to have similar grammatical status. Under (e) in §1.10, it was noted that Latin has one set of nominal inflections which fuse information on gender, number, and case. These three verbal categories have distinct semantic and grammatical roles—gender is an inherent property of a noun, number is a referential feature of an NP, and case marks the function of an NP in its clause. It also has a system of inflectional suffixes on verbs which combine information on tense, mood, and voice, plus person and number of subject argument. Like the categories fused together in nominal morphology, each of these has a quite distinct meaning and function.

Some languages have simpler inflectional systems—say, case on nouns and tense on verbs. They are similar in both being inflections (see the discussion of this in §5.3) but in no other way. Case marks syntactic function, while tense indicates the temporal reference of an event or state. It must be realized that the properties of one will not in any way parallel those of the other category.

Terminology can carry expectations of similarity which may or may not be upheld, and may engender confusion in the mind of a grammar reader. ‘Antipassive’ is the ‘opposite’ of ‘passive’ in some aspects of its syntax but not in semantic effect (see §3.20). The term ‘antipassive’ is not a particularly helpful one but is now well established, so that it would be foolish to try to change it. One should always bear in mind that terminology is just that—names. It is fine if the technical use of a term accords in part with its everyday meaning or with its etymology, but often it does not do so. ‘Passive’ itself, as a grammatical label, is detached in meaning from the adjective passive, meaning ‘inactive, inert’. ‘Dative’, as the term is used in many grammars, is not restricted to marking something that is given, which is what the etymology would imply.

One has necessarily to commence any linguistic study by examining the surface structure of utterances. But analysis must extend far beyond this. The underlying grammatical system of a language is not in terms of sequences of words but of relations between categories and construction types. Several apposite examples were provided earlier. Under (a) in §1.10 we saw how the English phrasal verb *hand over* can be used in *John handed over the documents* or in *John handed the documents over*. Although the two elements of the lexeme occur next to each other in *John handed over the documents*, it was shown that *John handed the documents over* has to be taken as the canonical construction in order to explain grammatical properties of this and other phrasal verbs. Under (c) in §1.10 we saw that it is not useful, when a core argument is realized both through an NP and through a bound pronominal element within the

predicate, to enquire which of these is the major instance of the argument. Neither is. The underlying argument is an item in grammatical structure, its manner of realization being a secondary issue. Under (d) in §1.10 we examined four completely different construction types which can each have a reduced version ‘verb(–object)–adjective’ (illustrated by *I consider John clever*, *I want the house clean*, *John retired rich*, and *He licked the plate clean*). The adjectives in these constructions have been referred to as ‘secondary predicates’. They have virtually no grammatical properties in common, save occurring in a superficially similar reduced realization.

A grammatical system most often does have all its terms realized at one place in surface structure. But it need not have, and one should always be on the lookout for this eventuality. The Australian language Tiwi has a complex structure for its verbal word, with twelve prefix and three suffix slots. There is a two-term system of temporal specification—‘happens in the morning’, prefix *-atə-*, and ‘happens in the evening’, prefix *-kə-*. As would be expected, only one of these may be included within a given verb. The interesting—and unexpected—fact is that the ‘in the morning’ form occurs in prefix slot 6, immediately before an object pronominal prefix (in slot 7), whereas the ‘in the evening’ form comes in slot 11, between the future-imperative suffix (slot 10) and the ‘do while walking’ suffix (slot 12). Slot membership is a matter of surface realization, with little or no relevance for the process of communication. What is significant is that these ‘time of day’ specifications make up a two-term grammatical system. (See Osborne 1974: 37, 46; and also Lee 1987: 151, 186.)

In summary, we find that mechanisms with quite different profiles may be combined to achieve some grammatical goal—as it were, homing in on it from different angles. And what is one unit or one category in terms of the underlying grammar may be mapped onto different places in surface structure. Basic linguistic theory sets out a typological paradigm, by inductive generalization from reliable grammars. But each individual grammar has its own personality, producing little or big issues of surprise and wonderment.

## 4.9 Explanation

Everybody seeks a reason. Joe Bloggs just died at an early age. Why? Because he had lung cancer. Why again? Because he smoked too much. And why did he do that? He started when young, due to peer group pressure, and although he often tried to give up the habit, could never summon up the requisite determination.

In English, *mouse*, /maʊs/, has an irregular plural *mice*, /maɪs/. Why? Because these reflect earlier forms /mu:s/ and /my:s/, where /y/ is a front

rounded vowel. First /y/ lost its rounding, becoming /i/, and then, by the Great Vowel Shift, /u:/ and /i:/ became /au/ and /ai/ respectively. But why were the earlier forms /mu:s/ and /my:s/? The root was /mu:s/ and plural added /i/, giving /mu:si/. Then an assimilatory process called umlaut applied, causing the /u:/ in /mu:si/ to be fronted, like the following /i/, becoming /y/. Then the final /i/ in /my:si/ gradually weakened and finally dropped. In summary, the singular form went from /mu:s/ to /maus/ and the plural one from /mu:si/ to /my:si/ to /my:s/ to /mi:s/ to /mais/.

Many explanations have been presented in the chapters above and more will follow in succeeding ones. It is useful to here summarize types of explanation. Some of the most intricate involve one part of the language system being organized in such a way that it helps meet a constraint in some other portion. This can be illustrated for the Australian language Yidiñ, where morphotactics and allomorphy within the grammar appear to be directed towards achieving a phonological goal.

Yidiñ has three vowels (*a*, *i*, *u*) each with a long congener. It is—like Warrgamay, described in §4.6—a syllable-counting (not a mora-counting) language, in which short and long vowels each count as one phonological unit. If there is no long vowel in a phonological word, stress goes on the first syllable and on every second syllable thereafter; for example *búña* ‘woman’, *ɲúnangárra* ‘whale’. If there is a long vowel, it must bear stress, and every second syllable—counting in each direction from it—must also bear stress; for example *ɲunáɲgarrá:ɲ* ‘whale+ERGATIVE’. Each long vowel must be stressed and adjacent syllables cannot both bear stress. Thus, if a phonological word includes two long vowels, they must be separated by an odd number of syllables.

There is a preference in Yidiñ for each phonological word to have an even number of syllables—that is, a whole number of disyllabic feet. These can either be trochaic (a stressed syllable followed by an unstressed one) as in *málan* ‘flat rock’ and *ɲúnangárra*, or iambic (unstressed syllable followed by stressed one) as in *malá:n* ‘right hand’ and *ɲunáɲgarrá:ɲ*.

Consider a short paradigm of three nouns in three forms—absolutive, with zero suffix, ergative, with suffix *-ɲgu*, and genitive, with suffix *-ni* (note that *j* and *ñ* are lamino-palatal stop and nasal):

| (1)     | ABSOLUTIVE | ERGATIVE                         | GENITIVE                        |
|---------|------------|----------------------------------|---------------------------------|
| ‘woman’ | búña       | buña-ɲgu → buñá:ɲ                | buña-ni → buñá:n                |
| ‘man’   | wagú:ja    | wágujá-ɲgu                       | wágujá-ni                       |
| ‘whale’ | ɲúnangárra | ɲunangarra-ɲgu →<br>ɲunáɲgarrá:ɲ | ɲunangarra-ni →<br>ɲunáɲgarrá:n |

As described in §3.13, two phonological rules are in operation. First, if a phonological word has an odd number of syllables, the vowel in the penultimate syllable is lengthened, and bears stress. Corresponding to root *waguja* the absolutive form is *wagú:ja*. Secondly, the ending on an odd-syllabled word may be truncated, so that there is an even number of syllables (a whole number of iambic feet). Thus:

*búña* plus *-ŋgu* becomes *buñá:ŋgu* by rule 1 and then *buñá:ŋ* by rule 2

And similarly for *buñá:n*, *ŋunáŋgarrá:ŋ*, and *ŋunáŋgarrá:n*. The reduction is only possible if there is a morpheme boundary immediately after the penultimate vowels; or if the consonant following the penultimate vowel is one that may occur word-finally (a nasal or liquid or *y*, not a stop or *w*). *Wagú:ja* does not reduce to *wagú:j* since neither of these conditions is satisfied. Of the nine words in (1), only one (*wagú:ja*) has an odd number of syllables. If the truncation rule had not applied, there would have been five.

That is one source for long vowels. There are also two derivational suffixes to verbs which involve an inherent long vowel. They have the following allomorphs after stems from the two main verbal conjugations:

|     |                                   |                |                |
|-----|-----------------------------------|----------------|----------------|
| (2) |                                   | -N conjugation | -L conjugation |
|     | antipassive                       | -:ji-N         | -:ji-N         |
|     | ‘do while going’ (or ‘go and do’) | -ŋali-N        | -:li-N         |

A major function of the antipassive suffix, *-:ji-N*, is to derive an intransitive stem from a transitive one. The ‘(do while) going’ suffix does not affect transitivity. Both suffixes form a stem belonging to the *-N* conjugation.

The past tense inflection is *-ñu* onto a stem from the *-N* and *-lñu* with one from the *-L* conjugation. Examples of its use with two disyllabic verb roots are:

|            |                                |                                 |
|------------|--------------------------------|---------------------------------|
| root       | <i>gali-N</i> ‘go’             | <i>baga-L</i> ‘spear’           |
| past tense | <i>gali-ñu</i> → <i>gali:ñ</i> | <i>baga-lñu</i> → <i>bagá:l</i> |

When the two derivations are added to *baga-L*, and then past tense, we get:

|                           |                  |                  |
|---------------------------|------------------|------------------|
|                           | antipassive      | ‘do while going’ |
| add derivation            | <i>baga:ji-N</i> | <i>baga:li-N</i> |
| add past tense inflection | <i>bagá:jiñú</i> | <i>bagá:liñú</i> |

Each of the inflected verbs has four syllables, consisting of two iambic feet.

A root may take both of these derivational suffixes. Since they are grammatically independent of each other, there is no a priori expectation



about the sequence they should occur in. We can follow through the two possibilities:

|   |                                   |                     |
|---|-----------------------------------|---------------------|
| A | root                              | <i>baga-L</i>       |
|   | apply ‘do while going’ derivation | <i>baga:li-N</i>    |
|   | apply antipassive derivation      | <i>baga:li:ji-N</i> |

This would result in there being long vowels in successive syllables, contravening the phonological constraint against such an eventuality.

|   |                                   |                       |
|---|-----------------------------------|-----------------------|
| B | root                              | <i>baga-L</i>         |
|   | apply antipassive derivation      | <i>baga:ji-N</i>      |
|   | apply ‘do while going’ derivation | <i>baga:ji-ŋali-N</i> |

Although *baga-L* belongs to the -L conjugation, the derived stem *baga:ji-N* is in the -N class and takes the *-ŋali-N* allomorph of the ‘do while going’ suffix. Thus, B accords with the phonological possibilities for Yidiñ; and this is the order in which the two derivational suffixes must occur. The order of derivations in A is avoided, since it would produce a form which would violate the phonological constraints of the language.

The ‘do while going’ suffix developed historically from lexical verb *gali-N* ‘go’. It has reduced to *-li-N* on a verb from the -L conjugation but retains disyllabic form (with no long vowel), *-ŋali-N*, on a stem from the -N class.

There are two aspects of the morphology of Yidiñ which appear to be oriented towards satisfying requirements of the phonology:

- (i) There is no grammatical dependency between antipassive and ‘do while going’ derivations and so they could occur in either order. The order in which they do occur (antipassive before ‘do while going’) ensures that we do not get a form with long vowels in successive syllables.
- (ii) The allomorphy of the ‘do while going’ suffix also bears on satisfying this phonological constraint. If ‘do while going’ had reduced to be *-li-N* also on a form from the -N conjugation, then sequence B would have given *baga:ji:li-N*, again with long vowels in successive syllables.

Note that (i) and (ii) are interrelated. If ‘do while going’ had reduced to *-li-N* on a form from the N conjugation but remained *-ŋali-N* with one from the -L class, then sequence B would have produced the unacceptable *baga:ji:li-N* and A would have given rise to *bagaŋali:ji-N* which is phonologically fine. In essence, it is necessary to have the ‘do while going’ suffix retain a disyllabic allomorph (with no long vowel) in *one* conjugation, and which conjugation this is must then determine the sequence in which the two derivational suffixes occur.

We have thus summarized an example of one type of explanation, which involves showing how one part of a language system is organized in a certain way in order that it should satisfy a constraint in some other portion.

Another kind of explanation can relate to physiological capacity for speech production. It was pointed out in §1.3 that if a language has a non-symmetrical system of consonants and/or of vowels, the most likely gaps will be in the dorso-velar series for consonants and among back vowels. The tongue has to move furthest from a rest position to produce such sounds, and there are likely to be more choices available where articulation is relatively easy than where it requires more effort.

Some explanations come from outside the language system itself; for example, in a contact situation which involves languages of different genetic affiliations. Under (b) in §1.6 we mentioned that Tariana was originally like other North Arawak languages in showing little grammatical marking of evidence. Then it moved into the multilingual community of the Vaupés River basin, into contact with Tucanoan languages, and—to be like them—developed a five-term grammatical system of evidentiality. Under (c) in §1.6 we showed how the environment in which a language is spoken may serve to explain some things about the make-up of its grammar. For example, demonstratives which refer to vertical displacement—such as ‘that higher than here’ and ‘that lower than here’—are typically found in languages spoken in mountainous terrain.

Many aspects of the structure of a language have explanation in the culture, lifestyle, and beliefs of speakers. This comes out clearly in study of the semantics of Dyirbal noun classes, in §1.9. ‘Moon’ is in the same class as men and ‘sun’ in the same class as women since in mythology they are husband and wife. And so on.

A language is affected by its neighbours, which may provide explanation for some structural traits. But it is also a product of its history, and a good deal of explanation lies in that direction. The pioneer linguist Baudouin de Courtenay wrote, in 1871 (translation from Stankiewicz 1972: 63): ‘The mechanism of a language (its structure and composition) at any given time is the result of all its preceding history and development, and each synchronic state determines in turn its future development.’ However, there have been misconceptions concerning the role of historical explanation in relation to synchronic description, which should be addressed.

Before the twentieth century, lots of linguistic-type work was done by philologists who would examine just a fragment of a language, perhaps a single word, and study its history. As mentioned in §1.8, linguist Ferdinand de Saussure decried this. He made the point—emphasized throughout the

present book—that a part only has significance with respect to the whole to which it belongs. The object of study must be the *overall system* of a language at some point in time, a synchronic study. One can then pursue diachronic comparison of synchronic systems (not of bits of them). Oddly, Saussure's dictum has been misinterpreted as saying that one should never take account of diachronic factors while undertaking synchronic analysis. Nothing could be more mistaken.

Insightful explanations are always possible when there are historical records—of reasonable time-depth—on a language. In §2.1 it was pointed out that in British English (at least until very recently) one could not form a comparative or superlative on adjective *little*; that is, there were no forms *\*littler* or *\*littlest* (neither could one say *\*more little* or *\*most little*). This has an historical basis. The old comparative and superlative forms of *little* have been grammaticized as general comparative and superlative markers, *less* and *least* (for example, *less good*, *least suitable*), while *little* itself remained a lexical item, leaving a gap for its own comparative and superlative forms.

In the absence of actual historical data, use can be made of reconstructed past stages of a language, by the comparative method. This was invoked in explanation of the irregular plural form, *mice*, of *mouse*, at the beginning of this section. Under (5) in §3.13, we posited for Samoan underlying verb roots with a final consonant (for example, *silaf* 'see'), even though in this language each word must end in a vowel; inflections then either delete the root-final consonant (as in the imperative *silaf-* 'see!') or add a vowel or vowel sequence (past tense *silaf-ia* 'saw'). By comparison with related languages, we can reconstruct a past stage in which words could end in a consonant, which provides an explanation for the present-day scenario.

Feminine/masculine forms of body-part nouns in Jarawara were illustrated in §2.1. We find *noki/noko* 'eye', where gender is shown by alternation of the final vowel, and *tame/teme* 'foot', where gender is shown by the value of the vowel in the first syllable. Comparative reconstruction across languages of the small Arawá family shows that the original forms were *noko-ni/noko-ne* and *tama-ni/tama-ne*, with constant root forms (*noko* and *tama*) to which were added gender suffixes: *-ni* for feminine and *-ne* for masculine. Regular rules of historical change derive the modern forms from these reconstructed ones.

In similar fashion, a puzzling feature in one dialect of a language may find explanation through comparison with what occurs in other dialects. Dyirbal, like most Australian languages, has three vowels—*a*, *i*, and *u*. The most northerly dialect, Ngajan, has developed long vowels in a manner which can be easily understood by comparison with cognate forms in central dialects such as Mamu. For instance:

|        |              |              |               |               |
|--------|--------------|--------------|---------------|---------------|
| Mamu   | <i>balga</i> | <i>bulal</i> | <i>wayñji</i> | <i>yalgay</i> |
| Ngajan | <i>ba:ga</i> | <i>bula:</i> | <i>wa:ñji</i> | <i>ya:ga:</i> |
|        | ‘hit’        | ‘firefly’    | ‘go uphill’   | ‘road’        |

Roughly, an *l* or *y* at the end of a syllable—before a consonant or at the end of a word—has been replaced by length on the preceding vowel. (Note that an *l* or *y* at the beginning of a syllable is not affected; as with the first *l* in *bulal/bula:*.)

The change applies to lexemes, and in all parts of the grammar. For instance, the applicative derivational suffix to verbs (added to a vowel-final stem) is:

|        |               |               |
|--------|---------------|---------------|
|        | Y CONJUGATION | L CONJUGATION |
| Mamu   | <i>-yma-</i>  | <i>-lma-</i>  |
| Ngajan | <i>:-ma-</i>  | <i>:-ma-</i>  |

However, there are places in the grammar of Ngajan where vowel length is not at first explainable. Compare two tense inflections:

|        |               |               |               |               |
|--------|---------------|---------------|---------------|---------------|
|        | PAST TENSE    |               | FUTURE TENSE  |               |
|        | Y CONJUGATION | L CONJUGATION | Y CONJUGATION | L CONJUGATION |
| Mamu   | <i>-ñu</i>    | <i>-n</i>     | <i>-ñ</i>     | <i>-ñ</i>     |
| Ngajan | <i>-ñu</i>    | <i>-n</i>     | <i>-ñ</i>     | <i>:-ñ</i>    |

Allomorphs are identical between the two dialects, save for vowel length in future tense for just the L conjugation in Ngajan. Comparative data from Mamu and other central dialects sheds no light on this.

However, I was also able to gather good data on Girramay, the most southern dialect of the Dyirbal language. Here the two tense inflections are (as for Yidiñ, *ñ* is a lamino-palatal nasal, and *j* the corresponding stop):

|          |               |               |               |               |
|----------|---------------|---------------|---------------|---------------|
|          | PAST TENSE    |               | FUTURE TENSE  |               |
|          | Y CONJUGATION | L CONJUGATION | Y CONJUGATION | L CONJUGATION |
| Girramay | <i>-ñu</i>    | <i>-n</i>     | <i>-njay</i>  | <i>-ljay</i>  |

We can take the Girramay forms of future as the original ones. The central and northern dialects shortened these forms by:

- Dropping the final *-ay*.
- Since words in Dyirbal (as in most Australian languages) cannot end in a consonant cluster or a stop, the ending *-nj* in the Y conjugation was simplified to *-ñ*, combining manner of articulation (nasal) of *n*, the first segment, with place of articulation (lamino-palatal) of *j*, the second segment. This form was then analogized to the L conjugation.

What happened in Ngajan is that the rule replacing a syllable-final *l* by vowel length preceded the truncation of the future inflection:

|                        | Y CONJUGATION | L CONJUGATION |
|------------------------|---------------|---------------|
| original forms         | <i>-njay</i>  | <i>-ljay</i>  |
| $l \rightarrow :/- C$  | <i>-njay</i>  | <i>-:jay</i>  |
| truncation and analogy | <i>-ñ</i>     | <i>-:ñ</i>    |

Thus the vowel length in future allomorph, *-:ñ*, for the L conjugation in Ngajan has its origin in the initial *-l-* of suffix form *-ljay*, which is retained in Girramay.

Occurrences of long vowels in Ngajan (they are not found in other dialects) can thus be explained through comparison with forms in other dialects. This shows the value of gathering basic information on every dialect of a language under study. Generally, a linguist will focus on one dialect, and produce a comprehensive description of this. But if at all possible a little time should be reserved for some degree of examination of other dialects.

The final kind of explanation is typological. In §12.8 we note a correlation between the grammatical properties of its adjective class and whether a language has bound pronominal elements in the predicate indicating the core arguments, or whether the function of a core argument is shown by case marking on an NP realizing it. Very roughly:

- (1) (a) Adjective classes which are similar in their grammatical properties to nouns tend to be found in languages which mark core arguments through a case system applying to NPs.
- (b) Adjective classes which are similar in their grammatical properties to verbs tend to be found in languages which involve bound pronominal elements in the predicate, and also in those which have neither a case system nor bound pronouns.

This can serve to explain what kind of adjective class a given language has, in terms of its profile with respect to the way core arguments are shown. It can also explain processes of change. Suppose that a language shifts its profile in the following way. Free pronouns are grammaticalized to become obligatory bound pronominal elements within the predicate. When this is accomplished, case marking on NPs may become redundant and be dropped. In due course, the adjective class is likely to reorient itself so that correlation (1) is maintained—adjectives will take on new properties similar to those of verbs, and may divest themselves of some properties which were similar to those of nouns.

Basic linguistic theory reveals a fair number of dependencies between grammatical categories and construction types, thus providing a degree of explanation for the kind of internal organization the grammar of an individual language shows, and also how a certain type of change in profile may trigger another change, to preserve a typological correlation.

That is, it has significant predictive power.

## Sources and notes

4.3. A fuller discussion of relative clauses in Fijian is in Dixon (1988: 43, 251–5); and see §17.4 below.

Fuller discussion of ergative and instrumental in Dyirbal is in Dixon (1972: 42, 93–6). The examples here have been simplified by the omission of noun markers which are frequent, but not obligatory, components of NPs in Dyirbal. This does not have any bearing on the discussion here. Noun markers are included for all examples in Dixon (1972).

4.4. A fuller account of number words in Fijian is in Dixon (1988; see particularly pages 141–50, 99, 198, and 238).

4.5. Information on Warekena and Tariana comes from published grammars and dictionaries by Aikhenvald and from detailed personal discussion with her. See Aikhenvald (1998: 357–8, 420; 2003: 66–76; 2002a).

4.9. The discussion here of long vowels in Yidiñ has merely skimmed the surface of parts of the grammar and phonology of this language. For fuller discussion see Dixon (1977a: 42–77, 227–33; 1977b).

## Terminology

It is useful to gather together explanations of a number of important terms. Some have been introduced above and are recapitulated here. Some will be explained more fully in chapters which follow. Others receive their major—or only—discussion in this chapter.

### 5.1 Grammar and lexicon

The distinction between the two components of a language description was drawn in §1.11. The **grammar** describes underlying categories and structures. It is concerned with small closed systems—gender, number, tense, case, demonstratives, pronouns, and the like. Each of these systems has limited size, with each term in the system being specifiable as—and having meaning as—the complement of the others. For example, I am thinking of a demonstrative in English; it is not *this*, *these*, or *those*—what is it? It must be *that*, the other term from the four-term system of demonstratives in English {*this*, *that*, *these*, *those*}. A term in a grammatical system can be realized through an affix (or some other morphological process), or as a clitic, or as a separate word, with the mode of realization being no reflection of its status in the grammar. *The* (almost always a clitic), *this* (always a free word), *to* (generally a clitic), *on* (a free word), and *-ing* (an affix) are all fully grammatical elements in English. Their meanings and functions are fully described within the grammar.

Lexemes fill slots in the grammar but this provides only partial specification. In English, *laugh* is an intransitive verb and *hit* a transitive one. But hundreds of other lexemes have the same grammatical specification as *laugh* (*weep*, *talk*, *yawn*, *hop*, and so on and so forth) and hundreds have the same specification as *hit* (*cut*, *eat*, *choose*, *take*, and on and on). The **lexicon**—called a **dictionary** when arranged alphabetically—distinguishes the meanings of lexemes by stating their reference. The transitive verb *peel* refers to taking off the separable, non-rigid outer layer of something (this may be bark, rind, or skin). *Shell*, also a transitive verb, refers to taking off a rigid outer layer

(of a nut, an egg, or peas). This layer may be called a *shell* (but note that not all objects with shells can be shelled—not a tortoise, nor an oyster, nor an explosive projectile). Each lexeme has a referential meaning, in terms of its relation to some real or imagined thing in the world. Each grammatical item has a different type of meaning, first in terms of its role in the grammar and secondly—for some but not all grammatical items—in terms of its reference (for example, masculine gender refers to people of the male sex, past tense inflection to time in the past).

As said before, a dictionary need not include grammatical items, since these are completely described in the grammar. There is a general misapprehension that a dictionary should include all ‘words’, whether lexical or grammatical. Or rather, there is lack of appreciation that there is a distinction between the two kinds of word. When a grammatical item is included in a dictionary it is treated like a lexeme; that is, as a singleton, without attention to the system it is a member of (look up *that*, and I’ll bet that your dictionary does not contrast it with *this*). Some dictionaries even include affixes. They try—unsuccessfully—to do the job of a grammar within a lexicon. In fact, each grammatical item must be considered as part of a system, and each system must be considered within the context of the grammar as a whole.

The idea that everything written as a word is lexical can be pervasive. In the early 1950s, Morris Swadesh suggested that ‘core vocabulary’ gets replaced at a slower rate than non-core items, and produced 100- and 200-word lists of core items by means of which the genetic relationship between two languages could be seen almost at a glance. This method of ‘lexicostatistics’ was deeply flawed and has long been discredited, but Swadesh’s lists are still used by some linguists. One trouble is that they are not *lexical* lists. Included are items such as ‘I’, ‘we’, ‘thou’, ‘ye’, ‘he’, ‘this’, ‘and’, ‘if’, ‘at’, ‘in’, and others that are indubitably grammatical.

The distinction between grammar and lexicon is a natural one, made in all languages. It is brought out particularly clearly in the avoidance style of Dyirbal (called ‘mother-in-law language’ by native speakers). In the presence of a taboo relative—mother-in-law, son-in-law, father-in-law, or daughter-in-law—one must employ a special speech style, called Jalnguy. This has exactly the same phonology and grammar as the everyday style, Guwal; all affixes are the same, and also all grammatical words such as pronouns and demonstratives and *gulu* ‘not’, *yamba* ‘might be’, and *biri* ‘could have happened but didn’t’. All proper names have identical form in the two styles. But lexemes are different—common nouns, adjectives, verbs, and time words such as ‘tomorrow’. (It appears that there are just four common nouns which do not vary—the four grandparent terms; for example, *babi* is ‘father’s mother’ and the reciprocal grandchild term, in both speech styles.)



Here is a short sentence in Jalnguy, and the corresponding sentence in Guwal:

- (1) Jalnguy: *giña-n=bi*                      *maṅgay-maṅgay*                      *gulu*  
 Guwal: *giña-n=bi*                      *wuygi-wuygi*                      *gulu*  
 THIS-FEMININE=TOO    REDUPLICATION-old.person    NOT  
**jayma-n**  
**wuga-n**  
 give-PAST  
 These old ladies (lit. feminine old people) were also not  
 given [any]

It will be seen the demonstrative *giña-* ‘this’, feminine suffix *-n*, clitic *=bi* ‘too’, the morphological process of reduplication (marking plurality), negator *gulu*, and past tense suffix *-n* are the same. It is just the noun *wuygi* ‘old person’ and verb *wuga-* ‘give’ that have different forms in Jalnguy, becoming *maṅgay* and *jayma-* respectively (these are shown in bold).

There is a fairly general convention—followed throughout this book—to use SMALL CAPS to gloss grammatical elements and lower case for lexemes in interlinear glosses. It will be seen that all the SMALL CAPS material is the same in Jalnguy, and items with lower-case are different. That is, grammatical elements are the same but lexemes differ.

Interrogative words—such as ‘who’, ‘what’, and ‘where’—have a special status in every language. Each relates to a particular word class, but they are linked together by their interrogative meaning and by functioning in content questions. Word class membership differs between languages. In Dyirbal, ‘where’ belongs in the grammatical system of locationals, with ‘here’ and ‘there’, and ‘who’ inflects like the closed system of pronouns. But ‘what’ inflects like members of the open class of nouns. Interestingly, ‘where’ and ‘who’ are like other grammatical items in having the same form for Guwal and Jalnguy, while ‘what’ patterns with lexemes in having different forms:

|                      |              |       |                                 |
|----------------------|--------------|-------|---------------------------------|
|                      | ‘where (at)’ | ‘who’ | ‘what’                          |
| Guwal }<br>Jalnguy } | wuñjay       | waña  | { <b>miña</b><br><b>mindirr</b> |

We can now focus on grammar. (There is further discussion of the lexicon in Chapter 8.) The status of **phonology** with respect to grammar varies in different traditions. For many European scholars, phonology is looked on as an integral part of grammar: ‘A grammar of X’ will include one or more chapters on phonology. For others, often those in North America, phonology is something apart from grammar; they will write ‘A phonology and grammar of X’. The difference is terminological, and of little academic consequence;

but it should be noted that it exists. Phonology is briefly discussed in Chapter 7.

As stated before, **word**—dealt with in some detail in Chapter 10—is a basic unit of any grammar. A grammar has two components—**syntax**, which reaches upward from the word, up into clause, the other basic unit, and beyond; and **morphology**, which looks down into the structure of the word. We can now examine some of the terminology associated with morphology and then with syntax.

## 5.2 Morphology

As outlined in §3.13, the basis of a word is a lexical **root**, to which various **morphological processes** may apply. The processes are compounding (in which two roots are combined), reduplication, shift of stress or change of tone, internal change, subtraction, and affixation.

Lexical roots can be categorized as ‘free’ or ‘bound’ (all affixes are, by their nature, bound):

- **Free form**—may function as a word without any morphological processes applying; generally extended to forms which must be subject to an inflectional system, one of whose choices has zero realization. For example, in many languages, nominative or absolutive in a nominal case system, or imperative in a tense-mood system, has zero marking.
- **Bound form**—cannot function as a word on its own; some morphological process (with non-zero realization) must apply.

In Latin almost all roots are bound; for example adjective root *bon-* ‘good’, part of whose paradigm is illustrated in §3.13. In English, all lexical forms—nouns, verbs, adjectives—are free. In other languages, roots in some word classes may be free and those in other classes bound.

The most common type of morphological process is affixation, involving the addition to a root of suffix, prefix, circumfix (combination of prefix and suffix), or an infix inserted into the root (see 6 in §3.13). When just **affixes** are involved, a word can be segmented into components, as in English:

un-friend-ly

This consists of lexical root *friend* plus two grammatical elements, prefix *un-* and suffix *-ly*. Generally, each of these three elements will be called a **morpheme** (typically defined as a ‘minimal meaningful unit’). However there are some linguists (particularly, at one time, in the French tradition) who use the term morpheme solely for grammatical elements, not for the root.

Since affixation is the most common process, some linguists regard it as the canonical scheme of word formation, treating other morphological processes as odd variants of it. This is a simplistic and unhelpful attitude. It is scarcely revealing to talk of a ‘morpheme of tone change’ or a ‘morpheme of subtraction’ (according to ‘morpheme’ a wider meaning). These are processes, and to try to interpret them as similar to segmental units confuses the nature of grammar.

A grammatical process may have different realizations in different circumstances. The variants of an affixal process are called **allomorphs**. These may be phonologically conditioned, as are /-z/, /-əz/, and /-s/ variants of plural inflection in English (§4.2). Or they may be essentially ad hoc, relating to specific lexemes (this is often called ‘morphological conditioning’ or ‘lexical conditioning’); for example plural *-ren* just on *child* and *-en* on *ox* in English.

### 5.3 Inflection and derivation

It is often helpful to distinguish between inflectional and derivational processes. However, this is not a useful distinction for all languages, and should only be invoked when it sheds light on the operation of a grammar. As outlined in §3.13, we have:

1. underlying **root(s)**
2. optional **derivational process** may apply to it, forming a **stem**

Step 2, applying a derivational process, may operate several times. After all derivational processes have applied:

3. obligatory **inflectional process**, forming a **word**

These steps were illustrated in §3.13, for *de-central-iz-ation* and *baby-sitt-er-s*. Properties distinguishing inflection and derivation include the following.

1. There may be a number of derivational processes applying (or there can be none); all are optional. There will only be one inflectional system, and it must be applied. The nature of the inflectional process is determined by the word class of the stem after the last derivation has applied. An inflectional process may involve a grammatical system one of whose terms involves zero marking. But this zero always carries a meaning; for example, one choice in the number inflection of count nouns in English has zero form, and it indicates singular number.
2. Since an inflectional process applies last, if it is realized by an affix this will be on the rim of the word—the last suffix or the first prefix.

3. An inflectional process must always involve a closed grammatical system, and the nature of the system associated with each word class varies from language to language. In English, count nouns inflect for number; in Dyirbal all nouns inflect for case (in this language number is shown by derivational processes).

Some derivational processes involve mutually exclusive choices from small grammatical systems; for example ‘do while going’ and ‘do while coming’, mentioned in §1.11. But if neither the ‘going’ nor the ‘coming’ derivation applies, this does not indicate that neither going nor coming is involved, simply that no specification has been made concerning this. The system applies optionally. However, most derivational processes are not within one system.

Several derivational processes may apply within a word, often in fixed order. Consider *un-friend-ly* in English, an adverb derived from noun *friend* by attaching prefix *un-* and suffix *-ly*. Which process applies first? Well, there is a word *friend-ly* but no word \**un-friend*. We infer that first of all *-ly* is added, forming an adverb *friend-ly*, and then the negative prefix *un-*, which does not affect word class. Sometimes derivational processes may apply in variable order—with different meanings according to which process is within the scope of which other process. For example, nominal derivational suffixes in Dyirbal include *-gabun* ‘another’ and *-jarran* ‘two’. Adding these, in either order, to the root *yara* ‘man’, we get *yara-gabun-jarran* ‘two other men (that is, any two men in addition to those already referred to)’ and *yara-jarran-gabun* ‘another two men (that is, another pair of men, where all the men are arranged in pairs)’. Case inflection will follow the last derivational suffix. (Further examples of alternative orderings are in Dixon 1972: 232–3.)

(Note that *yara-gabun-jarran* and *yara-jarran-gabun* each make up a single grammatical word, on the following criteria. First, the three components must occur in fixed order, whereas words may occur in any order within a sentence. Secondly, a case inflection is added to each word in a phrase, and here it is only added at the end of *yara-jarran-gabun* and of *yara-gabun-jarran*.)

4. A derivational process may change word class; for example, verb stem *central-ize* from adjective root *central*. Or it may not—verb stem *de-central-ize* from verb stem *central-ize*—then simply adding a semantic element (here *de-* indicates reversal of process). An inflection cannot change class.
5. A derivational process is often applicable to only one word class, but not infrequently it can apply to more than one. For example, in English derivational suffix *-y* forms adjectival stems from both nouns and verbs, as in *brain-y* and *choose-y*. An inflection is added to the stem of one

specific word class. The fact that it takes a certain inflection is a defining characteristic of a word class.

6. Because of their grammatical status, inflectional processes generally have a higher frequency than derivational processes. As a consequence, inflectional affixes tend to have shorter forms than derivations; for example, they are more often monosyllabic, or consisting just of a syllable-closing consonant.
7. Each derivational process is distinct from the others; derivational suffixes are not likely to fuse together. In contrast, inflectional suffixes often fuse with the preceding stem. And inflectional suffixes are often themselves fused forms, as in Latin where the inflection on an adjective combines information about gender, number, and case. Note that since a word may take only one inflection, this *has to be* a fused form if it is to carry information concerning several grammatical categories.

It has been suggested that inflection is always subject to ‘agreement’ (see §5.6) and derivations never. Neither suggestion is supported by the facts of languages. There is certainly a tendency for the modifiers to a noun, say, to agree with it in inflectional categories. But in some languages a case inflection goes only onto the last word in an NP (whatever this may be), or just onto the first word, and not onto other constituent words in the phrase. And there are examples of derivations showing agreement. This can be illustrated with derivational suffix *-bila* ‘with’ in Dyrirbal. In the sentence ‘Give it to the man with a big stick’, the beneficiary NP would be:

|                             |                         |                       |
|-----------------------------|-------------------------|-----------------------|
| [yara-gu                    | [yugu-bila-gu           | bulgan-bila-gu]]      |
| man-DATIVE                  | stick-COMITATIVE-DATIVE | big-COMITATIVE-DATIVE |
| to the man with a big stick |                         |                       |

Comitative suffix *-bila* goes on noun *yugu* ‘stick’ and also onto its modifier *bulgan* ‘big’; these make up a phrase within a phrase—*yugu-bila bulgan-bila* ‘with a big stick’ modifies *yara* ‘man’, the head noun of the beneficiary NP. The function of this NP in the sentence is shown by dative inflection *-gu* being added to each of its constituent words. There is thus agreement involving derivational suffix *-bila*, and also involving dative inflection *-gu*.

Information about core arguments can be fused within inflectional affixes, as in Latin where verb endings show person and number of subject, combined with information on tense, mood, and voice. But, more generally, it is often not helpful to try to classify systems of bound pronouns as inflections or derivations.

In his grammar of Sarcee, an Athapaskan language, Cook (1984: 120) comments: ‘the distinction between the derivational...process and the

inflectional . . . process is by no means clear.’ Similar remarks apply to quite a few other languages, including a number from South America (among others, the Tucanoan, Arawak, and Arawá families). In Jarawara, for instance, there are many grammatical systems realized through suffixes—mood, tense, evidentiality, modality, negation, etc.—but all are optional. It is not at all helpful to try to categorize these as derivational or inflectional.

And here lies the joy of basic linguistic theory. A theoretical apparatus is provided which can be drawn upon *as needed*, when it aids description and explanation. But it *does not have to be* employed, if not appropriate for a language. In contrast, some formalists maintain that since a distinction is valid for European (and other) languages, it should be universal. Pity a poor student, trying to describe a language from South America—for which the inflection/derivation dichotomy is inappropriate—and being told that each affix must be characterized in these terms. It *can’t be done*; there are no applicable criteria. Attention is diverted to this instead of being directed towards the unusual and fascinating syntactic properties of the language, quite unlike anything found in more familiar tongues.

## 5.4 Clitic, affix, and adposition

A **clitic** is a surface element part-way between a word and an affix in properties. Sentence (1) from Dyirbal, in §5.1, includes clitic =*bi* ‘too, in addition to what has already been said’. (It is convenient to use ‘=’ for a clitic boundary, and ‘-’ for a boundary between root and affix or between two affixes.) The following properties distinguish clitics from affixes:

1. An affix is added to a root or stem, the whole functioning as one grammatical word (and generally also as one phonological word—see Chapter 10).

A clitic is a separate grammatical element, often to be regarded as a separate grammatical word. But it cannot stand alone; that is, it cannot make up a phonological word by itself. Instead, a clitic must ‘lean’ (its name comes from the Greek *klin-* ‘lean’) on a full word which is its **host**. A **proclitic** precedes and an **enclitic**—like =*bi* in (1)—follows the host. (The term ‘enclitic’ was first used in English about 1660 and ‘proclitic’ around 1840. ‘Clitic’ appears to have been back-formed from these names by Eugene A. Nida in 1946, in the first edition of his classic text *Morphology*.)

2. An affix is an integral part of the word to which it belongs, for purposes of stress placement. Phonological rules such as assimilation

and elision will apply across a root–affix boundary or an affix–affix boundary.

In contrast, a clitic is generally just ‘added on’ to its host and is unlikely to count as part of the word for purposes of stress assignment and phonological rule application.

3. An affix is always restricted in its application, sometimes just to forms from one word class, sometimes to more than one (but never to all possible stems). A clitic is typically omni-locatable and can be added to any of a wide range of words (often, to a word of any type). For example, in my Dyrirbal corpus, =*bi* is added to demonstratives, as in (1), to nouns, to pronouns, to time words, to verbs, and also to grammatical operators such as *gulu* ‘not’.
4. As illustrated in the last section, derivational affixes occasionally enter into agreement, occurring on several words in a construction, and inflectional affixes typically do. Clitics never do.
5. We noted that an inflectional process will apply after all derivational processes so that an inflectional affix will always be on the rim of a word. A clitic is added after derivational and inflectional processes are completed so that a proclitic precedes the first prefix and an enclitic follows the last suffix.

These are indicative properties. Aikhenvald (2002b) provides a definitive characterization of clitics, dealing with fifteen parameters of variation, and taking account of all previous literature on the topic. She notes exceptions to some of the properties just listed. In essence, the criteria for recognition of clitics must be formulated on an individual basis within the grammar of each language, if there appears to be a unit intermediate in profile between word and affix.

It is important to clearly distinguish (and never to confuse) the following:

- (a) Grammatical systems; for example, case, tense, pronouns, gender, definiteness, demonstratives.
- (b) Realization of these, as affix, clitic, or separate grammatical word.
- (c) Positioning of the realization; for example, attached to each word of a phrase, or just to some specified word.

There are some *expected* dependencies between these three parameters. For example, a term in a case system which is realized as a clitic or word is unlikely to be associated with every word of an NP, whereas an affix may be. A clitic with scope over a whole clause—for instance, as marker of a polar question—may have a definite position within the clause; often, as an enclitic onto the last word of the first phrasal constituent, or else right at the end of the clause.

But there are no *necessary* restrictions. For example, languages are known in which a case affix must (or may) attach to every word in an NP, or just to the head, or just to the last word (whatever that may be) or just to the first word (Dixon 2002: 142–5 lists Australian languages showing all of these types of marking).

The realization of a grammatical category is to be classified as an affix or clitic not on the basis of its semantic content, or because of where it occurs, but simply in terms of whether or not it satisfies criteria for an affix or clitic. In English, the plural term from the number system and the marker of a possessive construction have totally different statuses in the grammar. Yet they have identical realization, with the same phonological conditioning of suffixes /-z/, /-əz/, and /-s/ (see §4.2). It is irrelevant that plural -s attaches to the head noun of an NP and possessive -'s to the last word of an NP. The two markers have identical surface profiles; each is clearly an affix.

Some linguists muddle the parameters, saying things like: 'if a case marking is added to the end of a phrase it must be a clitic'. But a case choice indicates the *function of an NP* in a clause. It is surely natural that it should be shown once in the NP, and where better than at the end? The less natural situation is for case to be marked on every word of the NP; we then have to say that an affix which applies to the NP as a whole is—rather pedantically—added to every word within the phrase. (However, this is likely to interrelate with some other property of the grammar, such as allowing words to occur in any order in a clause. One can then see which words belong together, as making up an NP, in terms of them all showing the same case marking.) It is perfectly possible that there would be a language all of whose dialects have the same process for marking accusative case. But in one dialect accusative goes onto the first word of an NP, in the second dialect onto the last word, in another onto the head, in another onto the head and onto a demonstrative modifier but not onto an adjective modifier, and in a further dialect onto every word in the NP. In each dialect the accusative marker is added to a stem and forms one coherent phonological word with it. It has in each instance the properties of an affix. Surely it would not be sensible to say it is an affix in some dialect(s) and a clitic in other(s)?

'Words' can be misleading. There is a temptation to treat each 'word' in the same way; for example, insisting on a dictionary entry for every 'word' (or for every item which looks like one). Some words are lexemes (or have a lexical root as their core). Some, like pronouns and demonstratives, belong to a closed grammatical system. And some simply mark the function of a constituent in a clause or NP. The latter are termed **prepositions** (a term adopted into Old English from Latin and Greek grammars) if they precede the constituent they mark, or **postpositions** (an analogic creation in the nineteenth century) if they



follow. Rather recently, the term **adposition** has been coined as a cover term for prepositions and postpositions.

A case system is used to mark the function of a core argument (S, A, O) and/or to mark spatial, temporal, and other relations on peripheral arguments. A few languages (well-known examples are Finnish and Estonian) have more than a dozen cases, distinguishing things like ‘towards the outside of’ and ‘towards the inside of’. But most case systems are smaller. They are often combined with some other grammatical mechanism. One recurrent means is a possessed noun with spatial orientation, within an NP; for ‘on the box’ one would say ‘at the top of the box’ (see §16.5).

In the classical Indo-European languages (and in some modern ones, such as Russian and German), cases are combined with prepositions. For example, with noun roots *patria* ‘fatherland’ and *Rōma* ‘Rome’ we get in Latin:

|                               |                                          |
|-------------------------------|------------------------------------------|
| ABLATIVE CASE FORM            | <i>patriā</i> ‘from the fatherland’      |
| PREPOSITION <i>ex</i> PLUS    |                                          |
| ABLATIVE CASE FORM            | <i>ex patriā</i> ‘out of the fatherland’ |
| ACCUSATIVE CASE FORM          | <i>Rōmam</i> ‘to Rome’                   |
| PREPOSITION <i>intrā</i> PLUS |                                          |
| ACCUSATIVE CASE FORM          | <i>intrā Rōmam</i> ‘within Rome’         |

This is a fascinating, double-barrelled method of marking. One can use just a case (of which there are half a dozen) or a case plus a preposition (of which there are a couple of score). (Noun stems in Latin are bound forms and must occur with a case-ending, so in this language a preposition cannot be used without case.)

The grammatical elements shown by case inflections and by prepositions (or postpositions) have the same status within the grammar. The difference is almost always just a matter of realization. In summary:

MARKERS OF GRAMMATICAL (INCLUDING SPATIAL AND TEMPORAL)  
FUNCTIONS

|                    |                       |                         |
|--------------------|-----------------------|-------------------------|
| <i>realized as</i> | <i>ideally called</i> | <i>typically called</i> |
| affix              | case inflection       | case inflection         |
| clitic             | case clitic           | case or adposition      |
| grammatical word   | case word             | adposition              |

The terms ‘preposition’ and ‘postposition’ are not strictly necessary (one could just say ‘marker of grammatical function’); they would probably not have come into use if our linguistics had not grown out of early grammatical work on Greek and Latin. And they can be more than a little confusing, as when we find statements such as ‘preposition *ex* “governs” ablative case’.

It is better to say that *ex patriā* simply has double marking, by case combined with preposition. The fact that the preposition is a word and the case a morphological process (it could be regarded as an affix fused with the root) does not mean that the preposition has any greater importance as a grammatical element. Recall a major theme of this volume—we are aiming to describe the *underlying* grammatical system of a language, not to just analyse its *surface* structure. (The confusing notion of ‘govern’ is further discussed in §5.6.)

‘Preposition’ and ‘postposition’ are well-known terms, and they are useful if employed in a careful manner, with the realization that they code grammatical elements. They do vary, from language to language, in character, number, and function. In Latin and Greek (and even more in Russian), a preposition is not necessarily associated with a single case. There may be two possibilities and each will, of course, have a different meaning. For example, in Latin:

preposition *super* plus accusative case: ‘motion over’

preposition *super* plus ablative case: ‘at rest over, upon’

Many languages do not have a system of case inflections and just employ prepositions or postpositions. That is, they show syntactic function not by a morphological process but by clitics and/or words. In such a language, adpositions could sensibly be called ‘non-inflectional case markers’. After all, the grammatical category of tense may be shown by affixation or another morphological process, or by clitics, or by grammatical words. There is no reason why case should not be treated in the same way.

A non-affixal grammatical marker may be a separate word. But often, although written as a word, it is (or can be) a clitic. In English, for example, subject and object pronouns may be separate words but are much more likely to have clitic form; *You hit him* will be pronounced /jə=hit=im/. Similarly with articles *the* and *a(n)*, and with conjunctions *or*, *but*, and *and*. The phrase *pears and apples* is most likely to be pronounced /pɛəz n=æplz/. English has many prepositions; ten of the most common monosyllabic forms are generally realized as clitics (these are *for*, *of*, *to*, *at*, *from*, *till*, *than*, *as*, *by*, and, for some speakers, *upon*). The phrase *to the beach* is pronounced as one phonological word, /tə=ðə=bɪ:tʃ/. (Other monosyllabic prepositions, and all the polysyllabic ones, are distinct phonological words—*up*, *in*, *on*, *down*, *through*, *below*, *beneath*, and so on).

Preposition and postposition are, like affix and clitic, means for surface realization. They must not be confused with underlying grammatical systems. There is further discussion of them in §5.6.

## 5.5 Morphological types

A number of terms are in use for describing the morphological make-up of a language. The main ones are, in alphabetical order: agglutinating, analytic, fusional, inflectional, isolating, polysynthetic, and synthetic. In fact they belong to two distinct sets which should not be muddled together (although they often are):

Set A (labels have Latin origin)

- **Isolating.** Each word consists of one morpheme. That is, each lexical root is a word and each grammatical element is shown as a distinct word (not as an affix). Vietnamese is often cited as a canonical example of this type.
- **Agglutinating** (or agglutinative). A word consists of a number of morphemes (roots and affixes) but is fully segmentable; that is, one can position a hyphen between root and affix and between each affix. Languages from the Turkic and Bantu groups are of this type.
- **Fusional.** A word includes a number of grammatical elements but their realizations are not segmentable, being instead fused together (and they may also be fused with the root). This type, of which Latin and Greek are familiar members, was originally called ‘inflectional’ or ‘inflective’ or ‘inflecting’, all of which are inappropriate since agglutinating languages may also involve inflection. Recently, ‘fusional’ has been introduced as a clearer label.

Like much work in linguistics, this typology seems to assume that the only morphological process is affixation. Sapir—within an exemplary discussion of sets A and B in chapter VI of his 1921 masterwork *Language*—notes this point and posits a fourth type of language, utilizing ‘internal changes (reduplication, vocalic and consonantal change, changes in quantity, stress and pitch)’. Rather oddly, he terms this ‘symbolic’, but the name has not caught on.

‘Isolating’, ‘agglutinating’, and ‘fusional’ are idealizations. No language exactly fits one type but is always a mixture, for example, ‘basically agglutinating but with some fusion in the nominal word’.

Set B (labels have Greek origin)

- **Analytic:** smallish number of components—root(s) plus grammatical elements—per word
- **Synthetic:** largish number of components per word
- **Polysynthetic:** very large number of components per word

The two typologies intersect at the bottom; a highly analytic language will be isolating. But from there on up they are basically orthogonal. Sapir

(1921: 142–3) describes a number of combinations of value from Set A and Set B, from which we can extract:

|               |  |           |                        |
|---------------|--|-----------|------------------------|
| A             |  |           |                        |
| POLYSYNTHETIC |  | Yana      | Algonquin              |
| SYNTHETIC     |  | Turkish   | Salinan                |
| ANALYTIC      |  | Chinese   |                        |
|               |  | ISOLATING | AGGLUTINATING FUSIONAL |
|               |  | B         |                        |

(Yana and Salinan are languages of southern and northern California respectively.)

As already mentioned, each of ‘isolating’, ‘agglutinating’, and ‘fusional’ specifies a type of language. But ‘analytic’/‘synthetic’ describes a continuum. A language may be more or less analytic (and is then less or more synthetic) than another. What then is the difference between synthetic and polysynthetic? Can a line be drawn, and if so where?

Languages of a highly synthetic nature may include all sorts of grammatical elements in one word (typically, in a verbal word). There can be bound pronominal markers of subject and object (and sometimes also of a third argument such as indirect object). There may be valency-changing elements such as causative, applicative, passive, antipassive, reflexive, and reciprocal. Incorporated nominal roots of various kinds (relating to the S or O argument and/or to an instrument). Perhaps several verb roots in sequence. Markers of tense and aspect and evidentiality, and of mood. Markers of direction. Various kinds of adverbial-type elements such as ‘in the morning’ and ‘in the evening’ (illustrated at the end of §4.8 from Tiwi, one of the most strongly synthetic languages in Australia).

But no language has every one of these components in a single word. What does it *have to have* to qualify as ‘polysynthetic’? There may be a dozen grammatical elements crammed into one word, but not including bound pronouns; some linguists maintain that having bound pronouns is a *sine qua non* for a language to be admitted to the polysynthetic club. Another language may have extremely complex words but no incorporation of nominal roots. But nominal incorporation has—by some linguists, although not by others—been suggested as a criterial property to be called polysynthetic.

An analogy may help. Life-styles can be ranged along a continuum from penurious to luxurious. The most highly luxurious would involve many of the following: a private plane, a yacht, a house in the country as well as one in town, one or more Mercedes or BMW cars, a chauffeur plus a number of other servants, travelling first-class, and staying in five-star hotels. Suppose we

wanted to define what it takes to be ‘poly-luxurious’ (to combine Greek prefix with Latin root). What would be considered criterial? How about a plutocrat with a yacht and a jet plane but only one car and that a Honda, which they drive themselves. Would this qualify as ‘poly-luxurious’? Or a magnate of small build with absolutely everything but they prefer always to travel economy (or coach) class? Would they qualify? As can be seen, such a discussion is unproductive and not at all worthwhile.

Similarly for the distinction between ‘synthetic’ and ‘polysynthetic’. Typologies A and B are, in truth, of limited interest and usefulness. But since B is a continuum, labels are needed just for the two poles—‘analytic’ and ‘synthetic’. No useful purpose is served by distinguishing between degrees of complexity towards the synthetic end. It is sufficient to say that Yana, Algonquin—alongside Tiwi, Eskimo, and other languages—are highly synthetic, but all in different ways. The term ‘polysynthetic’ is unnecessary and also a distraction.

## 5.6 Syntax

As described in §§3.1–2, **clause** is the central unit of syntax. (The importance of not using the label ‘sentence’ when ‘clause’ is intended was stressed under (a) in §2.5.) Each clause has a structure consisting of a **predicate**—which may be filled by a copula verb, or may be zero in a verbless clause (see Chapter 14)—and a number of **arguments**. In logic, the term ‘predicate’ takes in everything in a clause except the subject; the term is here used in the normal linguistic sense, to refer to the verb or equivalent and its modifiers, not including any argument; see (b) and (f) in §2.5. The number and nature of the arguments is determined by the nature of the predicate.

‘**Construction**’ is a useful term for referring to a type of clause with certain properties; for example, a ‘**complement clause construction**’ has a complement clause (rather than an NP) filling one of the core argument slots in clause structure, as in  $I_A \text{ know } [that \text{ John did it}]_O$ . One can also talk of a construction at the sentence level; in English a conditional construction has a clause marked with *if* combined with a main clause, as in *If it doesn’t rain, we’ll come tomorrow*. And also of a construction at phrase level; for instance a possessive construction, such as *[the King of Spain]’s new hat*.

Each clause has **core argument(s)**, which are obligatory; that is, they must either be stated or understood from the context. It may also include optional **peripheral arguments** (these are sometimes referred to as ‘adjuncts’). An **intransitive construction** has a single core argument, which is said to be in **intransitive subject function** (abbreviated as S). For some intransitive predicates, the referent of the S argument will initiate and/or control the activity (for example, ‘jump’); for others it will not (for example, ‘yawn’). A **transitive construction** has two core arguments. That whose referent does or potentially

could initiate or control the activity is said to be in **transitive subject** function (abbreviated as **A**). The other core argument, whose referent is often affected by the activity, is in **transitive object** function (abbreviated as **O**).

In some languages, S and O arguments (which may be marked by absolutive case) pattern together, differently from A (which may be marked by ergative case). In others, S and A (nominative case) pattern together, differently from O (accusative case). Even in ergative languages, S and A share a number of properties—as addressee in imperative constructions, as controller of reflexive, and so on (see Dixon 1994: 131–42). **Subject** is simply the association of S, the only core argument of an intransitive clause, and A, that core argument in a transitive clause which could initiate or control the activity.

The major discussion of transitivity is in Chapter 13. But we can here briefly draw attention to an **extended transitive**, or ‘ditransitive’, construction (mentioned in §3.2), with three core arguments. As with a simple transitive, that argument whose referent could control or initiate the activity is in A function. But which of the other arguments is to be in O function? Consider a typical extended transitive verb, ‘give’, which has three semantic roles: Donor, Recipient, and Gift. Donor is always A. In some languages, including English, there are two constructions available, one in which Gift is O—for example *Mary<sub>A</sub> gave [the rabbit]<sub>GIFT:O</sub> to John*—and the other in which Recipient is O—*Mary<sub>A</sub> gave John<sub>RECIPIENT:O</sub> [the rabbit]*. There is further discussion of this in §13.5.3.

A clausal argument may be realized (in some languages) by a bound pronoun and (in all languages) by a **noun phrase** (NP). In its simplest form this is a noun, plus a number of optional modifiers, such as demonstrative, possessive phrase, adjective(s), relative clause, spatial phrase, temporal phrase, etc. A **verb phrase**, filling predicate slot, is similar—generally a verb, plus a number of optional modifiers such as adverbs. (Noun and verb phrases were discussed in §3.4.)

One item in each phrase will be its **head**. Generally, the head is the only obligatory component and may make up a complete phrase on its own. It is the head which dictates agreement on other items in the phrase, and it is the head which determines the properties of the NP as a whole. This can be illustrated with gender. If the head of an NP is a noun of feminine gender, then if demonstratives, adjectives, and other modifiers within the NP show gender, they will be feminine, in agreement with the head. And the whole NP will count as feminine, perhaps for agreement with the predicate.

What the head is in a possessive construction can be a fascinating question. In English the answer is straightforward—the item possessed is head, both in alienable possession, such as *Mary's car*, and in the inalienable variety, such as *Mary's foot*. (The head is underlined.) But, as we showed in §1.3,

many languages have more than one possessive construction. In Jarawara, for example, the possessed noun is head in an alienable construction:

|                   |            |                 |
|-------------------|------------|-----------------|
| <i>mati</i>       | <i>kaa</i> | <i>jomee</i>    |
| mother (feminine) | POSSESSIVE | dog (masculine) |
| mother's dog      |            |                 |

Here the whole NP is masculine, taking the gender of its head; in Jarawara, tense and mood suffixes in the predicate agree in gender with the NP which is in pivot function.

For inalienable possession, involving a body-part term, we get:

|                   |                 |
|-------------------|-----------------|
| <i>mati</i>       | <i>tame</i>     |
| mother (feminine) | foot (feminine) |
| mother's foot     |                 |

Here the inalienable possessor is head of the NP. It dictates the gender of the inalienably possessed noun, so that the feminine form of 'foot', *tame*, is used (the masculine form is *teme*, as in *jomee teme* 'the dog's foot'). And the whole NP counts as feminine for tense and mood agreement in the predicate.

The established use of the term 'head' is to refer to a phrase. Recently, a new use has come into play, dubbing the predicate as head of a clause. None of the criteria given for being head of a phrase are applicable here. The predicate does determine how many and what sort of arguments there are in its clause, but it does not engender agreement with arguments (rather, the reverse). And—save in languages that include bound pronouns within the predicate, and some isolating languages with extreme ellipsis—the predicate may not make up a clause on its own, in the way that the head of an NP or of a verb phrase can generally make up a whole phrase on its own. It should be borne in mind that this custom of calling the predicate the 'head' of a clause is a markedly different sense of 'head'.

Another, quite different use of 'head' is for the fullest statement of the common argument in a relative clause construction. The unsatisfactory nature of this further sense is discussed in §17.2.

The terms **modifier** and **qualifier** are straightforward (they are basically equivalent). A non-head item in a phrase modifies/qualifies the head, and can be called a modifier/qualifier. **Concord** and **agreement** are also, in most cases, synonymous. A modifier in a phrase may agree with its head (say, in gender). There is then gender concord between them. All of the words in an NP may agree in case (this is a category associated with the NP as a whole). There can also be concord between an argument and the predicate, as in Jarawara where mood and tense specifications within the predicate agree in gender with the pivot argument.

The terms **govern** and **government** are fraught with difficulty. Neither is at all necessary; indeed, linguistics would be better off without them. There are quite a number of ways in which the terms are used. One of the most common is to say that a verb such as *cut* ‘governs an object’. But this is precisely because *cut* is a transitive verb, requiring a core argument in O function as well as one in A function. If it ‘governs’ its object why should it not also be said to govern its subject? In a language where an O argument is marked with accusative case, it is often said that a transitive verb ‘governs the accusative’. This is conflating levels; one should say that a transitive verb requires an O argument and this may be marked with accusative case.

Prepositions (and postpositions) were discussed in §5.4. We saw that for the Latin phrase *ex patriā* (PREPOSITION fatherland+ABLATIVE) ‘out of the fatherland’, syntactic function and meaning are shown by a combination of preposition and case. The preposition is a word but it has similar grammatical status to case, which is an affix fused with the stem. Working in terms of surface structure, grammarians have said that ‘preposition *ex*’ (which is a grammatical element) ‘governs ablative case’, in the same way that they say a transitive predicate (which has a lexical head) governs accusative case. Taken over into English, which has no cases on nouns, it is typically said that, in the sentence *She<sub>A</sub> took coals<sub>O</sub> [to the fireplace]<sub>PERIPHERAL</sub>*, there are two instances of ‘government’:

verb *take* governs NP *coals*  
 preposition *to* governs NP *the fireplace*

This is highly confusing. And it extends further; a preposition is saddled not only with being a governor, but also a head:

predicate *take* is head of the clause *She took coals to the fireplace*  
 preposition *to* is head of the phrase *to the fireplace* (this is called a ‘prepositional phrase’)

*To the fireplace* is a phrase involving a lexeme (*fireplace*), yet the grammatical marker *to* is said to be head of the phrase. Such statements—although seemingly endorsed by tradition—are confusing and best avoided.

Consider a sentence in Bardi, an Australian language:

- (1) aamba-nim<sub>A</sub> aarli<sub>O</sub> inambuna  
 man-ERGATIVE fish spear+3sg+PAST  
 The man speared a fish

What is the status of the ergative marker *-nim*? It could be an affix and should then be called a case marker. But it might be a clitic, and could then perhaps be



called a postposition. In terms of ideas about adpositions just presented, there is a world of difference between these analyses. If *-nim* is a case then *aamba-nim* is an NP consisting just of a noun, *aamba*, which is its head. If *=nim* is a postposition then *aamba =nim* is a prepositional phrase whose head is *=nim*.

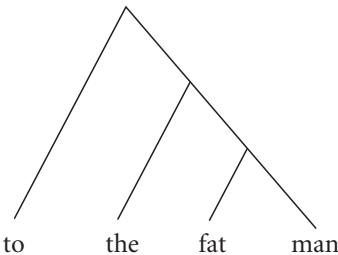
This surely indicates how unsatisfactory it is to treat a grammatical item such as a preposition as if it were a lexical form, just because it is realized as a word. In a more rational approach, *nim* is *always* a case marker—it may be a case affix or a case clitic. As suggested in §5.4, the terms ‘preposition’ and ‘postposition’ are encumbrances, the source of much confusion. If *ex patriā* ‘out of the fatherland’ is—like *to the fireplace*—to be called a ‘prepositional phrase’ then plain *patriā* ‘from the fatherland’ (involving ablative inflection) should surely be called a ‘case phrase’. Both labels are inappropriate.

**Constituent** is a handy term. It refers to anything which fills a slot in a syntactic structure—a clause within a complex sentence (or a relative clause within an NP), a phrase within a clause, a possessive NP (that is, an NP in possessive function) within an NP, an adjective plus modifier (for example, *quite hot* or *very tall*) within an NP, and so on. (It is of course possible—but not terribly fruitful—to apply the notion of constituent within a word; saying, for instance that *un-* and *friend-ly* are constituents of *un-friend-ly*, and then that *friend* and *-ly* are constituents of *friend-ly*.)

In the sentence *John slowly assembled the model*, the sequence of words *John slowly* is not a constituent since it does not fill a slot in clause structure, neither is *assembled the*. Constituents are nicely indicated by bracketing, and labelling the brackets:

[John]<sub>NP:A</sub> [slowly]<sub>SENTENTIAL.ADVERB</sub> [assembled]<sub>PREDICATE</sub>  
 [the=model]<sub>NP:O</sub>

A constituent consists of lexemes, or of referring grammatical elements that can be head of an NP, such as pronoun and demonstrative. This was illustrated in §1.11. It was explained that *to the fat man*, /tə=ðə=fæt mæn/, is sometimes represented as a ‘tree structure’:



This implies that *fat man* is a constituent of *the fat man*, and that *the fat man* is a constituent of *to the fat man*. A more appropriate representation is:

BENEFACTIVE RELATOR<sub>(to)</sub> [*fat man*]<sub>DEFINITE</sub>(*the*)

Lexeme *man* is head of the NP and is modified by lexeme *fat*. The term ‘definite’ is chosen from the grammatical system of definiteness and this is realized through a clitic *the*, /ðə=/. The function of the NP in its clause is marked by clitic *to*, /tə=/.

The principle is that grammatical markers—case, gender, tense, aspect, mood, etc.—are attached to lexemes or combinations of lexemes which make up a constituent. Neither definite marker *the*, /ðə=/, nor plural marker *-s*, nor a case affix, nor a preposition counts as a constituent itself. Following this principle, if a preposition cannot be a constituent of a phrase, it can scarcely function as the head of one.

When people talk of **constituent order** (sometimes mislabelled ‘word order’), they mean the order of core phrasal constituents in a clause—S and intransitive predicate, or A, O, and transitive predicate. In some languages, phrasal constituents can appear in virtually any order; this is called ‘free constituent order’. Often, the predicate must be first in a clause (as in Fijian) or last (as in Jarawara) but NPs in A and O function may occur in either order after or before it (see §2.4). Then there are languages with free **word order**, when words from different phrases may be scattered throughout a clause (sometimes, throughout a multi-clause sentence).

English has a fairly fixed order of phrases in a clause and of words in a phrase. The most peripatetic constituent is a type of sentential adverb; for example, *accidentally* can go at any of the places shown as *A* in *A Mary A spilt the milk A*. But not all of English’s lexical constituents have consecutive realization. We saw under (a) in §1.10 that a phrasal verb such as *hand over* is mapped discontinuously within clause structure—*John<sub>A</sub> handed [the documents]<sub>O</sub> over*. The second part of the lexeme, *over*, can be moved to the left over a full NP in O function (*John handed over the documents*) but not over an O consisting just of a pronoun (*John handed them over* but not \**John handed over them*). *Hand—over* is a complex lexeme consisting of two words, verb root *hand* and *over*, which has the form of a preposition but does not in this instance function as a preposition. (It is convenient to refer to it as a ‘preposition’ here, although this is not strictly accurate.)

Languages differ in how much freedom of ordering they allow, for phrasal constituents and for words. Working on Dyrbal, I discovered almost unlimited ordering of words—as I recorded and analysed texts, and as I tried to speak the language myself and was corrected when I erred. I often got things wrong and was gently put right, but I was never told off for using a wrong word order. In my 1968 dissertation I quoted a clause in which the constituents were all ordered together:

- (2) [bayi                                      wajal                                      [ba-ŋu-l  
 THERE+ABS+MASC boomerang+ABS      THERE-GEN-MASC+ABS  
 yara-ŋu                                      bulgan-u]POSS.NP]O      [ba-ŋgu-n  
 man-GEN+ABS      big-GEN+ABS                                      THERE-ERG-FEM  
 jugimbir-u]A      [bura-n]PREDICATE  
 woman-ERG      see-PAST

The woman saw the big man's boomerang

I then stated that the words could occur in any sequence and ordered them randomly as:

- (3) bayi yaraŋu jugumbiru buran wajal baŋgun baŋul bulganu  
 The woman saw the big man's boomerang

John Lyons, examiner for the dissertation, demurred at this: 'Surely, Bob, this is going too far?' His query made me doubt my competence (although I had already had sentence (3) checked and confirmed by consultants). On the next field trip, I put sentence (3) to the two main consultants, on separate occasions. Each time, I was gazed at with wonder: 'Why did you have to ask? You *know* that's alright!' Most of the sentences I made up to check some grammatical pattern had a subtle point, and exercised the consultants' minds just as they did mine. But why ask about (3), a matter of surface realization—was I losing my wits or what? (See Dixon 1968: 92; 1972: 107–8.)

Although a constituent, such as the A NP *baŋgun jugumbiru*, may have its words realized at different places in the surface structure of (3), it is still very much a constituent, as shown by the ergative case suffix, *-ŋgu ~ -u*, on its two words. Some linguists have said that in a sentence such as (3)—that is, in a language allowing freedom of word order—there are *no phrasal constituents*. That is, they determine that a phrasal constituent cannot be recognized unless its member words always occur in one block. This involves analysing surface structure, not good practice if one wants to discover how a language works. If an antipassive construction were derived from (2–3), the A argument would go into S function; that is *baŋgun jugumbiru* would become *balan jugumbil*, in absolutive case. This—and many other grammatical properties—show that *baŋgun* and *jugumbiru* make up one constituent. (In the antipassive, *balan* and *jugumbil* could appear next to each other or wide apart; it just doesn't matter.) A constituent, such as *baŋgun jugumbiru*, is a syntactic unit since it fills a slot in clause structure. How it is realized is entirely a secondary matter.

Other important terms were introduced earlier and are discussed in some detail later. They bear a brief mention here. Sentences may be linked by having a shared **topic**, which is a referent of some argument—in some grammatical

function—in each main clause (and sometimes also in certain kinds of subordinate clause). In some languages, the grammar determines which argument in each clause type may function as topic. This grammatically determined topic is called a **pivot**; see §3.21. Quite a number of languages allow any argument to be a topic, with no grammatical restrictions; they do not have a pivot.

It must be noted that topic is a discourse category, quite different from subject, which is a grammatical category. The subject of a clause very often does function as topic, but it does not have to. Some languages make use of **topicalization**, which may involve moving the topicalized constituent to the beginning of a clause, to highlight it.

There are a number of syntactic derivations which affect valency; see §3.20. **Voice** is the traditional term for derivations which reduce valency—passive and now also antipassive. It is preferable to restrict ‘voice’ to this sense, and not extend it to valency-increasing derivations such as causative and applicative, or to reflexive and reciprocal when they change valency. (Some linguists do accord a wider meaning to ‘voice’ and do so in a variety of different ways, which may lead to confusion.)

The term ‘middle’ was used in grammars of Ancient Greek for a type of voice when ‘the “action” or “state” affects the subject of the verb or [their] interests’ (Lyons 1968: 373). In recent years, the term has been used by a variety of linguists, all in different ways (sometimes including passive, antipassive, reflexive, or just a general intransitivizer). None of the new uses of ‘middle’ seem necessary and all are confusing; it is a term best avoided. (See Dixon and Aikhenvald 2000: 11–12, and further references therein.)

**Alienable** and **inalienable** types of possession, and the varying meanings accorded to these levels, were illustrated in §1.3; see also Chapter 16, especially §16.5. Under (e) in §2.5, there was discussion of how the term **ergative** has been misused by formalists, this being apparently due to a failure to understand the established use of the term; the error became, for a while, institutionalized as dogma.

## 5.7 Markedness

The notion of **markedness** can be useful in linguistic description and explanation, so long as it is defined and applied with care, and not overdone (as, unfortunately, it often is).

In his seminal 1939 study, *Principles of Phonology* (see §2.6), Trubetzkoy provided an exemplary account of types of phonological opposition (briefly mentioned in §2.4). He distinguished first between oppositions which are ‘bilateral’ (involving just two members) and those that are ‘multilateral’ (with

more than two). In English, *d/t* is bilateral since no other phoneme has the features shared by this pair, whereas *d/b* is part of a multilateral opposition—Trubetzkoy considered that everything which is shared by *b* and *d* is also shared by *g*.

The most important classification of oppositions suggested by Trubetzkoy is into the following three types:

- In a **privative** opposition, one feature is characterized by the presence, and the other by the absence, of a certain ‘mark’; for example, voiced/voiceless consonant, nasalized/non-nasalized vowel.
- In a **gradual** opposition, the features show different degrees or gradations of a property; for example, close/mid-close and then mid-close/mid-open tongue height in the vowel oppositions *i/e*, *e/ɛ*.
- In an **equipollent** opposition, the features are logically equivalent; there are neither gradations nor the presence or absence of any property. An example here is labial versus dorsal.

Privative oppositions are bilateral. The **marked** member of the opposition shows a certain mark—for example, glottalization, aspiration, voicing, nasalization, length on vowels—and the **unmarked** member lacks this mark. If the phonological opposition is neutralized, in a certain environment, it will be realized as the unmarked member. For example, in German the voicing distinction for stops is neutralized at the end of a word and here the voiceless realization is used—both *d* and *t* become *t* word-finally, and so on. The unmarked member of an opposition is almost always statistically more frequent, although this is a concomitant property and not a defining one—for example, long vowels are less common than short ones, aspirated stops less common than the unaspirated variety. (As mentioned in §2.4, Trubetzkoy’s longer-lived colleague Roman Jakobson treated all oppositions—in phonology and in grammar—as bilateral or binary, and subject to markedness on a +/– basis. This overuse vitiates the value of ‘markedness’.) There is further discussion of markedness in phonology in §7.2.

The idea of ‘markedness’ is also applied to grammar, although here it has an entirely different nature, so that a different label should ideally be used. But ‘grammatical markedness’ is an established term and can be useful if utilized in a cautious manner.

Markedness in grammar applies most appropriately to closed systems, and there are two distinct varieties of it. These sometimes correlate and sometimes do not. Each variety may be applied to systems of any size (with two members, or with more than two).

(a) **Formal markedness.** If a term in a system has zero realization, then it is formally unmarked. For example, singular within the {singular, plural} number system applying to count nouns in English (§3.13).

(b) **Functional markedness.** This relates to the situation of use—the marked term(s) may be used each in a restricted, specifiable situation, with the unmarked term being employed in all other circumstances. In English, singular number is functionally as well as formally unmarked. ‘Plural’ must refer to a set of two or more referents, whereas ‘singular’ may refer to just one referent, but may also be used in a general sense when no number specification is made; for example *The dog is the most companionable animal I know.*

A classic example of functional markedness concerns the inflectional system on verbs in Warrgamay which was described in §4.8. Quoting allomorphic forms used with intransitive verbs, we find:

- purposive, *-lagu*, used of a definite prediction or consequence
- irrealis, *-ma*, used for possible future, negative future, and apprehension
- perfect, *-gi*, used of events in the past which are irretrievably finished
- unmarked, *-y*

Suffix *-y* must be used when none of the conditions for employment of *-lagu*, *-ma*, or *-gi* are met; and it may also be used when one of them could apply, but the speaker does not wish to make such a specification.

The unmarked term from a system is always the one to be employed in the neutral situation of citation. When speakers of Warrgamay were discussing their language with me, a verb was always cited in unmarked inflection; they’d explain, for example, ‘*gaga-y* is how we say “go”.’

Pronouns in English retain the relic of a case system, as in 1st person singular forms *I* and *me*. *I* is the marked term since it is basically restricted to subject function, whereas *me*—the functionally unmarked term—is used in most other circumstances: as object of a verb, after a preposition, and when making up a sentence of its own (for example: *Who wants an apple? Me.*)

Gender and noun class systems—in most circumstances the two terms are interchangeable—vary between languages in their markedness. In a two-term gender system of {masculine, feminine}, for instance, either of these terms may be the functionally unmarked one.

- In Portuguese, a gender choice is made when the sex of the referent is known and unambiguous. But when it is unknown, or when there is a mix of referents of both sexes, then the functionally unmarked term, masculine, is employed. For example *o pai* (with masculine singular definite article, *o*) is ‘the father’ and *a mãe* (with feminine singular definite article, *a*) is ‘the mother’. But for ‘the parents’ (that is, ‘mother

and father’) one says *os pais* (with masculine plural definite article, *os*). The interrogative *quem* ‘who’ does not in itself indicate gender; it takes masculine agreement. (For further examples illustrating this markedness situation, see Aikhenvald 2000: 53.)

- Jarawara shows the opposite markedness. In this language, gender is shown by agreement within an NP and between clause constituents. Masculine, the functionally marked term, is only used for a masculine singular animate noun or for any masculine inanimate (there is no number system applying for inanimates). Feminine, the functionally unmarked gender, is used for everything else—all feminine nouns (animate or inanimate), all plurals (feminine and masculine), and also all pronouns. In addition, when a speaker uses interrogative *himata* ‘what’ without knowing what the sex of the referent might be, then it takes feminine agreement. (For fuller details, see Dixon 2004a: 80–1, 284–7.)

But markedness cannot be recognized for a gender or noun class system in every language. The system of four noun classes in Dyirbal—which can be roughly labelled ‘masculine’, ‘feminine’, ‘edible plants’, and ‘neuter’—was discussed in §1.9. We saw that the semantic basis of these noun classes is subtle but largely explainable. The ‘neuter’ class, IV, is formally unmarked, shown by zero suffix (as opposed to *-l* or *-yi* for ‘masculine’, *-n* for ‘feminine’, and *-m* for ‘edible plants’). But no term in the system may be recognized as functionally unmarked—neither ‘masculine’, nor ‘feminine’, nor either of the other two terms. During fieldwork—which extended over several decades—I tried to find some ‘neutral circumstances’ in which an unmarked term might appear, but with no success. A group of people of mixed sex, what noun class might be used to refer to them? It depended, people said, on the sex of the most senior member of the group. What about a child in the womb, which noun class would be used to refer to it? (In English we often use *it*, but the neuter noun class marker in Dyirbal cannot be used with human reference.) The answer was that one thinks of a foetus as being either male or female, and refers to it accordingly. There is simply *no functional markedness* in this system, as there is not in a fair number of other noun class and gender systems (see Aikhenvald 2000: 56).

It can be seen that formal and functional markedness do not always correlate. Masculine is the functionally unmarked gender for Portuguese and feminine for Jarawara, but in neither system is any term formally unmarked. Similarly for the verb inflections in Warrgamay; suffix *-y* is functionally unmarked but all four inflections have non-zero form. The opposite applies for Dyirbal noun classes, where there is no functional markedness but ‘neuter’ is formally unmarked.

Quite often, a certain grammatical system does show both functional and formal markedness. In most such cases, the unmarked terms coincide; for example, in the English number system, singular is both formally and functionally unmarked. But this is not always so. Consider the main verbal inflections in a sample northern and a sample southern dialect of Dyirbal:

|                    |                       | -Y conjugation | -L conjugation |
|--------------------|-----------------------|----------------|----------------|
| Mamu               | <b>future–present</b> | <b>-ñ</b>      | <b>-ñ</b>      |
| (typical of        | past                  | -ñu            | -n             |
| northern dialects) | positive imperative   | ∅              | ∅              |
| Jirrbal            | future                | -ñ             | -ñ             |
| (typical of        | <b>past–present</b>   | <b>-ñu</b>     | <b>-n</b>      |
| southern dialects) | positive imperative   | ∅              | ∅              |

It will be seen that future and past are the same for all dialects. But present falls together with future in the north and with past in the south. It is the inflection which covers present that is the functionally unmarked term in the system, shown here in bold. When a speaker of Mamu (or of any other northern dialect) cites a word they will always give it in the future–present form, saying: ‘our word for “eat” is *jaŋgañ*.’ But when a speaker of a southern dialect such as Jirrbal cites a word they will use the past–present form: ‘our word for “eat” is *jaŋgañu*.’ Note that imperative is shown by zero in both conjugations, for every dialect. This is the formally unmarked term, but it is never used in citation. In summary, the functionally unmarked term differs between dialect sets, but it is always different from the term which is formally unmarked.

There are a number of concomitant properties associated with grammatical markedness (these have often been taken, erroneously, to be defining criteria). These include:

- (i) The unmarked term in a system typically has greater textual frequency than any of the marked terms.
- (ii) The realization of the unmarked term is typically a shorter or simpler phonological form than the realization of marked terms. For example, the unmarked verbal inflection in Warrgamay consists just of a syllable-closing consonant, *-y*, whereas the marked terms are one or two syllables in extent: *-ma*, *-gi*, and *-lagu*. But note that in English the functionally unmarked pronoun *me* is phonologically heavier than the marked pronoun, *I*.

Neither of these properties necessarily holds, although they often do. It has also been suggested that—as in phonology—when a certain grammatical contrast is neutralized, it will be the unmarked term which is used. And



markedness may also play a role in dependencies between grammatical systems, discussed in §3.19. When system X depends on system Y, and there are more choices in X if term  $Y_1$  is chosen from system Y (rather than if  $Y_2$  or  $Y_3$  are chosen), then  $Y_1$  is likely to be the functionally unmarked term in its system. For example, there are often more choices in a tense or aspect or person or number system in positive than in negative clauses, and positive does appear to be the functionally unmarked member of the polarity system in every language.

As said above, the notion of markedness is most appropriate when applying to a grammatical system. But it may also be of use in describing construction types. It was mentioned in §3.2 that declarative is often (although by no means always) formally unmarked, and can typically be regarded as the default—that is, functionally unmarked—mood choice. In §3.20, ‘active’ was described as the unmarked construction type, and passive or antipassive as marked. Passive and antipassive will be employed in specific contexts, for particular semantic and/or syntactic purposes (for example, putting an argument into pivot function) with active being used in all other circumstances. In English (as in Latin and many other languages), all verbs may occur in active form, but quite a few are not found in the passive. (For details concerning English, see Dixon 2005a: 360–7.)

It can also be helpful to talk of functional markedness in the lexicon. We saw above that in Portuguese, *mãe* ‘mother’ is the marked and *pai* ‘father’ the unmarked term for referring to parents; when a male and a female parent (‘father and mother’) are referred to as a pair, the plural of *pai* is employed.

A number of adjectives in English constitute antonymic pairs. Consider:

- (1) How heavy is that suitcase?
- (2) How light is that suitcase?

Sentence (1) carries no presupposition as to whether the suitcase is light or heavy; it simply enquires what its weight might be. In contrast, sentence (2) presupposes that it is light and asks just *how* light it is. *Light* is the functionally marked term from this lexical opposition, indicating that something is of little weight. *Heavy* is the unmarked term; it can be used to refer to an object of great weight and also, in neutral circumstances, when nothing about the weight is known.

## Sources and notes

5.1 Swadesh’s 200-word list was consulted in Landar (1966: 186–91); it appears in many other publications. Lexicostatistics was soundly discredited by

Bergslund and Vogt (1962). See also Dixon (1997: 35–7) and further references therein.

5.3 Bauer (1983: 29) states, erroneously: ‘Inflection. Marks agreement’ and ‘Derivation. Does not mark agreement.’ Other discussions of inflection and derivation which are less than satisfactory include Payne (1990: 130, 135).

5.4 Illuminating accounts of double marking—which combine a case and a preposition—are in Benveniste (1971: 113–19) and Jakobson (1990: 332–85). For clitics in English, see Dixon (2007b).

5.5 Horne (1966) surveys early uses of the terms in Set A and Set B. However, he does not mention Duponceau (1819) who appears to have been the first to use the term ‘polysynthetic’.

C. E. Bazell, in his inaugural lecture at the University of London (1958: 17–18), welcomed the fact that the term ‘polysynthetic’ was at that time ‘hardly used any longer’. Sadly, there has been a crescendo of use in the decades since. Bazell commented that the term ‘deserved the self-contradictory definition of the Oxford Dictionary: “characterised by combining several words of a sentence into one word”.’

5.6 The Bardi problem was expounded by Claire Bowerin in a seminar presentation ‘Head and Dependent Marking in Bardi’ at the Research Centre for Linguistic Typology, on 17 December 2001.

The term ‘immediate constituent’ is also in use: X is an immediate constituent of Y if it functions in a structural slot of Y. For example *the very tall man* is an immediate constituent of *The very tall man fell over*, and *very tall* is an immediate constituent of *the very tall man*. But, although *very tall* is a constituent (at one remove, as it were) of *The very tall man fell down*; it is not an immediate constituent of it. As ‘constituent’ is used in this volume it generally has the meaning ‘immediate constituent’.

5.7 Worthwhile discussions of markedness include Matthews (1991: 234–45) and Aikhenvald (2000: 51–6). Also Andrews (1990), whose Chapter 4 is on ‘Myths about markedness’, discussing inter alia statistical frequency and neutralization. Greenberg (1966) and Croft (1990: 64–94) provide a number of helpful insights but are overall of mixed quality. Haspelmath (2006) conducts a wide-ranging survey into what is referred to by ‘markedness’ in the literature. He concludes that, in view of the vastly different ways in which ‘markedness’ is employed, the term is perhaps best avoided altogether.

In fact, there are two verbs in English which are pretty well restricted to occurring in the passive—*rumour* and *repute* (Dixon 2005a: 369).

## Doing Typology

Linguistic typology is ‘the classification of languages according to their general structure rather than according to their historical or geographical relationship’ (Bazell 1958: 3). It aims at putting forward hypotheses concerning universal characteristics of underlying grammatical categories and structures, and interrelations between them. Some of the earliest endeavours were the classifications of languages into isolating, agglutinating, and fusional (or inflectional) types, or along the continuum from analytic to synthetic (to which the label ‘polysynthetic’ has been unnecessarily added), as described in §5.5. (However, the actual term ‘linguistic typology’ did not come into circulation until relatively recently.)

Nowadays, typological classifications are attempted on a wide variety of features; among others, consonant systems, tense systems, genders, accusative and ergative characteristics, and—this being absolutely done to death—constituent order (often being misleadingly named ‘word order’).

A typological study can range over a limited set of languages—all in one geographical area or all in a genetic family—or over all human languages. For it to be significant, there should be a fair number of languages in the population considered (more than just a handful). When a typological classification is designed to cover a large number of languages, there are various potential problems. Around 4,000 languages are currently spoken or were recently spoken across the world. It would be impossible for even a team of linguists to accord close scrutiny to 4,000 grammars. But, in fact, there are good and reliable descriptions available for only a fraction of known languages. This all brings up the question of ‘sampling’, discussed in §6.6.

One can only compare two things which are described in similar terms. One piece of real estate is advertised as: ‘Period house in enviable suburban setting. This nineteenth-century bishop’s residence has been fully refurbished with a patio added; there are polished floors throughout. Offering a stylish and harmonious lifestyle for the discerning buyer.’ Another advertisement states: ‘Three bedrooms, two with built-in wardrobes; two bathrooms; sitting, dining, and kitchen areas with adjoining patio. Convenient to public transport.’

It is not possible to sensibly compare houses on the basis of such disparate descriptions. Similarly for languages.

§6.1 discusses how compatible analyses of languages are needed as the basis for typological comparison. The matter of what can judiciously be taken as the subject matter for typological study is taken up in §6.2. There is then discussion of phonological typology in §6.3 and of grammatical typology in §6.4. A brief note on typological study with respect to the lexicon is in §6.5.

## 6.1 Requirement for consistent analysis

Consider three languages, called 1, 2, and 3, and three sets of words in each for which labels ‘n’, ‘v’, and ‘a’ are used—thus 1n, 1v, 1a; 2n, and so on. The languages vary in the major syntactic properties of their sets of words, as shown in Table 6.1.

Now consider two methods for grammatical description, both operating in terms of the following three criteria:

- I Those words which can be head of an NP comprise the class of nouns.
- II Those words which can be head of a predicate comprise the class of verbs.
- III Those words which can be modifier within an NP comprise the class of adjectives.

The **first method** applies the criteria in sequence. We get

### Language 1

Noun class covers 1n

Verb class covers 1v and 1a

we can recognize 1a as a subclass that may be called ‘state verbs’ (since they refer to states) which can also function as modifier in an NP

TABLE 6.1. Syntactic properties for languages 1, 2, and 3

| CAN BE:           | LANGUAGE 1 |    |    | LANGUAGE 2 |    |    | LANGUAGE 3 |    |    |
|-------------------|------------|----|----|------------|----|----|------------|----|----|
|                   | 1n         | 1v | 1a | 2n         | 2v | 2a | 3n         | 3v | 3a |
| head of NP        | ✓          | –  | –  | ✓          | –  | ✓  | ✓          | –  | –  |
| head of predicate | –          | ✓  | ✓  | –          | ✓  | –  | –          | ✓  | –  |
| modifier in NP    | –          | –  | ✓  | –          | –  | ✓  | –          | –  | ✓  |

**Language 2**

Noun class covers 1n and 1a

we can recognize 1a as a subclass that may be called ‘quality nouns’  
(since they refer to qualities) which can also function as modifier in  
an NP

Verb class covers 1v

**Language 3**

Noun class covers 1n

Verb class covers 1v

Adjective class covers 1a

In the first method, Criterion I applies first. Criterion II applies for those words not assigned to a class by Criterion I. Finally, Criterion III applies for those words not assigned to a class by Criteria I and II. (For languages 1 and 2, there are no words remaining for Criterion III to apply to.)

The **second method** employs the same criteria but has them operating simultaneously rather than sequentially. For languages 1 and 2 a difficulty arises in that word sets 1a and 2a have two functions in Table 6.1. One then has to undertake further study, and decide which function should be considered primary. Suppose that, for each language, functioning as modifier within an NP is the most central (and most frequent) function, while functioning as predicate head (for 1a) or as NP head (for 2a) is of lesser frequency and importance. Under the second method, we thus get the following word classes for languages 1 and 2 (those for language 3 are unchanged):

**Language 1**

Noun class covers 1n

Verb class covers 1v

Adjective class covers 1a

this class may also function as head of a predicate

**Language 2**

Noun class covers 1n

Verb class covers 1v

Adjective class covers 1a

this class may also function as head of an NP

(This discussion is, of course, oversimplified. In deciding on word classes, many morphological and syntactic properties must be considered, for each

language. But to make the pedagogic point, data in Table 6.1 serves perfectly well.)

Now consider how these two methods of analysis—identical save that the criteria are applied in sequence in the first and simultaneously in the second method—feed in to a typological study concerning adjective classes. Under the second method, all three languages have an adjective class with the basic syntactic property of functioning as modifier within an NP. In languages 1 and 2 (but not in language 3) the class has a further function as head of a predicate or NP. In contrast, the first method of analysis would recognize an adjective class just for language 3.

It is likely that the subclass of state verbs in language 1, the subclass of quality nouns in 2, and the adjective class in 3 will have similar semantic content, including items from the Dimension, Age, Colour, and Value semantic types, and probably from Physical Property (see §1.11 and §3.6). This is a further reason for calling each of 1a and 2a an adjective class instead of a subclass of verbs and of nouns, respectively. But some linguists do prefer the first method of analysis.

A typologist examining these languages should not just reproduce the descriptions provided. It may be that the grammar provided for language 1 uses the first method and that for language 2 the second one; there would then be an adjective class for languages 2 and 3 but not for language 1. Or vice versa, when there would be an adjective class for 1 and 3 but not for language 2. While I was working on a typology of adjective classes (published as Dixon 2004b, substantially repeated in Chapter 12 of Volume 2) it was necessary to carefully study each grammar. In many instances, where a linguist spoke just of ‘state verbs’ or of ‘quality nouns’ it was possible to perceive criteria for recognizing these as a distinct adjective class rather than as a subclass of verbs or of nouns. (This applied for high-quality grammars, with detailed syntactic and morphological information provided. Such a reanalysis was not possible in the case of a grammar of lesser quality; but of course such grammars do not in any case constitute adequate base material for typological study.)

A number of similar situations were presented in previous chapters. For instance, the treatment of ‘adjectives’ in Warekena and Tariana (§4.5) and the question of whether Fijian should be recognized as having a distinct number class rather than numbers being regarded as a subclass of verbs (§4.4).

A survey of ‘passives’ states: ‘The analysis of the various constructions referred to in the literature as “passive” leads to the conclusion that there is not even one single property which all these constructions have in common’ (Siewierska 1984: 1) This shows that a typologist must take great care in

assessing what is described as a passive construction in each grammar under consideration. (See §3.20.)

When embarking on a typological investigation of category X, one must first of all adopt a working definition of ‘what X is’ and the types of criteria for recognizing it in a given language. Then, for each language which is studied, the available grammar(s) must be carefully assessed to see if what is there called ‘X’ accords to the definition and criteria (or, perhaps, whether something which is given a different label in the grammar does so). In some languages, what are called ‘passive’ will not satisfy the criteria and so should be excluded from the typological study.

It would be a mistake to look up the index at the end of a grammar, see that, say, ‘passives’ are discussed on pages 332–5, and then just read these four pages. But a part only exists with respect to the whole. Analytic decisions about passives are likely to depend on or interrelate with other kinds of analytic decisions, in various parts of the grammar (for example, concerning number and types of arguments). Proper procedure is to study the whole grammar in outline, to understand the context for this short discussion of passives, and then see how this links up with treatment of transitivity, other syntactic derivations which affect valency, discourse structure, and the like.

Sentence (1) of §5.6 illustrates the ergative marker *-nim* in the Australian language Bardi. One linguist analyses it as a case suffix and another as a clitic, which may be called an adposition. Which analysis is preferred makes a huge difference for a linguist working on the typology of case systems—is Bardi to be included or not?

The fundamental error here lies in embarking on a study of just one of the alternative means for the marking of grammatical relations within a clause. All languages need some way of communicating which argument is in A or O or S function, or is instrument or beneficiary or recipient, etc. There are various mechanisms for doing this:

- (a) By case-type markers, associated with NPs. These can be affixes, clitics, or words (the latter two sometimes called adpositions).
- (b) By bound pronouns, typically included within the predicate.
- (c) By constituent order.
- (d) By the pragmatics of the situation of utterance.

Or, of course, several of these mechanisms may be combined. If a typologist embarks on such a study—of how grammatical relations are marked, in general terms, not just of *one* mechanism for marking—then whether *-nim* in Bardi is regarded as an affix or a clitic matters little.

In summary, one should undertake typological study of some aspect of the underlying organization of grammar, rather than of its surface realization. And one should carefully consider the analysis of this feature provided by each grammar that is included in the survey. In an ideal world, writers of grammars would all employ similar criteria and make analytic decisions on a similar basis. I hope that the present volume on the principles of basic linguistic theory may engender a step towards this situation.

## 6.2 What should be compared

There are basically two modes of linguistic typology. The first is **intra-language typology**, comparing a feature in one language with the same or very similar features in other languages, in terms of an agreed set of theoretical parameters. The following kinds of item are available for typological study:

- **structures**—for example, the underlying make-up of a clause or phrase in grammar, or of a syllable (or other unit) in phonology;
- **systems**—for example, noun classes or tense-aspect choices in grammar; vowel systems in phonology;
- **construction types** in grammar—for example, relative clause or complement clause constructions.

To these can be added associated mechanisms, such as marking of syntactic relations within a clause, and of a possessive modifier within an NP.

What should *not* be compared are isolated entities, such as the occurrence of an unrounded central close vowel phoneme /i/ ('barred i') across a group of languages. The proper method is to compare complete vowel systems, not one isolated vowel. Its role within the language will vary with the size of the vowel system. For instance /i/ in a two-vowel system—consisting, say, of /i/ and /ə/—will be quite different in terms of phonological function and range of phonetic realization from /i/ in a fifteen-vowel system.

An example of inadequate typology in the field of grammar would be to study which languages within a given population have a plural marker. As demonstrated in Table 1.1 of §1.4, plural may be used for 'more than one', 'more than two', 'more than three', or 'more than a few'. What should be compared are complete number systems, and their function within each grammar. (For example, whether the number system is associated with nouns, adjectives, demonstratives, verbs, and so on.)

The other mode can be called **extra-language typology**. This involves considering something in the real world, outside of the abstract systems of phonology and grammar (which are the result of a linguist's analysis), and seeing how it may be coded within a grammar.



For example, how is reference to past time coded across a number of languages (or across all languages)? The possibilities include:

- (a) By a 'past' term in a tense system; for example, in a two-term system, {past, non-past}, or a three-term system, {past, present, future}. Some languages have a two-term system {future, non-future} with the 'non-future' term referring to both present and past. Which time frame is intended may be clear from the context of utterance, or by use of a lexical time word—see (c)—or by some other means. (As mentioned in §3.15, tense is generally marked within the predicate but in some languages may also attach to an NP; for example '[my husband]-PAST' would be 'my ex-husband'.)
- (b) There may be no tense system per se but some other indication of non-spatial setting can provide limited information about time reference. For example, in the inflectional system on verbs in Warrgamay, described in §4.8, 'perfect' refers to an event from the past that is irretrievably finished. A clause marked as perfect refers to past time, but not every clause referring to past time is marked as perfect.
- (c) All languages have time lexemes, typically including 'recently', 'yesterday', 'now', 'tomorrow', and 'soon'. If there is no tense system in the grammar, these may carry the major burden of time reference.

When an extra-language typology describes how something in the real world—time, or sex/gender, or definiteness, or whatever—can be shown through a certain type of grammatical system (as with a tense system for time reference), the linguist *should then* investigate the intra-language typology of the system. That is, an extra-language typology must not be regarded as an end in itself, but—if it is to have predictive power—it must be seen as a conduit to one or more intra-language studies.

At the other end of the spectrum, one can consider a certain vocal gesture in the mouth, and how this relates to phonological systems of languages. Consider an apico-alveolar tap sound, [ɾ], made by instantaneous closure between tongue tip and gum ridge. In one language [ɾ] may be the sole realization of a phoneme /ɾ/. In another it may be one allophone of a phoneme more typically realized as a trill /r/. In some languages, [ɾ] is one allophone of a phoneme also realized as a voiced stop, /d/, or a voiceless stop, /t/. In each instance, of course, the type of extra-language connection must be backed up by intra-language consideration of the types of consonant system involved.

In summary, extra-linguistic typology examines how something outside a language system is coded within the system. To have any scientific validity, it must be augmented by intra-language typological study of the coding mechanism(s).

### 6.3 Phonological typology

One of the commonest topics for typological study is **syllable structure** (see §7.4). Each syllable has a nucleus (V) which is generally a vowel but can be a syllabic nasal or liquid (as in English *bottle*, /bɒtl/, where the final syllable has /l/ as nucleus). Most typically, a syllable has a consonant (C) as initial onset; many syllables have a further consonant as final coda.

A simple and quite common syllable structure is CV. Some languages allow CV and V, abbreviated as (C)V. Other straightforward possibilities include CV(C) and (C)V(C). There are many kinds of complex syllable structures, with more than one consonant in onset and/or in coda slot. As mentioned in §1.3, English allows CCCVCCC syllables, exemplified by *strengths* /streŋθs/. One can also undertake typological study of the structure of foot, of phonological word, and of intonation group. (See Chapter 10 for discussion of phonological word.)

Every description of a language should specify its set of phonemes, divided into a vowel system and a consonant system. The **vowel system** corresponds to slot V in syllable structure. A small vowel system is likely to apply at every V slot in a word. However, a large system may only be operative in, say, a stressed syllable, with a smaller subsystem being used elsewhere (as happens in English).

The commonest vowel system is that enshrined in the Roman alphabet, {/i/, /e/, /a/, /o/, /u/}. If there are only four vowels, then a system {/i/, /e/, /a/, /u/}—here /u/ generally covers [u] and [o]—is commoner than {/i/, /a/, /o/, /u/}—where /i/ may cover [i] and [e]. A three-vowel system is almost always {/i/, /a/, /u/}. There are many kinds of variant, often including the unrounded central close vowel /i/; for example {/i/, /e/, /a/, /u/, /i/}. There are some languages analysed as having just two vowels—generally with a height distinction, either {/i/, /ə/} or {/ə/, /a/}—and some systems with more than a dozen members.

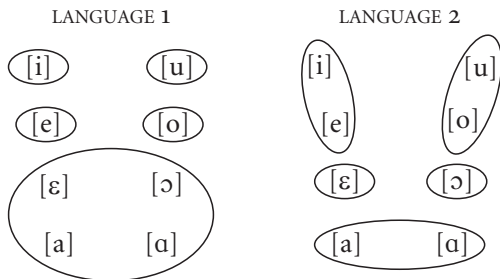
A typological survey of vowel systems across the Australian linguistic area found the following (full details are in Dixon 2002: 628–32):

- two vowels in just two or three languages (analyses vary somewhat)
- three vowels in about 150 languages
- four vowels in 20 languages, 12 of which have {/i/, /e/, /a/, /u/}
- five vowels in 47 languages, 40 of them having the standard system {/i/, /e/, /a/, /o/, /u/}
- more than five vowels in 15 languages

The Australian area is thus unusual in having such a large number of languages with a three-vowel system, whereas in most other areas a five-vowel system predominates.

There may also be a contrast between short and long vowels; often, only some of the short vowels have a long congener. And between plain and nasalized vowels; the nasalization prosody may only apply for some vowels.

A typology of vowel systems should pay attention to the phonetic values of the members of each system. Consider prototypical systems of five vowels. The mid vowels may be [e] and [o] in one language, but [ɛ] and [ɔ] in another. That is, in terms of cardinal vowel values, with circles and ovals indicating phonemes:



A comprehensive typology of vowel systems should take care to recognize these (and other) subtleties of five-vowel systems. They carry quite different potential for future change in the language.

A **consonant system** is typically about three or four times the size of the vowel system in that language (although there is great variation). In a language with syllable structure CV, the full consonant system is likely to apply at every C slot in a word. But with more complex structures, there are generally different subsystems of consonants occurring at different C position in the word. There are always greater possibilities for the onset slot, C<sub>1</sub>, than for the coda slot, C<sub>2</sub>, in a C<sub>1</sub>VC<sub>2</sub> syllable. In some Australian languages—for example Martuthunira (Dench 1995: 30–4; see also Dixon 2002: 553–7)—the full consonant system only applies at a medial position between vowels, C<sub>2</sub> in C<sub>1</sub>VC<sub>2</sub>VC<sub>3</sub>, with restricted systems at C<sub>1</sub> and C<sub>3</sub> (and at C<sub>4</sub> and C<sub>5</sub> in C<sub>1</sub>VC<sub>4</sub>C<sub>5</sub>VC<sub>3</sub>); see §7.4. Consonant clusters at the beginning or end of a syllable generally show severe limitations on what may appear at each position within the cluster. Saying that a certain set of languages ‘shows initial clusters CC-’, for example, is of little value. The types of consonant involved should be specified—for example stop plus liquid, or nasal plus stop, or whatever.

Consonants are classified by place and manner, with secondary specification for things like voicing, aspiration, glottalization, palatalization, labialization, etc. A typological study of one feature should always take account of the whole consonant system to which it belongs. Whether there is a contrast of aspiration must be related to the places and manners to which aspiration applies and how it ties in with the voicing distinction (if there is one); for

example, whether there are aspirated and non-aspirated versions of just voiced stops, or just voiceless stops, or both.

In some languages stress is contrastive—moving the position of stress may create a new word. In others it is predictable, and may then go on the first syllable of a word, or the last, or the second to last, or there may be some more complex rule for stress assignment. We saw in §4.6 that in Fijian stress goes on that syllable which includes the penultimate mora of the word. A typology of stress placement must pay attention to every factor which may be relevant. For example, whether syllables or moras are involved and, if moras, how they are defined. In some languages a syllable of form CVV or CVC is regarded as having two moras, in other languages only a syllable of the structure CVV has two moras; see §7.6.

Other topics for phonological typology include whether there is a system of tone contrasts; since the nature of tone systems varies widely, such a study will be of limited use unless it is specified whether the tones are realized in terms of pitch height or pitch contour (or a combination) and their segmental scope.

There are many other topics within phonology ripe for typological study. For instance, the occurrence of a prosody of vowel harmony. It is important to specify the scope of each prosody—whether it applies over a syllable or just a part of a syllable, or over all or part of a phonological word, or over a phrase. And also the phonetic parameter involved in the harmony—whether front/back, or high/low, or advanced/retracted tongue root, or whatever.

§6.1 emphasized how typological study of a group of languages is dependent on the ways in which they have been analysed. As a further example, a phonetic sequence [ai] may be interpreted phonologically as a VC sequence, /ay/, or as a VV sequence, /ai/. Which analysis is selected could affect the phonotactic profile provided for the language. For instance, Fijian is restricted to CV syllables, where V can be a short vowel, a long vowel, or a vowel sequence. If [ai] were interpreted as /ay/ then the language would have to be accorded a second syllable type,  $C_1VC_2$  (with only  $y$  being possible in slot  $C_2$ ).

However, as pointed out in §4.6, there is generally some clear line of argumentation within the phonology for preferring one type of interpretation over the other. If a phonological analysis is sound, then all will be well with typological study based on it. (This does of course carry a warning—should any be needed—against using, in any typological work, half-baked descriptions of languages which have not been properly thought through and/or are based on inadequate field investigation.)

## 6.4 Grammatical typology

Although **word classes** have lexical content—in contrast to closed systems which are entirely grammatical—the establishment of word classes is a

grammatical matter. And they constitute a fine field for typological study. Do all languages have noun and verb classes? This important question is discussed in Chapter 11, and answered in the affirmative. Chapter 12 concludes that an adjective class can be recognized for every language. But, as mentioned before (in §3.6), an adjective class can be small and closed, or large and open; it may have grammatical properties similar to those of verbs, or similar to those of nouns, or to both, or to neither. These are prime topics for comparative enquiry.

Underlying **grammatical systems** are also suitable fodder for careful typological scrutiny—tense and/or aspect, evidentiality, gender/noun classes, demonstratives, pronouns, and the like. For each category, in every language, care must be taken to note its interactions with other systems. For example, an evidentiality system always applies in past tense and declarative mood. Does it also apply (in full or abbreviated form) in other tenses (see §3.19), and in other moods?

As stated before, a single term in a system—such as dative case—is not a suitable topic for typological examination. What is called ‘dative case’ should have a common core of meaning and function, but its range will vary in different languages. However, it is always one item within a case system, and has meaning and function with respect to other terms in the system. Any comparative study should, at the least, compare complete case systems. And, as noted in §6.1, a case system is just one of several possible ways of marking grammatical relations—a proper typology should focus on the latter as a holistic enterprise.

As part of study of a grammatical category, the typologist will include information on the realization of the category. Take gender. First of all, how many terms are there in a gender system and what is the semantic content of each? Then, on what elements in an NP and in a clause is gender marked? How does the gender system fit into the overall grammatical scheme of the language? Finally, how is it shown—by prefix or suffix or in some other way (and is it fused with the realization of some other category)?

What is less useful is to begin with realization. For example, asking which languages have genders shown by suffixes and which by prefixes. Or, a neophyte linguist may notice that feminine is shown by an affix including a dental or alveolar nasal, *n*, in a handful of languages, and embark on a study to see in how many languages the realization of feminine gender includes *n*. Unless there were some sound-symbolic basis to genders—which, as far as I know, there isn’t—this would not be a useful task to undertake.

Grammatical categories are typically realized through morphological processes applying to a root or stem. Following Sapir, the processes were listed in §3.13 as (leaving aside compounding): reduplication, shift of stress or change of tone, internal change, subtraction, and affixation. Within the

typological examination of a category, its types of realization must be considered. What is not useful is to start at the other end—that is, to look at a type of morphological process, and ask what it may realize. The most common process is affixation. Surely affixation can realize just about any underlying category—not much to study there. What about the less common processes? Shift of stress or change of tone, for example? Well, these mark derivation between word classes in some languages, tense or aspect in others, syntactic function in others. There is no limit on the grammatical categories which can be realized by a particular kind of process, and relatively little insight into the basic functioning of language can be gained by considering the matter.

The process of reduplication holds a particular fascination for linguists. Partly because it is so obvious and useful a process but is largely lacking from the familiar languages of Europe. And there is a semantic reason. The meaning attached to reduplication is quite often—but by no means always—iconic with morphological repetition. Reduplication of a verb can indicate an action repeated several times, or done in several places, or performed continuously, or done with vigour. With a noun, reduplication may indicate plural or distributive (scattered over a region). But in addition to these sorts of meanings, reduplication may mark derivations (from a verbal root to a nominal stem, or from a noun root to an adjectival stem, and so on), or tense, or possession (see §3.13), or syntactic function. For example, in Alutor—spoken on the Kamchatkan peninsula—some nouns form nominative singular by final reduplication of the initial syllable; for example *kalti* ‘beetle’, nominative singular form *kalti-kal*. It appears that, like other grammatical processes, reduplication has the potential to realize any underlying grammatical category.

There are different forms of reduplication, which must be taken into consideration—repetition of all or part of a stem, either before or after (or, occasionally, in the middle of) the stem. As described in §3.13, in Jarawara initial reduplication of the first syllable of a stem means ‘do a bit’, initial reduplication of the first two syllables is ‘do with force’, and final reduplication of the last syllable indicates ‘many participants’. It is doubtful if these links between phonological form and the meaning of reduplication would be replicated elsewhere.

An affix of a certain form may apply to nouns, and to verbs, in a given language, carrying a different meaning with each word class; we have, in effect, homonymous affixes. (For example suffix *-s* in English, which can indicate the possessive relation or plural number with nouns, and 3rd person singular present with verbs.) In similar fashion, reduplication may apply to two or more word classes, showing a different form and meaning with each. Indeed, this may constitute a criterion for distinguishing word classes; it is discussed further in §§12.5.1–2, in connection with the recognition of adjective classes.

To summarize, although the meaning carried by reduplication sometimes refers to repetition or some kind of plurality, it often does not. It may mark augmentative in one language and diminutive in another (see §3.13). Typological study of types of reduplication is likely to shed only a little light on the underlying structure of human language, and will provide minimal input to the development of basic linguistic theory.

**Underlying structures of clauses and of phrases** are a fit topic for typological study. For example, if there is a copula construction, it always allows two arguments, copula subject and copula complement, as described and exemplified in §3.2 (and Chapter 14 in Volume 2). Some languages have a further copula construction with only one argument, the copula subject (as in Latin *deus est* ‘god is’, meaning ‘there is a god’). Which languages have a single-argument copula construction, and how it interrelates with other aspects of its grammar, would be a fruitful question to pursue.

The structure of a verbal clause is, fundamentally, an investigation of transitivity. All languages have intransitive and transitive clause types, as described in §3.2. Many also have extended transitive (or ditransitive) and some also have extended intransitive clauses. In a number of languages there are clauses which appear to have two objects. The number and types of peripheral argument allowed in a clause vary from language to language. All these are fine topics for typological scrutiny.

Study of underlying clause structure will pay attention to how one argument is distinguished from another in surface structure. As mentioned several times (including in §6.1) there are various possible mechanisms here, including case or similar markers on an NP, bound pronouns, and contrastive constituent order.

That is, constituent order can be—in some languages—one way of showing which argument is in A and which in O function in a transitive clause, and so on. Following the work of Greenberg (1963), the linguistic world has during recent times been besmirched by an overpowering addiction to study of constituent order (often mislabelled ‘word order’). Anyone working on this topic must first of all establish the role of constituent order in each language under consideration. The two major possibilities are:

- fixed—or almost fixed—order, indicating the syntactic function of NPs;
- fairly free order, which can be used to pick out that argument which is in topic/pivot or focus function.

It is not sensible to ignore the distinction; yet this is often done. Some languages are like English in showing a rather fixed constituent order. Others allow considerable fluidity. Yet, for each language of the second type, some practitioners feel the need to establish some ‘basic order’, as if it were a

language of the first type. Enough was said about this in §2.4 (which could well be reread at this juncture).

Phrase structure is also a topic for comparative study. What can be the head of an NP? What modifiers are possible with various kinds of heads? Can there only be one adjective modifier or more than one? And allied to all this is the question of the order in which head and the various modifiers must be, or can be, or are likely to be.

Construction types are a most important topic for comparative study—imperative and interrogative clauses, and multi-clausal constructions involving relative clauses, complement clauses, and the like. Reciprocal and reflexive underlying structures can be shown in a number of different ways; before studying these topics one must first investigate the transitivity system of a language, and its pronominal paradigm. Discussion of these and similar topics is planned for inclusion in Volume 3.

A further matter which is ripe for typological study involves dependencies between grammatical categories. For example, there may be fewer person distinctions under negation, or fewer in past tenses than in non-past tenses. This kind of typology was broached in Aikhenvald and Dixon (1998), the major points from which were summarized in §3.19. It will provide a fertile field for future research.

## 6.5 Lexical typology

One could compare the range of meaning of, say, ‘head’—that lexeme whose central reference is to the round upper part of the human body (which includes ears, eyes, nose, and mouth)—across a number of languages. For example, in four languages that I know well:

- **Yidiñ**—*dungu* is head of person, animal, or grub, top of mountain, top of boil, the position of the sun at midday, etc.
- **Fijian**—*ulu-* is head of animate being, which is considered to be the most important part of the body (if a commoner should touch the head of a chief, even accidentally, it is considered a heinous offence); the most important part of anything; a mountain is ‘head of the land’.
- **Jarawara**—*tati* is head of person or animal; upper part or top of an object, such as pineapple plant or tree; roof of house; prow of canoe; either end of an aeroplane runway; etc.
- **English**—*head* is upper portion of body of person or animal; the mind; person in charge (of company or school or department); front of a queue; top part of an object (for example, stairs, screw, river); froth on top of a glass of beer; etc.



Comparison of the full range of meaning of ‘head’ across a variety of languages might be interesting, but anecdotal. For a study to qualify as lexical typology it should involve comparison of a tightly knit set of terms, the meaning of each being with respect to the meanings of the other terms in the set (just as in a grammatical system). In §1.7, the six terms in English to describe the body from shoulder to finger (*shoulder, arm, elbow, wrist, hand, fingers*) were ranged against the seven terms in Dyirbal for this part of human anatomy. A comparison of this set of terms across languages would constitute a kind of typological study. It should be tied in with consideration of the corresponding lower part of the body, from thigh to toes. One could then consider questions of the following sort: If there is a separate term for ‘fingers’ is there also one for ‘toes’? Are ‘elbow’ and ‘knee’ dealt with in similar fashion? And so on.

Certain areas of the lexicon do show systematic organization which makes them suitable for lexical typology. Numbers are one field, and kin terms another. One must first work out the nature of the basic kinship system (on which there is substantial anthropological literature), and then look at terms used to code underlying kin categories. The set of colour terms is also a fine field for typological study, building on the theoretical framework introduced by Berlin and Kay (1969).

Each language will have a number of well-integrated sets of terms for fields of cultural specification. For instance, the main verbs to describe ‘piercing’ or ‘spearing’ in Dyirbal are:

- *bagan*, the most general term for piercing something with a sharp-pointed implement; including: dig with a yamstick; sew clothes; row boat (with oar piercing water); squeeze boil; bird makes nest. Used to describe throwing a spear with the aid of a woomera (spear-thrower).
- *jinban*, throw a spear from the hand, not using a woomera
- *jurrgañu*, spear something that can be seen (typically, fish), holding on to the spear; prickle or splinter enters skin
- *wagañu*, spear something that cannot be seen, holding onto the spear; for example, jab a spear into long grass in which there has been a movement, indicating the presence of a small animal; jab a spear between submerged tree roots to try to impale an eel (since eels tend to hide there)
- *ñirran*, poke something sharp into (meat, ground, log,) to see whether it is cooked or soft, etc.
- *ñuban*, poke a stick into the ground to test for the presence of something (for instance, wild yams or snails)

It would be fascinating to compare this with similar sets of verbs in other languages. Unfortunately, other languages—even neighbouring tongues

in Australia—do not have anything comparable. The possibilities for lexical typology are rather limited simply because each community tends to structure its lexicon in a way that suits its cultural needs, and these differ immensely.

One suitable topic for lexical typology is the set of verbs referring to posture, with key members 'sit', 'stand', and 'lie' (see §8.3.3). There are various kinds of additions to and variations on the basic schema. One language may include a verb 'hang' in the set; another may have two verbs for different varieties of 'lie'. But all languages do have a set of posture verbs which may be compared—in terms of their range of referential meaning and their other functions in the language. For example, some posture verbs may also function as auxiliaries, or as verbs of existence. In Enga (Papuan area) different posture verbs are used to describe the existence of various classes of beings: women sit, men stand, wasps hang, and eels lie.

Lexical typology is particularly fruitful when it gazes towards grammar. It was mentioned in §1.11 and §3.6 that even a small adjective class is likely to include lexemes from the Dimension, Age, Colour, and Values semantic types. It would be a valuable exercise to consider a semantic type—say Age, with lexemes such as 'old', 'senior', 'young', 'new'—and see which word class or classes it relates to, and concomitant grammatical properties, on a cross-linguistic basis.

## 6.6 The question of sampling

Someone may wish to attempt some generalization about a linguistic feature across all the languages in a given region or family. They find it hard to examine every single language, so take a sample of them. However, there are among linguists today many mistaken ideas about sampling.

'Contrary to popular belief, the size of a sample does not depend primarily upon the size of the population from which the sample is taken. Instead, the sample size depends upon the degree of accuracy demanded in representativeness of the sample statistics as an estimate of the population parameter' (Reichmann 1964: 231).

Suppose that an investigator wants to discover what percentage of the population in various urban centres owns a dog, and looks at both a city with three million inhabitants, and a town of 30,000 people. The *same size* sample will be appropriate for each location, to achieve a certain level of accuracy. If it turns out that about one-quarter (25 per cent) of people has a dog, then a sample size of 1,875 will produce a standard error of 1 per cent. That is, there is a 95 per cent confidence level that the value over the whole population will be within twice the standard error on either side of the sample score; there is a 95

per cent chance that over the whole population the proportion of dog-owners is between 23 per cent and 27 per cent. With a smaller sample size of 500, the standard error is 1.9 per cent, so there is a 95 per cent chance that between 21.2 per cent and 28.8 per cent of the whole population has a dog. Employing a much smaller sample, of 100, the standard error is 4.3 per cent, so that there is a 95 per cent chance that the number of dog-owners in the population as a whole is between 16.4 per cent and 33.6 per cent. As can be seen, such a small sample produces results that are really too vague to be of much use, making the enterprise scarcely worthwhile.

Another important prerequisite is that correct information should be provided by all respondents in the sample. If some of them do not properly understand the question, the information they give may be inaccurate. For example, they may have thought they were being asked whether they liked dogs, or if they had ever owned a dog in the past. Care must be taken to weed out such responses.

Most linguistic work which involves sampling deals with matters of surface structure, which in the earlier part of this chapter were not recommended as likely to shed light on the basic nature of human language. For instance, whether or not an NP in A function must or is likely to precede one in O function, within a clause. Nevertheless, we should enquire whether the results they provide have any validity. A typologist may consider the 900 or so languages of sub-Saharan Africa, or the 880 languages of New Guinea, or the 250 languages originally spoken across Australia, or the approximately 400 languages currently spoken in South America. They often adopt a different-sized sample for each area, proportional to the total number of languages there. As stated above, this is bad practice. Basically, the same size sample is needed for each area, irrespective of the number of languages spoken there, to achieve a required degree of accuracy.

Suppose that a 'yes' or 'no' question is asked (such as 'does an A NP precede an O NP?'). Let  $p$  be the chance of answer 'yes' and  $q$  (which is  $1-p$ ) the chance of 'no' in the sample; 'n' is the size of the sample. Then we have:

$$\text{standard error} = \sqrt{\frac{p \times q}{n}}$$

There is a 95 per cent chance of the figure for the whole population being within two standard errors of the sample value. Table 6.2 shows these standard errors for 50/50 ( $p = q = 1/2$ ) and 25/75 ( $p = 1/4, q = 3/4$ ) situations.

That is, if, in a sample of 100 languages, A precedes O in about half of them, there is a confidence level of 95 per cent that over the whole population A will precede O in between 40 per cent and 60 per cent of languages. With a sample of 25, the 95 per cent confidence level applies for a

TABLE 6.2. Standard errors for different population sizes

| SAMPLE SIZE | 50/50 distribution in sample<br>( $p = 50\%$ ) |                                                                          |  | 25/75 distribution in sample<br>( $p = 25\%$ ) |                                                                          |  |
|-------------|------------------------------------------------|--------------------------------------------------------------------------|--|------------------------------------------------|--------------------------------------------------------------------------|--|
|             | STANDARD ERROR                                 | 95% CONFIDENCE LEVEL THAT FIGURE FOR $p$ OVER POPULATION WILL BE BETWEEN |  | STANDARD ERROR                                 | 95% CONFIDENCE LEVEL THAT FIGURE FOR $p$ OVER POPULATION WILL BE BETWEEN |  |
| 200         | 3.5%                                           | 43% and 57%                                                              |  | 3.1%                                           | 18.8% and 31.2%                                                          |  |
| 100         | 5%                                             | 40% and 60%                                                              |  | 4.3%                                           | 16.4% and 33.6%                                                          |  |
| 50          | 7.1%                                           | 35.8% and 64.2%                                                          |  | 6.1%                                           | 12.8% and 37.2%                                                          |  |
| 25          | 10%                                            | 30% and 70%                                                              |  | 8.7%                                           | 7.6% and 42.4%                                                           |  |
| 10          | 15.8%                                          | 18.4% and 81.6%                                                          |  | 13.7%                                          | zero and 52.4%                                                           |  |

figure over the entire population that is somewhere between 30 per cent and 70 per cent.

It will be seen that the smallest sample which would provide useful results for a simple 'yes/no' question is about 200 (and, if possible, a rather larger sample than this is to be preferred). Of course, for more complex questions—such as those indicated as the proper subject matter for typology throughout this chapter, involving interrelations between categories—considerably larger samples would be required. In essence, the soundest approach is to examine all appropriate languages. This would be a huge task. But proper scientific method does not allow for short cuts.

However, as pointed out before, only a small portion of the 4,000 or so languages currently or recently spoken across the world have been provided with thorough and reliable descriptions. These are available as a reliable basis for typological study. The many partial and defective grammars should be avoided. They are like the respondents in the dog-owner questionnaire who did not correctly understand the question. Unreliable data must be excluded from any set of materials for typological study.

We can focus on one unfortunate attempt at sampling (note that there are many others of similar quality). Bybee, Perkins, and Pagliuca (1994) survey 'tense, aspect and modality in the languages of the world' on the basis of a sample of 94 languages. These were not chosen at random, but (sensibly) in terms of picking a number of languages from each of a set of purported genetic groupings. Sadly, the genetic groupings adopted were deeply flawed. They included 'Andean-Equatorial', which involves the Arawak and Tupí families, which can *not* be linked genetically. Similarly for 'Ge-Pano-Carib'. Worst of

all is 'Indo-Pacific' which links the sixty-odd language families, plus many isolates, of New Guinea, the two language families of the Andaman Islands (for none of which is a sound grammar available), and the languages of Tasmania (about which virtually nothing is known for certain).

Within each group, one or more languages were chosen by a random method (such as looking at every fortieth language from a list). If there was no material available for a selected language, the next on the list was chosen. If there was *some* grammatical information available for this language, it was included in the survey, even though the data available might be scanty and unreliable, and even though there was a well-described language next to it (which was not included in the survey). In Australia, for example, Worora was included mainly on the basis of materials by a missionary who knew Latin but not linguistics. However, for the neighbouring language, Ungarinjin, there is available a fine grammar. Another unfortunate choice was Gugada, a dialect of the Western Desert language for which an extremely poor grammar is available. (It appears that 'language' and 'dialect' were not properly distinguished.) There is one grammar of the highest quality available for another dialect of this language, and good grammars for five other dialects.

There are so few really good grammars available that any typological survey which fails to take account of the best exemplars is, in effect, shooting itself in the foot. A recent volume entitled *The world atlas of language structures* (Haspelmath et al. 2005) is open to virtually every criticism there is, one being that its database includes languages for which the materials available are poor and unreliable, but omits some excellent sources.

The idea has spread that using sampling techniques, and quantified correlations, and chi-square tests and the like, makes linguistics 'scientific'. However, what is needed, at the present time, is not quantitative comparison of superficial bits of surface structure, but rather qualitative study, providing careful and fine-grained analysis of the underlying structures of some of the more than 3,000 languages which are still spoken and are in need of documentation.

What then is the optimal procedure in typological enquiry? Is any type of sampling appropriate? The answer is: 'not really'. Instead, one should carry out an extensive survey, and then home in on intensive studies. For instance, when Alexandra Aikhenvald began working on a cross-linguistic study of evidentiality, she first surveyed languages across every linguistic area and genetic family. The category is missing from some regions, spottily represented in some, rampant in others. Having identified critical areas, she then embarked on a detailed examination of all the well-described languages there. She also looked at poorer grammars of some of the languages which might be expected to include an evidentiality system, seeing if she could spy

any symptomatic features, and, where possible, corresponded with authors of the grammars. Finally, as the result of six or seven years of fairly intensive study, in 2004 she published a 479-page monograph, called *Evidentiality*. I followed a similar procedure in working on ergativity (Dixon 1994), on adjective classes (Chapter 12 of Volume 2 of this work), and on other topics.

One vitally important point which has been mentioned before but deserves emphasis is that not everything which has appeared in print—been written down—is equally good. When embarking on a comparative study of the indigenous languages of the Americas, Peter S. Duponceau—in 1819—set out his method of working:

I left no book or manuscript unconsulted that came within my reach; but I examined the assertions of each writer with a critical eye, fully determined in no case to swear on the word of a master. I tried to discover the sources from which my authors had derived their knowledge; the opportunities which they had of acquiring it; the time which they had spent among the Indians, or in the study of their languages; the degree of attention which they had bestowed upon it, and the powers of mind by which they had been enabled to take a just and an accurate view of their subject. Finally, I rejected every thing that came in the shape of mere assertion, and paid attention only to those specimens of the different idioms in which their grammatical structure was sufficiently exhibited.

A typologist requires the ability to discern whether a grammar is or is not reliable. People who have themselves undertaken extensive immersion fieldwork, and written good grammars, naturally acquire this ability. Those who have chosen a different lifestyle are less likely to develop the necessary discrimination. And one has to be prepared—like Duponceau—to be honest: to say that such-and-such is lacking in quality, even if it risks offending a colleague (or perhaps someone who exerts power within the establishment).

The chapters in Volumes 2 and 3 of the current work are based on typological study of the kind described in this chapter. I hope that they stand as testimony to the method.

## Sources and notes

Bazell's 1958 essay is highly recommended as an introduction to linguistic typology.

6.2. Most languages have a three-term person system. Some also have what has been called a 'fourth person pronoun'. How interesting a topic! Let's do

a typological study of fourth person. But beware! As described in §15.1.6, the term ‘fourth person’ is used of a number of quite different phenomena which are scarcely comparable (see the summary in Fleck 2008). Employment of the same name does not imply that similar phenomena are being described.

6.3. A prime source for cross-linguistic information on vowel systems is Maddieson (1984).

6.4. The richest source of information on reduplication is Key (1965). There are a number of useful studies of reduplication in individual languages in Hurch (2005). That volume also includes a general survey article by Rubino, who does not seem aware of the existence of Key’s classic paper. Rubino includes some information that was not in Key, but overall Key is the superior source. Reduplication for nominative singular in Alutor is in Kibrik, Kodzasov, and Muravyova (2004: 217, 291).

6.5 On pages 522–9 of Haspelmath et al. (2005) there is some information, by Cecil H. Brown, on languages which have one lexeme covering both ‘hand’ and ‘arm’ and those which have two, and similarly for ‘hand’ and ‘fingers’. No account is taken of how ‘leg’, ‘foot’, and ‘toes’ are treated, or of any other part of the anatomy, which would be needed for the work to be of much consequence. See further comments on this volume below.

On kinship structures see, for example, Keesing (1975) and further references therein. A full description of the use of posture verbs as existentials in Enga is in Lang (1975). The chapters in Newman (2002) provide useful studies on verbs of sitting, standing, and lying in various languages.

6.6 Note also: ‘The relationship between sample size and population size does have some effect, but this is negligible unless the sample represents a large proportion of the population. It is in fact generally ignored in practice unless the sample size represents at least one-tenth of the total population, and it is also often ignored unless it represents one-fifth of the population. At this latter level the standard error, unless adjusted, would be overstated by about 11 per cent’ (Reichmann 1964: 235). In fact, I have never seen a sample used in linguistic work which is as large as one-fifth of the population.

The excellent grammar of Ungarinjin is by Rumsey (1982). Bybee et al. used Platt’s 67-page grammar of Gugada (1972). The finest grammar of a Western Desert dialect is Goddard’s 408-page study (1985). Other worthwhile grammars of Western Desert dialects—all better than Platt’s—include Douglas (1964), Hansen and Hansen (1978), Marsh (1976), Glass and Hackett (1970), Vászolyi (1979), and Eckert and Hudson (1988).

Immediately evident faults of Haspelmath et al. (2005) include: employing unreliable sources; employing only some (not all) of the good sources available

---

for a sample language; omitting all mention of certain areas where a given feature occurs; quoting sources erroneously; misinterpreting sources.

A further point can be made concerning the proper methodology for pursuing linguistic typology. Sometimes it is not necessary to examine absolutely every language with a certain feature. For example, there are over 300 languages in the tight-knit Bantu genetic group, virtually all showing systems of noun classes of recurrent types. When Aikhenvald was working on a typology of classifiers and noun classes (published in 2000) she deemed it sufficient to examine a representative selection of good grammars of Bantu languages.

There has recently appeared a peculiar label 'convenience sample', to describe making use of every available good-quality grammar in a typological study. This is not a sample at all, and the label is misbegotten.



# Phonology

This book focuses on grammar. But every description of a language must, of course, include a section on phonology, which interrelates with grammar in a myriad ways. As pointed out in §4.7, the first chapter to be drafted in any fieldwork-based endeavour must be phonology, since phonemes need to be recognized as the basis for an orthography in which to write the language. But the phonology chapter will be the last to be finalized, since the range of morphological alternations and conditioned variants will feed into a definitive statement of phonological structures and processes. This chapter presents just a brief overview of the most basic points which must be taken account of, to ensure a decent phonological description.

## 7.1 Phonemes

The ‘phoneme’, the basic concept of phonology, is amongst the most fundamental and important of linguistic units. The function of phonemes is to distinguish words of different meaning. If one phoneme is replaced by another, a new word is produced. Thus, replacing the initial /p/ in English *pin*, /pin/, by /b/ gives *bin*, /bin/, a quite different word.

Each phoneme has a number of variant pronunciations, called allophones. These variants depend on the position of the phoneme in an utterance (its phonological and phonetic environment), on the social context, on the speech style, and on whether the speaker is feeling tired or lazy, among other factors. Substituting one allophone of a given phoneme for another does not affect meaning; it will simply lead to a funny-sounding pronunciation of a familiar—and recognizable—word. When in word-initial position, English phoneme *p* is generally pronounced as an aspirated voiceless bilabial stop, [p<sup>h</sup>]. But after a consonant it lacks aspiration, [p]. One normally says and hears *pin*, [p<sup>h</sup>in] and *spin*, [spin]. If the allophones were interchanged, one would still recognize [pin] as the word *pin* and [sp<sup>h</sup>in] as *spin*.

Although the term ‘phoneme’ was only introduced in about 1880 (see §3.13), the idea goes back a very long way. The first alphabet is thought to have been devised about 3,500 years ago, near the eastern shores of the Mediterranean.

Almost all modern alphabets derive from this in one way or another. A notable exception is the Han'gŭl alphabet for Korean, commissioned by their king in 1446. The basic principle of an alphabet is to have one letter for each distinct phoneme. This property holds for the Han'gŭl alphabet, the Greek alphabet, and the Roman alphabet. But an alphabet which is fine for one language may be adopted for another for which it is less appropriate. And the way of speaking a language may change, while the convention for writing remains unchanged. When the Roman alphabet was first used for English it was reasonably adequate. Tradition fixed the written representation of words. And, although various processes of natural change caused their pronunciation to alter markedly, they continued to be written in the same way. In Chaucer's time the word pronounced as [na:mə] was appropriately written *name*. By Shakespeare's time it was said as [ne:m] and today we say [neim]; but the spelling has been kept the same.

An alphabet—originally with one letter for each phoneme—enables people to write. But then the permanence of writing, as opposed to the fluidity of speech, serves to make the alphabet less iconic as centuries roll by.

A speaker of a language hears it in terms of its phonemes, not the phonetic form a phoneme has in each environment. Speakers of English will aver that the *p* in *pin* is 'the same as' the *p* in *spin*. In a classic paper, 'The psychological reality of phonemes' (1933), Sapir described how this assisted him when working on Southern Paiute (from south-west Utah). He taught his consultant to write the language and, to Sapir's surprise, both [p] and [β] were represented by 'p', showing that they are variant pronunciations of one phoneme. Sapir then searched for—and discovered—factors which condition the occurrence of the two allophones.

The only sensible and effective way of working out the phonemes of a previously undescribed language is to work in close harmony with a native speaker. The same word, heard on different occasions, may sound different to a non-native linguist, but it must have the same phonemic form. When commencing work on Dyirbal, I transcribed 'stone' as [tiban] on one occasion and [diban] on another. Perhaps the contrast of voicing is not phonologically significant for this language? I tried pronouncing the word in four ways—[diban], [dipan], [tiban], [tipan]. 'Yes, those are all exactly the same.' My teacher sounded a trifle impatient that I could not see the obvious. She went on to give the word for man [yaɾa]. Having heard a trill in the language, I repeated the word with this, [yara]. 'No,' I was reprimanded, 'that means "fishing line".' In this language there is a contrast between two rhotic phonemes, an apico-alveolar trill /r/, as in /yara/ 'fishing line', and an apico-postalveolar continuant /ɾ/, as in [yaɾa] 'man'. Rhotic trill (as in Scottish pronunciation) and rhotic continuant (similar to that in standard English pronunciation) are allophones

of one phoneme in English. Whether one hears [ærou] or [ærou], the word is recognized as *arrow*. But in Dyrirbal they constitute different phonemes. In contrast, /p/ and /b/ are contrasting phonemes in English, but [p] and [b] are allophones of a single bilabial stop phoneme in Dyrirbal (and similarly for /t/ and /d/, and /k/ and /g/).

Nowhere in linguistic work is there a uniquely correct analysis for a given set of data. Competing analyses have different virtues and drawbacks. In balancing these, one analytic decision may be most appropriate for a certain purpose, and another for another. However, for most languages deciding on systems of consonant and vowel phonemes does come fairly close to constituting an analysis about which almost all linguists would be likely to agree. Nevertheless, there can be some issues for which alternative analyses—all plausible—may be provided. A phonetic long vowel [a:] could be treated as a separate phoneme /a:/, contrasting with short vowel /a/, or as a sequence of two short vowels, /a/ followed by /a/. Some linguists treat English word *chin* as commencing with a single consonant, a lamino-alveolar affricate (written /č/ or /tʃ/). Others say that this word commences with a consonant cluster, of apico-alveolar stop, /t/, followed by lamino-alveolar fricative (written /š/ or /ʃ/). Arguments can be given for both stances (see §7.4).

It should go without saying that two languages may have very similar sets of data, which require totally different analyses, depending on the overall structures of the languages. This was illustrated in §4.6 for phonetic long vowels, and for phonetic diphthongs. We saw that in Dyrirbal [a<sup>1</sup>] must be analysed as a vowel-plus-consonant sequence /ay/ whereas in Fijian it has to be treated as a sequence of two vowels /ai/. Sapir's classic paper 'Sound patterns in language' (1925) provides a more complex, and highly educative, example of this.

## 7.2 Consonants

Every language has a system of consonants. The idea of 'system' is critical. Each term in a phonological system has its phonetic realization—and its functional possibilities—relative to those of the other terms in its system. The system recognized for Dyrirbal includes a stop consonant written as /b/, and so does that for English. But in English /b/ contrasts with /p/, that is, it represents a *voiced* bilabial stop, whereas in Dyrirbal /b/ simply represents a bilabial stop, which can have voiced or voiceless pronunciation, [b] or [p]. The Dyrirbal phoneme could be written as either /b/ or /p/; the most frequent allophone is the voiced one and for that reason /b/ was selected.

One has to specify more than just the total inventory of consonants for a language. There are likely to be a number of subsystems, each corresponding

to a different slot in the structure of a syllable or of a phonological word. As mentioned in §5.7, the subsystem at word-initial and word-medial positions in German includes both /b/ and /p/; word-finally, there is a single bilabial stop, without any voicing contrast (and similarly for other places of articulation). In Mandarin Chinese, twenty-one of the twenty-two consonants (all save /ŋ/) occur syllable-initially, but at the end of a syllable we find only /n/ (which also occurs initially) and /ŋ/. As mentioned in §6.3, in the Australian language Martuthunira, the full set of twenty consonants only contrasts between vowels. Word-initially there is a subsystem of just ten and word-finally six; only the lamino-palatal nasal, /ɲ/, may occur both initially and finally; see §7.4 (Dench 1995: 30–4; Dixon 2002: 553–7).

In any statement of consonants—as of vowels, of tones, and so on—the phonetic nature of each must be clearly stated. There is nothing worse than being presented with a tidy table of consonants which is entirely lacking in labels for its rows and columns; or being told that ‘language X has phoneme *d*’, with no statement whatsoever of the phonetic nature of this *d*—whether it is apico-dental or apico-alveolar or lamino-dental or lamino-alveolar or whatever.

It is the established custom to present the full set of consonants in a language in a matrix, with the columns showing place and the rows manner of articulation. A sample matrix (for no language in particular) is in Table 7.1; note that this only illustrates some of the possibilities.

TABLE 7.1. A sample consonant matrix

| ACTIVE<br>ARTICULATOR         | apico-         |          |              | lamino- | dorso- | labio-         |        | glottal |
|-------------------------------|----------------|----------|--------------|---------|--------|----------------|--------|---------|
|                               | dental         | alveolar | postalveolar | palatal | velar  | labial         | dental |         |
| unaspirated<br>voiceless stop | t              |          |              |         | k      | p              |        | ʔ       |
| aspirated<br>voiceless stop   | t <sup>h</sup> |          |              |         |        | p <sup>h</sup> |        |         |
| voiced stop                   | d              |          |              | ʃ       |        | b              |        |         |
| voiced affricate              |                |          |              | dʒ      |        |                |        |         |
| voiceless fricative           |                | s        |              |         |        |                | f      | h       |
| nasal                         | n              |          |              | ɲ       | ŋ      | m              |        |         |
| lateral                       |                | l        |              | ʎ       |        |                |        |         |
| rhotic                        |                | r        | ɻ            |         |        |                |        |         |
| semi-vowel                    |                |          |              | y       | w      |                |        |         |

**(a) Place of articulation**

An active articulator is brought into contact with—or into approximation with—a passive articulator. There is a pervasive tendency (one could say, a tradition) to state just the passive articulator; hence ‘place of articulation’. The presumption is that each active articulator will associate itself with that passive articulator which is closest to it. This presumption is unfounded—there are several possible alternative destinations for each active articulator. Omitting to provide a full specification, as is shown in Table 7.1, is simply lazy (no matter that such laziness is a tradition).

PASSIVE ARTICULATORS are:

|                                      |          |          |
|--------------------------------------|----------|----------|
| upper lip                            | labelled | labial   |
| upper teeth                          |          | dental   |
| gum ridge                            |          | alveolar |
| front (hard) part of palate          |          | palatal  |
| back (soft) part of palate, or velum |          | velar    |
| uvula                                |          | uvular   |

(Note that the uvula is passive articulator for dorso-uvular stops and fricatives. But in a uvular trill, the uvula is in motion and should then be considered the active articulator.)

ACTIVE ARTICULATORS are (together with the passive articulators they typically combine with):

|                                                                                     |                                                |                                                                                 |
|-------------------------------------------------------------------------------------|------------------------------------------------|---------------------------------------------------------------------------------|
| lower lip (labio-)                                                                  | with upper lip:                                | bilabial (that is, labio-labial)                                                |
|                                                                                     | with upper teeth:                              | labio-dental                                                                    |
| tip of tongue (apico-)                                                              | with upper teeth:                              | apico-dental                                                                    |
|                                                                                     | with gum ridge:                                | apico-alveolar                                                                  |
|                                                                                     | with hard palate:                              | apico-palatal                                                                   |
| tip of tongue turned back so that the under side touches the back of the gum ridge: |                                                | apico-postalveolar (also called ‘retroflex’, describing the bend in the tongue) |
| blade of tongue (lamino-)                                                           | with upper teeth (or with both sets of teeth): | lamino-dental                                                                   |
|                                                                                     | with gum ridge:                                | lamino-alveolar                                                                 |
|                                                                                     | with hard palate:                              | lamino-palatal                                                                  |
| back of tongue (dorso-)                                                             | with hard palate:                              | dorso-palatal                                                                   |
|                                                                                     | with velum:                                    | dorso-velar                                                                     |
|                                                                                     | with uvula:                                    | dorso-uvular                                                                    |

Moving from the back of the mouth, down the throat, there are two further ‘places’ of articulation. Fricatives may be pharyngeal, with the root of the

tongue (or the epiglottis just below it) being pushed towards the back wall of the pharynx. Further down are the ‘vocal lips’ (also called ‘vocal cords’ or ‘vocal folds’) which vibrate to create a voiced sound. The vocal lips may be brought tightly together and then released, producing a glottal stop. Or the lips may be held just apart, the noise of air as it is squeezed between them producing a glottal fricative.

These brief descriptions are far from exhaustive. More intricate specifications invoke labels with ‘pre-’ and ‘post-’. Some sounds have ‘double articulation’; for example, a stop may involve simultaneous dorso-velar and bilabial closure and/or release. There are various kinds of combinations of mechanisms—involving glottalization, palatalization, labialization, and more.

### (b) Manner of articulation

The basic manners are, with useful labels which group them together as natural classes:

|           |   |           |   |        |
|-----------|---|-----------|---|--------|
| stop      | } | obstruent |   |        |
| affricate |   |           |   |        |
| fricative |   |           |   |        |
| nasal     | } | sonorant  |   |        |
| lateral   |   |           | } | liquid |
| rhotic    |   |           |   |        |
| semivowel |   |           |   |        |
|           |   |           |   |        |

(Note that, for some linguists, vowels are also classed as sonorants.)

An obstruent involves an obstruction to the air flow at or above the larynx. For a stop, there is complete closure of active against passive articulator. For a fricative, a small gap is left between articulators, creating audible friction—or turbulence—as air is pushed through it. (This is a little like rapids on a river.) An affricate is a stop released into a fricative.

For a sonorant, the air flows freely, with no obstruction, but is directed in a particular way. For nasals the velum is lowered so that air exits through the nose; there is a closure at the lips or in the mouth forming a closed chamber against which the air passing into the nose resonates. (Since there is closure, nasals are sometimes called ‘nasal stops.’)

Laterals involve the centre of the tongue being raised so that air flows past one or both sides. For prototypical rhotics the sides of the tongue are raised so that air flows through a trough in the middle. ‘Liquid’ has come into use as a convenient cover term for the natural class of laterals and rhotics. (When young children cannot yet produce a rhotic, they typically substitute a lateral.)

The notion ‘rhotic’ is somewhat complex, and not without controversy. It is most usefully regarded as the label for a ‘family of sounds’, each of which has

resemblance to some of the others without there being any one factor common to all. The most prototypical—and probably cross-linguistically the most common—rhotic is an apico-alveolar trill in which the sides of the tongue are raised and the tip, placed close to the gum ridge and in a relaxed mode, vibrates as air is pushed past it. A trill with just one vibration is a tap. A uvular trill is similar in that vibrations are involved, but here the uvula vibrates against the back of the tongue. The most common rhotic across dialects of English is a continuant, where air flows smoothly through a trough in the tongue.

All of these (and other) varieties of rhotic have naturally been written with variants of the letter ‘r’ (r, ɾ, ɹ, ɻ, R, and so on), showing an intuitive connection between them. They tend to alternate as allophones of a single phoneme, to play similar roles in phonological rules (such as dissimilation), and often to occur in the same slots both within consonant clusters and in syllable and word structure.

A semivowel has similar articulation to a high vowel, but belongs to the consonant system. That is, it occurs not in the nucleus of a syllable, as a vowel does, but precedes or follows this (as syllable onset or coda). The laminal semivowel—/j/ in the IPA convention, but very often written /y/, as it is in Table 7.1—is similar to unrounded high front vowel /i/. In a couple of Australian languages there are two series of stops and nasals made with the blade of the tongue, and corresponding to these, two *y*-type semivowels, lamino-dental and lamino-palatal.

The semivowel written as ‘w’ typically involves both the back of the tongue raised towards the velum and the lips brought together (either flat or rounded). It thus relates to both dorso-velar and bilabial columns in a matrix such as that in Table 7.1. If the bilabial column is on the left and the dorso-velar column towards the right of the matrix, this leads to difficulty of placement. Linguists sometimes write *w* in both columns, often with parentheses around one occurrence to indicate that a single phoneme is involved. Table 7.1 shows a way out of this dilemma—position the bilabial column next to dorso-velar and write ‘w’ between them. (‘Peripheral’ is a cover term for bilabial and dorso-velar, sounds made at the periphery of the mouth’.)

There can be further semivowels—bilabial contrasting with dorso-velar, a semivowel corresponding to a rounded front high vowel (as in French), and so on.

Some linguists—paying more attention to form than to function—associate a continuant rhotic with semivowels (calling this class ‘glides’ or ‘approximants’). This has been suggested by a minority of the linguists working on Australian languages, most of which have two rhotics, one apico-alveolar (typically realized as a trill) and the other apico-postalveolar (typically a continuant). That is, instead of analysis I—that followed here—they prefer II, which does not recognize any commonality between the two r-sounds:

|            |                      |         |                |                    |
|------------|----------------------|---------|----------------|--------------------|
| I          | dorso-velar/bilabial | laminal | apico-alveolar | apico-postalveolar |
| rhotic     |                      |         | r              | ɹ                  |
| semi-vowel | w                    | y       |                |                    |
| II         | dorso-velar/bilabial | laminal | apico-alveolar | apico-postalveolar |
| trill/tap  |                      |         | r              |                    |
| glide      | w                    | y       |                | ɹ                  |

There are some arguments in favour of analysis II but many more for I. The two rhotics do act as a natural class in phonotactic structure and in diachronic changes. In addition, only analysis I provides a satisfactory framework for the phonetic realizations encountered. Consider, for example, the two rhotic phonemes in Yidiñ:

- r apico-alveolar articulation—generally a trill, sometimes reduced to a single tap
- ɹ apico-postalveolar articulation—most often a grooved continuant but can be a trill, typically at the end of a stressed syllable (in one dialect there is a fair chance of encountering an apico-postalveolar trill in any position within a phonological word)

This shows that the critical factor distinguishing the two phonemes is place, not manner, of articulation.

Sounds are made by air moving through the mouth or nose or both. Something has to set the air in motion. All the sounds described above involve air being pushed out by the lungs, called the ‘pulmonic airstream mechanism’. Air movement can be initiated at the glottis, producing ejective stops and fricatives (sometimes misleadingly called glottalized consonants) when the air is pushed out, and implosive stops when it is pulled in. Or an ingressive airstream can be initiated by the back of the tongue against the velum, producing clicks.

The various manners of articulation may be combined in a variety of ways, giving lateral fricatives and affricates, prenasalized stops, prestopped nasals, and many more.

### (c) Associated parameters

- (i) VOICING. For every sound made on a pulmonic airstream, the glottis can be open (voiceless) or vibrating (voiced). Many languages have contrasting voiced and voiceless phonemes for some or all of the obstruent manners (stop, fricative, affricate). There may be a voicing contrast for sonorants but these are fairly rare. We also get whispered voice (with a narrowed glottis) and creaky voice (with the vocal lips vibrating very slowly), but these are only occasionally phonologically contrastive.



- (ii) **ASPIRATION.** There may be friction at the glottis as a lip or mouth closure is released, providing an aspirated tinge at the end of the stop (which may be voiced or voiceless). These are aspirated stops, which occur as allophones of voiceless stops in English but constitute contrastive phonemes in other languages (including Sanskrit). Or glottal friction may precede closure, producing pre-aspirated stops.

Each phoneme naturally has a number of components—the physical elements of place of articulation (active and passive articulator), manner of articulation, voicing, aspiration, and so on. These have acoustic correlates. In addition, the sounds made at the peripheries of the mouth—bilabial and dorsovelar—share acoustic properties; we have seen how semi-vowel *w* typically relates to both.

These component features form systems of oppositions. We owe to Trubetzkoy (1939) the classification of oppositions as ‘privative’, involving the presence or absence of a feature, ‘gradual’, showing degrees of a property, and ‘equipollent’, the features being logically equivalent. These were described in §5.7. As stated in §2.4, it is a grave error to try to reduce all oppositions to binary form; some naturally are so, others definitely are not.

In a privative opposition, the member showing a certain feature is said to be ‘marked’ with respect to the other—the ‘unmarked’ member—which lacks the feature. (The difference between formal and functional markedness is explained in §5.7.) A privative opposition which applies at some structural positions may be neutralized at others. It was mentioned that in German voiced and voiceless stops contrast initially and medially but not finally. In this position just the voiceless articulation, the unmarked member of the opposition, occurs (the unit resulting from neutralization of two established phonemes in the language is often called an ‘archiphoneme’).

It is sometimes said that if a language completely lacks a voicing distinction at the phonological level—with [b] and [p] being allophones of one phoneme, and so on—then one should use the letter corresponding to the unmarked member of the opposition in languages for which it is contrastive. That is [b]/[p] should be written as /p/. But as was mentioned in §7.1, for Dyrbal the [b]/[p] phoneme is written as /b/—and similarly for /d/, /ʒ/, /g/—since by far the most frequent allophone is voiced. In fact, the normal state of affairs in Dyrbal is for the glottis to be in vibration for all speech—always for vowels, semi-vowels, rhotics, laterals, and nasals, and almost always for stops (just occasionally, glottalic vibration may commence slightly after the beginning of an utterance, producing a voiceless stop allophone). In Dyrbal, the unmarked

state of the glottis is to be vibrating. The lesson here is that each language must be treated as a system in its own right, and described accordingly.

A phonological description should commence with statement of the systems of consonants and vowels. And then the range of phonetic realizations of each phoneme, in every phonological environment, must be stated. It may be considered appropriate to quote minimal pairs which demonstrate phonemic contrasts that are unusual cross-linguistically—for instance, if there are three or four contrasting lateral phonemes. Generally speaking, however, it is not necessary to provide ‘minimal pairs’ for every possible pair. A grammar should always include a vocabulary, ideally including all words and affixes which occur in examples and texts throughout the work. The interested reader will be able to use this to investigate any contrast which interests them.

### 7.3 Vowels

There are three major parameters involved in the production of vowels, those sounds which occur as the nucleus of a syllable:

- (a) The height of the tongue—how far it is raised towards the roof of the mouth. Labels used here are ‘close/open’ or ‘high/low’ (they are equivalent).
- (b) The horizontal position of the part of the tongue that is raised. Most languages have a two-way distinction between ‘front’ (towards the hard palate) and ‘back’ (towards the velum). There may be a third ‘central’ position between these.
- (c) Whether the lips are rounded or unrounded.

Parameters (b) and (c) interrelate. Front vowels are generally unrounded; there may in addition be rounded front vowels (seldom or never do we find rounded front vowels without there also being an unrounded series). Back vowels are typically rounded, but in some languages there is just one series of back vowels, which is unrounded. There may be two series of back vowels, one rounded and the other unrounded. The eight Cardinal Vowels (of Daniel Jones) are set out in Table 7.2.

Just as with consonants, so it is essential to provide full articulatory specification for each vowel phoneme. One sometimes reads that a language has ‘back vowels *u* and *o*’ without it being specified whether they are rounded or unrounded. (The IPA symbols for unrounded close and close-mid back vowels are /u/ and /ɤ/. It is often convenient to employ regular letters of the Roman alphabet ‘u’ and ‘o’, but it should be stated that they are being used, in this instance, for unrounded back vowels.)

TABLE 7.2. Cardinal vowels

| front     |                      | back    |
|-----------|----------------------|---------|
| unrounded |                      | rounded |
| i         | close = high         | u       |
| e         | close-mid = high-mid | o       |
| ɛ         | open-mid = low-mid   | ɔ       |
| a         | open = low           | ɒ       |

There may be a close central vowel, either unrounded /ɨ/ or rounded /ʉ/; this is especially common in languages from Amazonia. The schwa vowel, /ə/, is central on both vertical and horizontal axes (and unrounded). In many languages one or more vowels are pronounced as /ə/ in certain environments (typically, in an unstressed syllable), such that the underlying identity of the vowel can only be ascertained by, say, adding an affix and changing the stress pattern of the word. In addition to this, /ə/ may also function as a contrastive phoneme; it can be a subtle matter to distinguish these two kinds of schwa.

There are further features which can characterize vowels. The velum may be lowered so that air escapes through nose as well as mouth, giving nasalized vowels. In some languages (especially in Africa) there are two positions for the tongue root—advanced or retracted, creating distinct series of vowel phonemes (the vowels of each series may be linked by their role in vowel harmony). Other minor features of vowel quality are discussed in Ladefoged and Maddieson (1996: 298–320).

A full description of the allophones of each vowel phoneme, and their conditioning factors, must be provided. It is often convenient to map vowel realizations onto the Cardinal Vowel diagram. This can either involve a cross for the major allophone of each phoneme (indicating its position with respect to back/front and close/open parameters), or an area, indicating the overall allophonic range. In a small system of, say, three vowels, each phoneme may occupy considerable phonetic space (for example, /i/ may include [i], [e], and [ɛ]), whereas for a larger system the phonetic space of each vowel will be consequently smaller. In essence: the realizational possibilities of vowel phonemes tend to expand to fill the phonetic space available.

There can be overlapping realizations of phonemes. For example, in a three-vowel system:

|     |                                               |                        |
|-----|-----------------------------------------------|------------------------|
|     | allophone after a<br>lamino-palatal consonant | allophone<br>elsewhere |
| /i/ | [i]                                           | [ɛ]                    |
| /a/ | [ɛ]                                           | [a]                    |

These never lead to confusion. One can tell which phoneme the sound [ɛ] relates to from the environment it occurs in. There may also be allophonic overlap between consonants. In American English, an apico-alveolar flap can be an allophone of /t/ intervocalically after a stressed vowel, and an allophone of /r/ after /θ/, as in *three* /θri:/. Again, the allophones occur in totally different environments, so that there is never any uncertainty of interpretation.

Semivowels were described, in §7.2, as having similar articulation to high vowels but functioning within the consonant system. (Interestingly, in all the languages I know in which the high back vowel is unrounded, the corresponding bilabial/dorso-velar semivowel is also unrounded.) Other sonorant consonants may also on occasion function as syllabic nucleus. For example, in English we find syllabic /l/, /n/, and /m/, as in (using '˘' to mark syllable boundary) *bottle* /'bɒt.l/, *sudden* /'sʌd.n/, and *rhythm* /rið.m/. And, in rhotic dialects, continuant /r/ can be syllabic, as in *butter* /'bʌt.r/.

## 7.4 Phonotactics and syllables

An essential component of the phonological description of a language is statement of which systems and subsystems of consonants and vowels correspond to the various slots in the structure of syllable and of phonological word. This is 'phonotactics'.

In many languages, the structure of a syllable is just CV—or (C)V—with a phonological word consisting of one or more of these syllables. Here 'C' indicates a consonant and 'V' a vowel (which can be a short vowel, a long vowel, or a diphthong). It is then likely that every consonant will correspond to each C slot, and every vowel to each V slot. But, even with this simple a structure, there can be constraints. For example, it might be that laterals and/or rhotics may occur only in a non-initial C slot; that is, a word cannot commence with one of these segments.

When syllable structure is CV(C)—or (C)V(C)—we get sequences of two consonants in the middle of a word. The possibilities are:

monosyllabic word:  $C_1V_1(C_2)$

disyllabic word:  $C_1V_2C_3C_4V_3(C_2)$  or  $C_1V_2C_5V_3(C_2)$

There may be the same possibilities at  $V_1$ ,  $V_2$ , and  $V_3$ . Or these may vary. If the language has initial stress we would expect the same vowel possibilities at  $V_1$  and  $V_2$ ; if it bears stress on the final syllable, we would expect the possibilities for  $V_1$  and  $V_3$  to coincide. (There may be a fuller system of vowels in a stressed than in an unstressed syllable; never the reverse.)

Looking now at consonants, the possibilities at word-initial position,  $C_1$ , are likely to be the same for all three structures, and similarly for word-final

position,  $C_2$ . If the same syllabic template applied throughout, we would expect  $C_1$  and  $C_4$  to coincide, and also  $C_2$  and  $C_3$ . In fact, this is seldom the case. The Australian language Martuthunira was mentioned in §6.3 and §7.2. Here, every phonological word must have at least two syllables. There are two apical, two laminal (lamino-dental and lamino-palatal), and two peripheral (bilabial and dorso-velar) series for stops and nasals; two apical and two laminal series for laterals; two apical series for rhotics; and the standard two semi-vowels—twenty consonants in all. The subsystems for consonant slots in phonological word structure are:

- $C_5$  All twenty consonants.
- $C_1$  Just ten—laminal and peripheral stops and nasals, plus the two semivowels.
- $C_4$  Just six—peripheral stops and nasals, plus the two semi-vowels.
- $C_2$  and  $C_3$  Just seven—stops and laterals at apico-alveolar, apico-postalveolar, and lamino-palatal positions, plus the apico-alveolar rhotic.

Martuthunira has three vowels, which contrast at both  $V_2$  and  $V_3$  slots (Dench 1995: 30–4; Dixon 2002: 554).

Very many languages have distinct subsystems of consonants applying at different structural positions. There are a number of recurrent cross-linguistic tendencies. For example, nasal-stop consonant clusters in the middle of a word are likely to be either predominantly or strictly homorganic, having the same place of articulation (*-mb-*, *-nd-*, *-ŋg-*) rather than heterorganic, with different places of articulation (*-md-*, *-ŋb-*, etc.). With regard to initial and final positions within a phonological word:

### Initial Position

- PLACE OF ARTICULATION. Apicals are likely not to feature at all, or else in a fairly small number of words.
- MANNER OF ARTICULATION. Stops are frequent, laterals and rhotics infrequent or missing.
- VOICING. If there is a voicing contrast, it is likely to apply initially.

If a language has prenasalized stops or pre-aspirated stops (as phonemes or allophones of phonemes), these are most likely to occur at initial position.

### Final Position

- PLACE OF ARTICULATION. Apicals are very common. There may be few or no peripherals.

- MANNER OF ARTICULATION. A favoured position for laterals and rhotics, disfavoured for stops.
- VOICING. Voicing contrast is often neutralized finally.

Nasals generally occur in both initial and final slots. In a few languages—including English and Mandarin Chinese—dorso-velar nasal  $\eta$  has special status, being restricted to syllable- or word-final position. In other languages,  $\eta$  is initial segment for many words.

Some languages do, of course, allow consonant clusters at the beginning and end of words. In §1.3, the CCCVCCC pattern in English was illustrated with *strengths* /streŋθs/. (Some people even pronounce it as /streŋkθs/, with a final sequence of four consonants.) As a rule, the more consonants there are in a cluster, the more limited the possibilities are likely to be at each structural slot within it. For example, initial  $C_1C_2C_3$ —in English allows only *s* at  $C_1$ , *p*, *t*, or *k* at  $C_2$ , and *r* or *l* at  $C_3$ .

The natural structure of the phonotactics of a language can help resolve queries concerning phonological analysis. The initial sound in English words like *chin* was mentioned in §7.1—it could be analysed either as one phoneme or as a sequence of two, stop plus sibilant. One criterion which can help in deciding between these alternatives comes from phonotactics: there are in English no other initial sequences of voiceless stop plus fricative. We do find other initial clusters commencing with /t/, for example /tr/. Parallel to this are /pr/, /kr/, /θr/, and /fr—every voiceless stop and non-sibilant fricative may be followed by /r/. The lack of pattern supporting a cluster of /t/ plus /ʃ/ suggests that /tʃ/ should here be treated as a unit phoneme. And speakers of English do have intuition that this /tʃ/ is a ‘single sound’ (one way in which this is shown is when speakers dictate slowly, letter by letter). We can get /t/ at the end of one word and /ʃ/ at the beginning of the next, as in *heat sheets* /hi:t ʃi:ts/. But speakers clearly distinguish this sequence from the *ch* in *he cheats* /hi: tʃi:ts/, providing further justification for /tʃ/ as a phoneme. (This would be written /č/ in Americanist phonetic notation, which is a clearer reflection of the decision not to treat it as a sequence of phonemes.)

In a number of languages, different grammatical word classes show varied phonotactic possibilities. For the Australian language Nyawaygi, only verbs (not nouns) may have monosyllabic roots. In Nyawaygi’s northerly neighbour Warrgamay, the only monosyllable roots are nouns (plus *ŋa:* ‘not’). Manambu, from New Guinea, is similar to Warrgamay. In the north Amazonian language Tariana the only roots of more than one syllable are nouns.

And types of morphological element may have different possibilities. Merlan (1994: 6, 15) reports a standard five-vowel system for the Australian language Wardaman, but states that the mid vowels, *e* and *o*, do not occur in

inflectional morphology (perhaps suggesting that *e* and *o* may have come into the language relatively recently).

The lungs expel air in a sequence of pulses (often called 'chest pulses'). It is generally accepted that the crest of each pulse is the nucleus of a syllable, realized by a vowel or a syllabic sonorant consonant. (This is something of a simplification, there being a fair amount of debate about what exactly a syllable is, in articulatory terms.)

Every syllable must have a nucleus, and the great majority of syllables across the languages of the world have a consonant or consonant cluster as onset. Even if a language does allow V or V(C) syllables, there are invariably a much greater number of syllables of CV or CV(C) type. The Australian language Olgolo was mentioned in §1.3 as having lost all initial consonants, giving rise to a V(C) profile. But then what were generic classifiers reduced to be single-consonant prefixes to nouns, restoring the CV(C) template.

In §4.6 we described how phonetic diphthong [a<sup>i</sup>] demands different phonological analyses in different languages. Dyirbal has no vowel sequences and here [a<sup>i</sup>] is phonological /ay/. This analysis is justified by the fact that a word ending in [a<sup>i</sup>] takes the case allomorphs which follow consonants, not those which come after vowels. In contrast, Fijian has entirely open syllables, (C)V, and here phonetic diphthong [a<sup>i</sup>] is analysed as a sequence of two vowels /ai/. Each word in Fijian must be of at least two moras, with a short vowel counting as one and a long vowel as two moras; the verb 'see' is /rai/, [ra<sup>i</sup>].

The number of phonetic syllables corresponding to a phonological sequence of vowels may also vary, both within a language and between languages. In Fijian, certain vowel sequences (*a*, *e*, or *o* followed by *i* or *u*, plus *i* followed by *u*) are pronounced as diphthongs with the remainder constituting a sequence of two phonetic syllables. For example, /niu/ 'coconut' is pronounced as a single syllable [n<sup>i</sup>u] where /nui/ 'hope' is disyllable [nu.i] (syllable boundary being shown by '·'). Jarawara, from southern Amazonia, has identical syllable structure to Fijian, (C)V. But here there are no diphthongs, each sequence of vowels being a sequence of phonetic syllables; thus /sai/ 'be audible' is [sa.i].

One can always tell how many syllables a word has, but it is not always easy to decide where one syllable ends and the next begins. Dyirbal has a single consonant at the beginning of a phonological word and an optional single consonant at the end, but up to three consonants between vowels, as in *gulmbin* 'sword' (and now also: 'cross-cut saw') and *gulnbin* 'noise of children splashing in water'. We would in each case take *l* to end the first syllable and *b* to begin the second one. But what about medial *m*, at the same place of articulation as the following *b*, and *n*, homorganic with the preceding *l*? No principled decision can be made as to *where* the syllable boundary should be placed, and in fact this does not matter. How many syllables a word contains

is critical, for stress placement and allomorphic selection, but the actual point of division is of little consequence. (One common way of marking stress is to place an accent on the vowel of a stressed syllable. The other is to write ' at the beginning of the stressed syllable; this would present difficulties in the case just described.)

## 7.5 Prosodies

Words are distinguished from each other by the substitution of consonant and vowel phonemes in structural slots. And there are other ways. Some phonological contrasts have scope over a sequence of segments—for example, the beginning or the end of a syllable, a complete syllable, a full phonological word. These are called 'suprasegmental systems' or, more felicitously, 'prosodic systems'. They include tone, glottalization, nasalization, rhoticization, labialization, and vowel harmony (and also stress, discussed in §7.6).

Every syllable has a pitch, the rate of vibration of the vocal cords; this is raised by greater tension in the cords and/or increased air flow through them. Pitch differences characterize an individual's accent (females tend to have higher pitch than males) and may vary with emotional state. In some languages there is a contrastive system of pitch patterns, called 'tone', such that substituting one tone for another produces a different word (in the same way that substituting one consonant or vowel for another does). Some languages have 'register tone'—each syllable must choose between high and low tone, or between high, mid, and low. Others have contrasts involving gliding movement, 'contour tones'. For example, in Mandarin Chinese the syllable *ma* can mean 'mother' with high level, 'hemp' with high rising, 'horse' with low falling rising, and 'scold' with high falling tone.

Generally, in a tone language each syllable of a word will have its own tone. Although properly a prosody of the syllable, it is often convenient to write tone as an accent or diacritic to the vowel which is syllable nucleus. When writing a language, people usually represent the consonant and vowel phonemes, but not untypically omit marking other features. If vowel length is contrastive, it is essential to mark it. Stress, when it is not predictable, must be marked. And not to specify the tone of each syllable is like missing out a feature which distinguishes consonants or vowels (for example, writing *raN* for all of *ram*, *ran*, and *rang*).

Just as languages may have varied segmental phonology for different word classes, so tonal assignment may vary. In Lango, from Uganda, each noun, adjective, and preposition has its own lexical tone, whereas a verb has tone determined not by the root but through aspect/mood marking.

As mentioned several times before, any set of linguistic data is likely to be open to competing analyses, which must be weighed and evaluated. For Sanskrit,



W. S. Allen (1951) suggests that ‘retroflex’ is not a place of articulation, but instead a prosody which may extend over a sequence of segments, up until the end of a word, that include apico-alveolar consonants. In *tāṇ ḍimbhān*, the fact that retroflex extends over two contiguous segments is shown by:

tāṇ ḍimbhān  
R

where R represents the prosody of retroflexion, and the line above its scope. Allen gives the rule for determining the scope of a retroflex prosody over a number of apico-alveolar consonants, even though they may be separated by a vowel or one of a limited set of consonants. For example, *niṣanna* is written:

niṣanna  
R

There are many other types of prosody which apply over a phonological extent which is more than a single segment. These include glottalization. In the Australian language Djapu, the glottal stop appears at first sight to be a segmental phoneme which may occur only at the end of a syllable. But examination of the mechanics of case marking on nouns reveals quite a different picture. First consider nouns which do not involve a glottal stop. Ergative allomorphs are *-y* after a vowel, *-thu* after a nasal, and *-yu* after a liquid:

|          | ROOT     | ERGATIVE FORM |
|----------|----------|---------------|
| ‘eye’    | maṇutji  | maṇutji-y     |
| ‘hand’   | goŋ      | goŋ-thu       |
| ‘tongue’ | ŋu:rnarr | ŋu:rnarr-yu   |

Now consider roots that end with a glottal stop:

|         | ROOT     | ERGATIVE FORM |
|---------|----------|---------------|
| ‘knife’ | yikiʔ    | yiki-yʔ       |
| ‘hook’  | bekaŋʔ   | bekaŋʔ-thu    |
| ‘fat’   | djukurrʔ | djukurrʔ-yu   |

It can be seen that the allomorphs are determined by the final segment of the root, ignoring ʔ. And, most significantly, the glottal stop comes after the case suffix *-y* on ‘knife’. In fact, glottal stop always occurs at the end of a syllable, irrespective of where a morpheme boundary comes. In this language, glottalization is most appropriately treated as a syllable prosody, realized at the end of the syllable.

## 7.6 Stress

For the great majority of languages which do not have a contrastive prosody of tone (and for some of those that do), there is a prosody of stress (or accent)

applying to syllables. A stressed syllable is perceived as involving more energy than an unstressed one. Any or all of a number of factors may contribute to this impression—loudness (due to greater muscular effort), vowel quality (generally, less centralized), (higher) pitch, and length. In languages which lack a contrast between long and short vowels, greater length is likely to be one feature of a stressed syllable; this is less likely when there is a length contrast for vowels. Where stress is shown largely by tone, the language is said to have ‘pitch accent’ (this may well develop into a prosody of tone). The complementary term for a language in which pitch is not a major factor is ‘stress accent’. (These terms are confusing, but they are in general use.)

As mentioned in §6.3, in some languages stress is contrastive. That is, if it is shifted from one syllable to another, a new word is created. For example, in English we have related noun *import* /'ɪmpɔ:t/ and verb *import* /ɪm'pɔ:t/, and the unrelated *billow* /'bɪləʊ/ and *below* /bɪ'ləʊ/. Some well-known languages have contrastive stress but do not mark it in their writing system (for example, Russian and English). Others—including Portuguese and Spanish—do. In writing a grammar of a previously undescribed language for which stress is contrastive, it is essential that it be systematically marked.

Stress applies to syllables (generally made up of several segments) and is a prosody, quite distinct from consonants and vowels, which are phonemes. This essential difference has been confused by some linguists (now mostly in the past) who talk of ‘suprasegmental phonemes’ of stress and also tone (sometimes calling these ‘tonemes’). Indeed, people who refer to every kind of contrast as phonemic naturally—and unfortunately—use the label ‘phonemics’ for what is more generally called ‘phonology’.

In most languages, stress is predictable; it always falls on a specifiable syllable (or syllables) of a word. The rules determining which syllable bears stress must be stated. Sometimes this is straightforward—on the first syllable, or the last, or the penultimate. Other times the matter is more complex, and depends on the internal structure of syllables. The unit ‘mora’ is useful for describing the placement of stress. There are two main ways in which ‘mora’ is used:

- (a) Each short vowel counts as one mora (in a sequence of vowels, each counts as one mora), and a long vowel counts as two moras.

This applies for Fijian, as described in §4.6. Here the syllable which includes the penultimate mora of a word bears stress; for example, intransitive *córi* ‘be tied’ and transitive *coríta* ‘tie’; intransitive *tádra* ‘dream’ and transitive *tadráa* ‘dream of’; intransitive *rái* ‘look’ and transitive *ráica* ‘look at, see’. In *rái*, the penultimate mora is the *a*, and stress goes on the diphthong *ai*; in *ráica*, the penultimate mora is the *i*, and stress again goes on the diphthong *ai*.

For a language with syllable-final consonants, ‘mora’ may be defined:

- (b) An open syllable with a short vowel counts as one mora, and with a long vowel as two moras. A closed syllable (ending in VC) counts as two moras.

In Latin, stress is on the syllable which includes the penultimate mora before the final syllable. Thus we get *dóminus*, with stress on the antepenult, when the penultimate syllable has a short vowel and is open, thus consisting of one mora. And we get *festí:na* and *agénda*, with stress on the penult when this has a long vowel, or is closed, and consists of two moras.

In §7.1, we broached the question of whether a long vowel should be described as a sequence of two identical short vowels, or as a distinct entity, contrasting with a short vowel. Different analyses are appropriate for different languages, depending on the manner of operation of phonological rules, including stress placement. If, say, stress goes on the penultimate syllable irrespective of whether that or any other syllable has a short or a long vowel, this is cause for treating a long vowel as a unit on a par with a short vowel, and writing it as, say, *ā* or *a:* (the latter is easier to handle in terms of what is available on a standard keyboard). But if a long vowel counts as two moras and a short vowel as one, as in the Fijian and Latin examples just described, then it is appropriate to describe it as, effectively, a sequence of two short vowels, and so write it by a doubled letter, say *aa*.

A longish word will often have more than one stressed syllable. Typically, alternate syllables may be stressed, counting from the specified first stressed syllable (which is determined by a rule similar to those described). This organizes a phonological word in terms of ‘feet’, each made up of one stressed and one unstressed syllable. There may be constraints; for example it is not uncommon for there to be a rule that a final syllable may not be stressed. It is sometimes useful to distinguish between ‘primary stress’ (the syllable selected by the stress rules) and ‘secondary stress’ (those in alternate syllables counting from the primary stress, or determined in some other way).

It must always be borne in mind that every component of the description of a language interconnects with the others. In some languages certain morphemes bear inherent stress while for others stress may be assignable by phonological rules. And stress rules can relate to morphological divisions within a word, with specification of which syllables are stressed then assisting in grammatical analysis. In the Australian language Warlpiri, stress falls on the first syllable of a root and on the first syllable of a suffix, unless it is monosyllabic and occurs word-finally. The root *yaparla* ‘father’s mother’ may take ablative suffix *-ngurlu*, yielding *yaparlangurlu*. The root *yapa* ‘person’ can take derivational suffix *-rlangu* ‘some, for example’ and then ergative case

inflection *-rlu*, also yielding *yaparlangurlu*. These words are disambiguated by whether stress falls on the second or third syllable:

'yaparla-'ngurlu    'yapa-'rlangu-rlu

A vital prosody for every language is intonation, which generally applies over clause and sentence (see the discussion in §2.3). There is today a tradition among grammar writers (followed by the present author, to his shame) of more or less ignoring intonation, something which will surely be wondered at and bemoaned by future generations. A few remarks may be popped in about rising intonation for questions, but little about how this relates to intonation patterns for non-questions. What is needed, in every grammar, is explicit characterization of the intonation tunes for all clause types.

Any language which is in social contact with other languages will borrow from them (and they from it). Loans are assimilated, in varying degrees, to the phonological system of the borrowing language. Some are fully assimilated, and phonologically indistinguishable from native words. Others are scarcely assimilated, being used like items from a foreign language. In between there are different levels of assimilation. For example, a loan may be adapted to the existing consonants and vowels of its new language, but with non-native phonotactics. Sometimes, a language includes a considerable number of words of significantly different structure from the main body of vocabulary. In essence, these are 'coexisting systems' of consonants and/or vowels and/or phonotactics and/or prosodies. Seminal studies on this topic include Henderson (1951) and Fries and Pike (1949).

One other point to be noted is that interjections often have different segmental and prosodic phonology from the bulk of the language. For example, in English these include the alveolar click sound indicating disapproval (often written *tsk tsk* or *tut tut*) or the sound indicating hesitation or impatience (often written *hmm* or *h'm*), which is a bilabial nasal commencing voiceless and then becoming voiced. For another example, Dyirbal, like almost all Australian languages, has no sibilant phonemes, but sibilants do feature in interjections. There may also be special phonological and phonetic features found in onomatopoeic forms, and in distinctive styles such as the register used to talk to babies.

## 7.7 Balancing parsimony and clarity

In any scientific endeavour, a guiding principle is to employ the minimum number of analytic units such that description and explanation are fully served. However, the principle of parsimony can be taken too far, to the extent

that it impinges on clarity of analysis. The following two discussions exemplify the need to balance parsimony against clarity.

(a) *w* and *u*, *y* and *i*

Once, when in Tonga, I planned to visit the island called 'Eua (here ' represents an initial glottal stop). My tongue got the shivers at the prospect of pronunciation. Three vowels together—a triphthong, or what? In fact, nothing could be more straightforward; the word has a CV.CV structure [ʔe.wa]. It is just that those who devised the Tongan writing system decided they could do without separate letters for *y* and *w* (which in fact only occur medially, not word-initially), making *i* and *u* each do double duty as vowel and as semi-vowel. It is possible to achieve this for Tongan, without ambiguity, but by so doing the major principle of any writing system—that it should be easy to comprehend and to use—is somewhat vitiated.

Making one letter stand for a vowel, which occurs as the nucleus of a syllable, and also for a consonant, which functions as onset and/or coda, scores high on the scale of parsimony. But it impedes clarity. Consider the following from a language which has syllable structure CV(C), never permits a sequence of two vowels, and writes *w* as *u* and *y* as *i*:

- (i) *uabai*. This is unambiguous but it does take a little working out to understand what precisely it is. For each apparent vowel sequence, one of the letters must represent a semi-vowel. There is no semi-vowel corresponding to *a*, and so the word must be (with syllable boundary marked) *wa.bay*.
- (ii) *baiui*. This could represent two quite different structures, *bay.wi* or *ba.yuy*. The only way to show the structure of the word would be to indicate where the syllable boundary is: *bai.ui* for *bay.wi* and *ba.iii* for *ba.yuy*. Application of the parsimony principle has led to there being two fewer consonants, but it is now necessary to mark syllable boundary (in at least some words), a complexity which cancels out some or all of the gain. With respect to clarity, there is simply loss.

(b) [d] and [r] in Nyawaygi

Almost every Australian language has a symmetrical consonant system in that there is a nasal corresponding to each stop consonant. (Just a few have more stops than nasals, never the reverse.) An earlier stage of Nyawaygi had three vowels (as does the modern language) and a canonical set of consonants, set out in Table 7.3. At this stage the functional possibilities of phonemes /d/, /r/, and /ɹ/ were:

TABLE 7.3. Consonant system at an earlier stage of Nyawaygi

|            | apico-alveolar | apico-postalveolar | lamino-palatal | dorso-velar | bilabial |
|------------|----------------|--------------------|----------------|-------------|----------|
| rhotic     | r              | ɹ                  |                |             |          |
| stop       | d              |                    | ʃ              | g           | b        |
| nasal      | n              |                    | ɲ              | ŋ           | m        |
| lateral    | l              |                    |                |             |          |
| semi-vowel |                |                    | y              |             | w        |

|     | word-initial | between vowels | before a<br>consonant | after a<br>consonant | word-final |
|-----|--------------|----------------|-----------------------|----------------------|------------|
| /d/ | ✓            | ✓              |                       | ✓                    |            |
| /r/ |              | ✓              | ✓                     |                      | ✓          |
| /ɹ/ | ✓            | ✓              | ✓                     |                      |            |

Then the following changes took place:

- In initial position: apico-alveolar stop *d* became apico-alveolar rhotic trill *r*
- Between vowels: before *u*, *d* became apico-postalveolar rhotic continuant *ɹ*; before *i* and *a*, *d* again became *r*
- After a consonant (only *-nd-* is attested): *d* remained an apico-alveolar stop

Once these rules applied, the functional occurrences of sounds [d], [r], and [ɹ] were:

|     | word-initial | between vowels | before a<br>consonant | after a<br>consonant | word-final |
|-----|--------------|----------------|-----------------------|----------------------|------------|
| [d] |              |                |                       | ✓                    |            |
| [r] | ✓            | ✓              | ✓                     |                      | ✓          |
| [ɹ] | ✓            | ✓              | ✓                     |                      |            |

Originally, all three sounds contrasted in one environment—between vowels—showing that there were then three phonemes /d/, /r/, and /ɹ/. After the changes just described, [d] occurs only as second element of the consonant cluster *-nd-*; it is in complementary distribution with [r] and [ɹ], neither of which can occur in this environment. Since /d/ is phonetically more similar to [r], the most appropriate analysis is to take them as allophones of a single phoneme, which is pronounced [d] after [n] and as [r] elsewhere. The ‘psychological reality’ of the [r]/[d] phoneme is shown in the phonology of loans. The place name *Cardwell* [ˈkɑːdwəl] is rendered as *Gariwul* in Nyawaygi,

English *d* being interpreted as the [r] allophone—that which occurs between vowels—of the [d]/[r] phoneme.

What letter do we use for this new phoneme, [d]/[r]? If /d/ were chosen, Nyawaygi would retain its symmetrical consonant matrix, with a stop corresponding to every nasal. But this would surely be inappropriate since [r] is plainly the major allophone, occurring in four environments, while [d] is restricted to just one.

It is neither necessary nor appropriate to make a fixed decision on this. One can write the top part of the consonant chart as:

|        | apico-<br>alveolar | apico-<br>postalveolar | lamino-palatal | dorso-velar | bilabial |
|--------|--------------------|------------------------|----------------|-------------|----------|
| rhotic | / ( [r] ) /        | /ɹ/                    |                |             |          |
| stop   | / ( [d] ) /        |                        | /tʃ/           | /g/         | /b/      |
| nasal  | /n/                |                        | /ɲ/            | /ŋ/         | /m/      |

Here the squashed circle drawn around [r] and [d] indicates that they are allophones of a single phoneme. The language does have an apico-alveolar stop, corresponding to the apico-alveolar nasal (so that nasal/stop symmetry is not entirely broken), it is just that it is allophonically linked with the apico-alveolar rhotic. When writing a grammar of Nyawaygi I actually used orthographic letter *d* for the [r]/[d] phoneme after /n/ and letter *r* when it occurs in the other four environments, mirroring their allophony. This was done for clarity.

It can be seen that the [r]/[d] phoneme occurs in all five functional slots, more than any other consonant. At the earlier stage /r/ was in three and /d/ in three slots; they both occurred between vowels, and this is the only functional contrast that has been lost.

## 7.8 Orthography

Standard phonetic symbols should be used in the phonological chapter of a grammar, but elsewhere a practical orthography may be appropriately employed. It is best—as far as possible—to use letters of the Roman alphabet, avoiding (as much as can be) digraphs and diacritics.

When Tongan was first provided with a writing system, letters for most consonant phonemes were straightforward: *f*, *h*, *k*, *l*, *m*, *n*, *p*, *s*, *t*, and *v*. There is also a dorso-velar nasal phoneme /ŋ/. This is written by *ng* in English but it was sensibly decided to avoid a digraph and employ letter *g* for /ŋ/. Missionary work in Tonga then extended to Fiji, with the alphabet in successful use for Tongan being augmented to cover the additional phonemes in Fijian. Letter *c* was used for the voiced apico-dental fricative /ð/. Fijian has three prenasalized

stop phonemes, /<sup>m</sup>b/, /<sup>n</sup>d/, and /<sup>ŋ</sup>g/. Letters *b* and *d* were conveniently used for the first two. For consistency, *g* should really have been used for /<sup>ŋ</sup>g/, but this was already bespoken for /ŋ/. In consequence, *q* was adopted for /<sup>ŋ</sup>g/. This orthography has now been in successful use for much more than a century.

It is sensible to use letters of the Roman alphabet for sounds that are similar—but not necessarily identical—to those conventionally represented by them. The voiceless bilabial fricative / $\Phi$ / is often written as *f*, which in English is used for the voiceless labio-dental fricative. Where there are five vowels, /i/, /e/, /a/, [o]/[u], and /i/, some linguists use *o* for the phoneme whose allophones are [o] and [u], freeing letter *u* to represent /i/.

Many languages can be written entirely using letters of the Roman alphabet (employing a bit of creativity). Others can't be—they simply have too many vowels and/or consonants. Possible recourses are to employ either digraphs or diacritics. Suppose a language has the following three contrasting voiceless stops: lamino-dental / $\underset{d}{d}$ /, apico-alveolar /*d*/, and apico-postalveolar / $\underset{d}{d}$ /, plus corresponding nasals. A word could be written with these phonetic symbols; for instance *ḍanḍan*. Or it could be written with digraphs (using *dh* and *nh* for the lamino-dentals, and *rd* and *rn* for the apico-postalveolars), as *dharnrdanh*. The digraph version is long and fiddly, not easy to use. The diacritic version is fine, save that experience has shown that speakers habitually miss off things like diacritics. Which type of orthography to employ is not an easy matter. Each linguist has to weigh the pros and cons, and make their own decision.

It may be that different orthographies are considered appropriate for writing a grammar, and for producing literacy materials for the speech community. But for every purpose, one overriding principle must be adhered to. All contrastive distinctions in the language—whether involving consonants, vowels, stress, tone, or other prosodies—must be shown. (See also §2.3.)

## Sources and notes

This chapter deals just with spoken languages. Sign languages have their own modes of realization, which are in many ways more complex than the phonologies of spoken languages.

7.1. Three of the clearest and best publications on the phoneme are Jakobson (1962), Swadesh (1934), and Chao (1934). On this and many other aspects of phonology, Trubetzkoy (1939) is without peer; see especially pages 31–65. Pike's *Phonemics* (1947b) remains an invaluable work, although one must beware of taking too literally the idea of analytic 'procedures'.

There are many useful texts providing good descriptions of the articulation of consonants, vowels, tones, and so on; no one stands out as head and shoulders above the rest. Abercrombie (1967) is a fine account of the foundations,



while Ladefoged (1975, and later editions) provides a perceptive introduction. Ladefoged and Maddieson (1996) is an outstanding account, covering every speech sound found across the world's languages (see §2.6).

7.2. There is an excellent account of rhotics in Ladefoged and Maddieson (1996: 215–45), building on Lindau (1985). For an account of the variety of rhotics in Australian languages, see Dixon (2002: 573–81).

Australian languages with two *y*-type semi-vowels are described in Wordick (1982: 10–11) and Rumsey (2000: 40–2), and summarized in Dixon (2002: 552). The pros and cons of analyses I and II are discussed in Dixon (2002: 574–5). Rhotics in Yidiñ are described in Dixon (1977a: 33).

7.3. Overlapping phonemes involving apico-alveolar flap in American English are discussed in Bloch (1941).

7.5. Mandarin Chinese examples from Ladefoged (1975: 228). Other good discussions of tone languages include Pike (1948) and Hyman (1975). Lango from Noonan (1992: 69). Copious further exemplification of prosodies is in Palmer (1970). The Djapu materials are from Morphy (1983: 18); see also Dixon (2002: 616–17).

7.7. As an exercise in parsimony, Kuipers (1960) attempted to reduce the number of vowel phonemes recognized for the North-West Caucasian language Kabardian from three to two to one to none. The first step was plausible—long low vowel *ā* was analysed as *ha* when syllable-initial and as *ah* elsewhere. (This has the added advantage of simplifying the phonotactics, since all syllables now commence with a consonant.) But Halle (1970) shows that Kuipers's attempts to eliminate high vowel *ə* and short low vowel *a* entail the inclusion of new specifications (such as marking the position of stress and then adding a phonetic vowel at that position). Besides this, they would greatly reduce clarity.

7.8. In 1943 the Privy Council of Tonga laid down a number of decisions concerning orthography. One was that phoneme /ŋ/ should henceforth be written by *ng* rather than by *g* (Churchward 1953: viii–ix). However, in Fijian, letter *g* has been retained for /ŋ/.

## Lexicon

A language is made up of two interlocking components—grammar and lexicon. As discussed in §1.11 and §5.2, the grammar deals with closed systems: pronouns, articles, tense, gender, and so on. These are exhaustively listed and fully described within the grammar. The lexicon consists of open classes. The members of a lexical class have the same or similar grammatical properties, but differ in terms of their ranges of referents. The lexicon comprises all the lexemes (alternatively called lexical items) in the open and semi-open classes of the language—nouns, verbs, adjectives, and so on. ‘Vocabulary’ is an alternative name for lexicon; a lexicon arranged in alphabetical order is a ‘dictionary’.

Since grammar and lexicon are so closely linked, the same scholar (or group of scholars) should be equally involved in the two endeavours. This is generally the case for description of out-of-the-way languages. But for the best-known languages a tradition has arisen whereby quite different people take responsibility for the tasks of grammar writing and dictionary compiling, to the detriment of both.

The methodology of grammar writing has improved immeasurably over the past 150 years, but that for dictionary making has virtually stood still. Many compilers of dictionaries know little concerning the principles of descriptive linguistics.

Dictionaries of English and other major languages provide some information about the referential meanings of lexemes, but scarcely indicate basic grammatical characteristics. They don’t tell us that *wish* may take a THAT or a TO complement clause (for example, *I wish that I could go*, and *I wish to go*) but *want* is restricted to a TO complement (one can say *I want to go*, but not \**I want that I should go*). Some information of this sort may be included (in not always very coherent form) in ‘Advanced Learners’ Dictionaries’ (why not in all dictionaries?). But even these pay no attention to quite critical bits of grammatical information. The entries for *church* and *school* do not mention that the definite article may be omitted from before these nouns—one can say *I’m going to church/school* but only, for example, *I’m going to the factory* (not \**I’m going to factory*). The entry for *know* does not mention that, in Standard

English, this verb may not be used in ‘imperfective aspect’—one cannot say \**She is knowing that information*.

Many languages are like English in having a number of complex lexemes. ‘Phrasal verbs’, such as *take after X*, *take X up on Y*, and *take X over*, are entirely distinct from the simple lexeme *take*. Yet dictionaries list these under *take*. Attention is paid to the orthographic word, not to the actual lexemes themselves.

Some languages have many homonyms—lexemes with the same form but different meaning—while others have very few. This is not an entirely arbitrary variation; it often depends, in part, on the number of distinct word forms which the phonology of the language permits. Dyirbal has thirteen consonants and three vowels, in word structures CV(C)(C)CV(C)—giving almost 10,000 disyllabic word forms. This language has very few homonyms. Jarawara has eleven consonants and four vowels, in structures (C)V(C)V—yielding only about 1,500 disyllabic word forms. It has many homonyms. There certainly is a tendency for there to be more homonyms in languages with fewer possibilities for distinct word forms. But there may be alternative reasons for multiple homonymy. If a certain phonological contrast is lost (say, that between voiced and voiceless stops), then words which had been differentiated just by this contrast will come to have the same form. Similarly, if final vowels are lost from polysyllabic forms, words which were distinguished only by this segment will become homonyms.

Where there are many lexical homonyms, the language often has grammatical means for distinguishing them. For instance, gender. In Jarawara, there are two lexemes *fowa*, one meaning ‘manioc’ and the other ‘mortar (for grinding)’. *Fowa* ‘manioc’ takes masculine agreement on the predicate it occurs with, while *fowa* ‘mortar’ selects feminine agreement, ensuring that the two lexemes are unlikely to be confused. In Russian there are homonymous verbs *ostrít’* ‘sharpen’ and *ostrít’* ‘make a joke’. They are distinguished grammatically, in that *ostrít’* ‘sharpen’ is transitive, taking an object NP in accusative case, while *ostrít’* ‘make a joke’ is intransitive. Tariana, from north-west Amazonia, has homonyms *-wa* ‘play a flute’, *-wa* ‘try to do’, and *-wa* ‘enter the forest’. We find that *-wa* ‘enter the forest’ only occurs in a serial verb construction with directional meaning (and may then be the first element in the construction); *-wa* ‘try to do’ may occur in a modal serial verb construction (in which it is second element). In contrast, *-wa* ‘play the flute’ may not be used in any kind of serial verb construction.

A certain concept may be coded within the grammar of one language, but dealt with lexically in another. This was illustrated in §1.11 for ‘try’, ‘start’, ‘make’, ‘want’, among others. In some languages, same-subject ‘want’ is coded

within the grammar, while different-subject ‘want’ involves the lexicon. The sentiment ‘We want to go’ may just involve a desiderative suffix attached to the verb ‘go’, whereas ‘We want them to go’ may have to be said as, literally, ‘“They should go”, we said.’ This is discussed further in §18.5.2.

## 8.1 Organizing the lexicon

The anthropologist Bronisław Malinowski (1935: 17) maintains that ‘it is impossible to define a word by mere equation’. His view is that the meaning of a lexeme can only be shown by ‘placing it within its context of culture, by putting it within the set of kindred and cognate expressions, by contrasting it with its opposites, by grammatical analysis and above all by a number of well-chosen examples’.

Consider the verb *balgan* in Dyirbal. A single-word translation into English is ‘hit’. But *balgan* means much more than this, as can be seen by placing it within the set of ‘kindred expressions’:

- *balgan* ‘hit with a long rigid implement, held in the hand’
- *minban* ‘hit with a long rigid implement, that is thrown’; used also for ‘lightning strikes’, and now also for ‘shoot with gun’
- *bunjun* ‘hit with a long flexible implement’ (this is necessarily held in the hand since it would not be effective when thrown); e.g. ‘whip’, ‘spank’
- *bijin* ‘hit with a rounded implement (such as a stone), either held in the hand or thrown’; including ‘punch with the fist’, and ‘heavy rain pelts down on a person’

By examining these four verbs, it can be seen that the major parameter involves the type of implement used: ‘long rigid’, ‘long flexible’, or ‘rounded’. And just for the ‘long rigid’ instrument, there is a further choice between ‘held in the hand’ and ‘let go of, i.e. thrown’ (this parameter recurs at other places in analysis of the lexicon of Dyirbal; see §6.5, a mention later in this section, and Dixon 1973: 457; 1982: 96).

It is convenient to arrange a lexicon in alphabetical order, so that any required lexemic form can easily be found. But it is a serious error to actually *compile* it initial letter by initial letter. Sometimes a team is employed, dividing up the alphabet between them. If this technique were applied for Dyirbal, a quite different lexicographer might be assigned to letter *m* (including *minban*) from the person assigned to letter *b* (with *balgan*, *bunjun*, and *bijin*). The semantic contrast between the four verbs of hitting might not then be brought out.

The meaning of one word begins where the meanings of other words finish. The only sensible way to appreciate the range of reference of colour adjectives

such as *orange*, *scarlet*, *red*, and *vermilion* is by contrasting them physically (using objects of various hues, or else a colour chart). Similarly for verbs such as *explain*, *explicate*, *elucidate*, *expound*, *interpret* (and note their antonym, *confuse*), and for nouns such as *gorilla*, *chimpanzee*, *orang-utan*, and *gibbon*.

Almost all dictionaries of major languages are compiled alphabetically, with each word considered on its own, not as part of a semantic set. This inevitably leads to many vicious circles—lexeme A is defined in terms of B, B in terms of C, C in terms of D, and D in terms of A, completing the circle and leaving the reader no wiser. (This is a generous scenario—many vicious circles in dictionaries have a cycle of two, A defined in terms of B and B in terms of A.)

Suppose you did not know the meaning of *consequence* and wished to ascertain this by looking it up in a recent edition of a 1,600-plus page Oxford English dictionary:

- *consequence*—the result or effect of an action or condition

Now follow through on the critical lexemes used in this definition:

- *result*—consequence, issue, or outcome of something
- *effect*—the result or consequence of an action, etc.

*Effect* takes us back to *result* and *consequence*. What about *outcome* and *issue* in the definition of *result*?

- *outcome*—result or visible effect

This again spirals back into the vicious circle.

- *issue*—a result, an outcome, a decision

*Result* and *outcome* we've had before. But *decision* is new. Suppose that the looker-up knows this word (but doesn't know *result*, *effect*, *outcome*, or *issue*). So, *consequence* is 'result of an action or condition', *result* is 'issue of something', and *issue* is 'a decision'. The learner infers that *consequence* means 'a decision'. This is, in truth, completely misleading.

The reader can amuse themselves with looking up any word, following through on the words employed in its definition, and so forth. Vicious circles will be found to abound.

Dictionaries copy each other's entries; that is freely admitted. And almost all modern dictionaries are organized on the same principles, principles which have not changed over hundreds of years, and are absolutely unsatisfactory.

A lexicon or dictionary will only succeed if there is a directionality to its definitions—a word with a more specific meaning (which is likely to be relatively uncommon in occurrence) should be defined in terms of words of more general meaning (which are likely to be of common occurrence).

Different languages provide different insights into the general nature of language as a human characteristic. It will be instructive to briefly outline an unusual insight gained from study of Jalnguy, the special avoidance language style of the Dyirbal language community. As mentioned in §5.1, Jalnguy has the same phonology and grammar as the everyday language style, Guwal, but every lexeme (save for the four grandparent terms) differs. There are fewer lexemes than in Guwal; Jalnguy has the minimum number of lexemes consistent with it being possible to say in Jalnguy everything which can be said in Guwal. A many-to-one correspondence holds between the Guwal and Jalnguy lexicons. For example, Guwal has separate terms for each species of kangaroo and wallaby; these include *yuri* ‘eastern grey kangaroo (*Macropus giganteus*)’, *barrgan* ‘agile wallaby (*Macropus agilis*)’, *jabali* ‘whiptail kangaroo (*Macropus parryi*)’, and *mabi* ‘tree-climbing kangaroo (*Dendrolagus lumholtzi*)’. Jalnguy has one lexeme, *jajjanbarra*, corresponding to the four Guwal terms *yuri*, *barrgan*, *jabali*, and *mabi*. If further specification were required in Jalnguy, a modifier or relative clause could be added.

Examination of the correspondences between verbal lexemes in Guwal and Jalnguy reveals a distinction between a small set of ‘nuclear verbs’, with general meanings, and a larger set of ‘non-nuclear verbs’, with more specific meanings. Non-nuclear verbs may be defined in terms of the nuclear items (not the reverse). It will be useful to briefly describe how—by following a systematic fieldwork plan—this distinction came to light.

After a year’s fieldwork, I had a corpus of 500 or 600 Guwal verbs and went through them one by one, putting each in a simple sentence, and asking what the Jalnguy correspondent would be. A card was made out for each Jalnguy verb. The card for *wuyuban* showed that it had been given as Jalnguy equivalent for four Guwal verbs:

*wuyuban*

|                  |                                                                     |
|------------------|---------------------------------------------------------------------|
| <i>buwañu</i>    | tell                                                                |
| <i>jingañu</i>   | tell a particular piece of news                                     |
| <i>gindimban</i> | warn                                                                |
| <i>ɲarran</i>    | tell someone one hasn’t a certain thing (e.g. food)<br>when one has |

The second stage of this enterprise was to ask things the other way around. The consultant was asked what the Guwal correspondent would be for Jalnguy verb *wuyuban*. There were a number of possibilities open to her. She could have said ‘*buwañu*, *jingañu*, *gindimban*, *ɲarran*’, listing the verbs for which she had given *wuyuban* in the Guwal-to-Jalnguy stage, or she could have just

mentioned *one* of the Guwal verbs. In fact her response was of the second type; she simply gave *buwañu* as the equivalent of *wuyuban*. I then followed the same procedure with the other main consultant, and received exactly the same response—again, just *buwañu* was given as the Guwal correspondent of *wuyuban*. And the *same thing* happened for *every one* of the Jalnguy verbs. Although from two to twenty verbs were listed on each card—those Guwal items for which the Jalnguy verb had been given as correspondent—each consultant gave *just one* of these as Guwal correspondent in the second stage, and in each case the two consultants picked *the same item*.

It became clear what is happening here. Each card includes one nuclear Guwal verb and a number of non-nuclear verbs. When the nuclear Jalnguy verb was put to a consultant, they *always* chose the nuclear Guwal verb as correspondent, *never* one of the non-nuclear items. These results provide not only justification for the nuclear/non-nuclear distinction, but also a procedure for checking which of the everyday-style verbs are nuclear.

I next checked with the consultant that Guwal verbs *buwañu* ‘tell’ and *jingañu* ‘tell a particular piece of news’ do differ in meaning, despite the fact that they had both been rendered by the Jalnguy verb *wuyuban*. The consultant was then asked how the difference in meaning could be expressed in Jalnguy, if it were necessary to do so. She replied that *buwañu* would just be rendered by *wuyuban* but that *jingañu* could be expressed by *wuyuwuyuban*, with the verb reduplicated. Verbal reduplication in Dyrbal means ‘do it to excess’; thus *wuyuwuyuban* perfectly conveys the meaning of *jingañu*—calling everyone together to listen as one *rather deliberately tells* some particular story or news item. Similarly, when confronted by *buwañu* and *gindimban* ‘warn’, the consultant said that for *buwañu* the Jalnguy translation would be just *wuyuban*, but that for *gindimban* could be *ñungulmban wuyuban*, involving a transitively verbalized form of the Jalnguy number adjective *ñungul* ‘one’; *ñungulmban wuyuban* is literally ‘tell once’, an adequate definition of *gindimban* within the context of Dyrbal culture. In the case of the pair *buwañu* and *ɲarran* ‘tell someone one hasn’t a certain thing when one has’, the consultant again gave just *wuyuban* for *buwañu* but volunteered *wuyuban jilbuɲga* for *ɲarran*. *Jilbu* means ‘nothing’ and *-ɲga* is the locative inflection; *wuyuban jilbuɲga* is literally ‘tell concerning (i.e. that there is) nothing’. At this stage the field notebook read:

*wuyuban*

|   |                  |                           |
|---|------------------|---------------------------|
| N | <i>buwañu</i>    | <i>wuyuban</i>            |
|   | <i>jingañu</i>   | <i>wuyuwuyuban</i>        |
|   | <i>gindimban</i> | <i>ñungulmban wuyuban</i> |
|   | <i>ɲarran</i>    | <i>wuyuban jilbuɲga</i>   |

Effectively, the non-nuclear verbs *jingañu*, *gindimban*, and *ɲarran* had been defined in terms of nuclear *buwañu/wuyuban*. Nothing had been added to the nuclear verb in distinguishing between Guwal pairs in Jalnguy; *buwañu* had, each time, been left simply with correspondent *wuyuban*. The same results were obtained for all the cards for both the main consultants. The everyday style nuclear verb was always left with just the Jalnguy verb as correspondent, and ‘definitions’ were given for non-nuclear verbs. The ‘definitions’ can be seen to be of different syntactic types. Consultants sometimes gave identical definitions for a non-nuclear verb but often rather different ones. However, they agreed in *always* giving some definition for a non-nuclear verb, and *never* attempting one for a nuclear verb. (A summary of the types of definitions employed is in Dixon 1973: 458–68, 1982: 96–112.)

The parameter ‘hold on to’ versus ‘let go of’ is pervasive in the verbal lexicon of Dyrirbal. The verb pair *balgan* ‘hit with a long rigid implement, held in the hand’ and *minban* ‘hit with a long rigid implement, that is thrown’ has been mentioned. This parameter also applies to verbs of spearing, quoted in §6.5. And to the nuclear verbs:

GUWAL JALNGUY

*madan nayɲun* ‘set in motion in a trajectory, letting go of (throw)’

*baygun bubaman* ‘set in motion in a trajectory, holding on to (shake, wave, or bash something on something else)’

Arguments on matters linguistic can be highly revealing. One of my Dyrirbal consultants disagreed with his wife on how the Guwal verb *darrbin* ‘shake (say, a blanket) to dislodge something (could be dirt or crumbs) from it’ should be translated into Jalnguy. He said *bubaman* while she preferred *nayɲun*. And they presented reasons for their choices. ‘It must be *bubaman*, because he keeps hold of the blanket.’ ‘No—*nayɲun*, because the crumbs fly off it.’ Both speakers identified *darrbin* in terms of ‘set in motion in a trajectory’, but they interpreted the action involved from different points of view, in applying the ‘let go of/hold on to’ criterion.

In compiling and analysing the lexicon for any language, words which describe general concepts should first be identified, similar to nuclear verbs in Dyrirbal. Other lexemes may then be defined in terms of them. The basic or nuclear lexemes are not open to definition, but can be explained. For example, *say* should be recognized as a nuclear verb for English, with *declare*, *assert*, *state*, *affirm*, *announce*, *proclaim*, *mention*, *remark*, and others being defined in terms of it, the definitions framed within the language. No attempt should be made to define *say*; its meaning can be shown by a combination of means—description, demonstration, translation.



## 8.2 Constructing a lexicon

The most useful lexical technique, when working on description of a language, is a thesaurus, with the lexemes grouped into semantic types. This should be augmented by an alphabetical list, which need only provide a short gloss for each lexeme, and reference to where it occurs in the thesaurus. It is within the thesaurus that detailed information on meaning and usage is given. All the terms for colours will be in one place, adjectives of value ('good', 'bad', 'nice', 'lovely', 'attractive', 'pleasing', 'atrocious', 'terrible', and so on) in another place, verbs of burning in one spot and verbs of tying in another, nouns referring to parts of trees in one location, nouns referring to utensils and weapons in a further location, and so on.

It is convenient to have a scheme of thesaurus organization ready in advance. One such scheme began in 1979 in the first volume of the *Handbook of Australian languages* and has since been quite widely adopted. I have used it as the basis for work on various Australian languages, and also on Fijian, and on Jarawara in Brazil. Working within a consistent scheme means that I know immediately where to place each lexeme, and can retrieve anything I need in an instant. The outline headings are:

|                        |                                       |
|------------------------|---------------------------------------|
| NOUNS                  | N Flora                               |
| A Body parts           | O ADJECTIVES                          |
| B Human classification | VERBS                                 |
| C Kinship              | P Motion                              |
| D Mammals              | Q Rest                                |
| E Reptiles             | R Induced position (including giving) |
| F Birds                | S Affect                              |
| G Fishes               | T Attention and mental                |
| H Insects, etc.        | U Speaking and thinking               |
| I Language, ceremony   | V Corporeal                           |
| J Artefacts, etc.      | W ADVERBIAL                           |
| K Fire, food, water    | X LOCATION                            |
| L Celestial, weather   | Y TIME                                |
| M Geography, dwellings | Z INTERJECTIONS                       |

Fieldwork techniques vary, depending on the social nature of the language situation and, to some extent, on the predilections of the fieldworker. But certain principles are a matter of common sense. If there is no existing lexicon, this should be built up by semantic sets. Good results are obtained by—at the end of a session—discussing with the consultant what should be on the next day's agenda. 'That story you mentioned about the two brothers who gave people their foods—could you perhaps record that one tomorrow?' 'Yes, I'll go over it in my head tonight, make sure all the bits are in the right places.'

Similarly for the names of things. A good consultant will appreciate the linguist following a systematic approach. ‘Let’s write down all the names of birds tomorrow.’ ‘Good, I’ll think them through before I go to sleep.’ And, when tomorrow comes, work steadily through the different varieties of birds—pelican, storks, ducks, hawks, pigeons and doves, and so on. Don’t hurry the consultant. ‘Any other birds like that one?’ You may record several dozen names in a session. Then, a day or two later, the consultant could volunteer a few more bird names which were overlooked before.

In pre-computer days I’d fill in a 5 inch by 3 inch (12.5 cm by 7.5 cm) card for each lexeme. I continue with the same procedure today, as do many other linguists. Electricity supply is non-existent in many field locations, or else it is intermittent and unreliable. Reliance on a computer leads to frustration and a feeling of impotence when it ceases to work properly. This is why many linguists, working in difficult field situations, prefer to leave their computer back at the university.

Every bird or animal does something. ‘That little lizard, you walk up to the tree he’s on, he go round other side, hide from you.’ Write that down (in the language) on the card. When eliciting names of birds with Dick Moses, teacher nonpareil of the Australian language Yidiñ, we came to pigeons. ‘*Gambi:n*, he come from the west, topknot pigeon they call him in the English’ (This is *Lopholaimus antarcticus*.) ‘What does he do?’ Moses gave the following sentence (to which I have added hyphens, brackets, subscripts, glosses, and translation):

|                          |                                   |                    |       |
|--------------------------|-----------------------------------|--------------------|-------|
| Gambinu-ŋgu <sub>A</sub> | [mayi burrmbur gamu] <sub>O</sub> | buga-ŋ,            | wañja |
| pigeon-ERGATIVE          | fruit palm                        | flower eat-PRESENT | WHEN  |
| nuba                     |                                   |                    |       |
| ripe                     |                                   |                    |       |

The topknot pigeon eats flowers of the palm tree, when they are ripe

Thus, in addition to the bird name, this sentence provides quite a lot of grammatical information. When the noun *gambi:n* takes ergative inflection it assumes the form *gambinu*, shortening the long vowel and adding *-u*. There is an NP consisting of classifier *mayi* ‘non-flesh food, fruit’, noun *burrmbur* ‘palm tree’ (probably the Alexandra palm, *Archontophoenix alexandrae*) and inalienably possessed part noun *gamu* ‘flower, fruit’. In addition, the unstated subject of the verbless (copula-like) clause ‘it is ripe’ is the same as the object argument of the previous transitive clause, indicating that Yidiñ has an S = O syntactic pivot (that is, it is syntactically ergative). All these bits of information are entered in appropriate places within the grammar notebook, coded ‘[gambi:n]’, the square brackets indicating that the information comes from the *gambi:n* card.

I would copy the names onto the left-hand pages of a spiral-bound thesaurus notebook, under ‘F, Birds’, leaving the right-hand pages blank, available for additions at a later date. The advantage of cards is that you can do things with them, maintaining total control over what is happening. Sort alphabetically. Enter an alphabetical list on the left-hand pages of a dictionary notebook, again leaving right-hand pages for later additions. With the lexicon organized in two ways—in the thesaurus by semantic types, and alphabetically—access is instantaneous. Nowadays, similar results can be achieved with a computer.

Some birds can co-occur with classifier *miña* ‘non-plant food’; others are not regarded as edible. This—like every other piece of relevant grammatical or other information—must be noted, for each bird, mammal, reptile, fish, and insect.

One can ask names for body parts (see the discussion in §1.7), fauna, flora, and other concrete things. Some verbs and adjectives may be obtained in this way, but many will only be encountered in texts and through participant observation. My aim in the field is each evening to go through the day’s materials, for new uses of lexemes already in the corpus (marking these on the existing cards), and for new lexemes, each being assigned a card. New cards gradually accumulate, and are sorted into semantic fields. Every couple of weeks I would go over the new items in a certain area. Five new bird names—are any of them dialect variants of previously documented names? And so on.

A new verb *wirryal* comes up during fieldwork on Yidiñ, glossed as ‘pull’. But that page of the thesaurus already has two verbs translated by ‘pull’, *mundal* and *burrgal*. Need to follow Malinowski’s advice, and compare kindred forms:

- *mundal*, pull or drag with a steady motion. For example: pulling a branch down to reach the fruit on it; leading a horse by the reins; pulling a canoe over land; pulling on a vine leading to a yam (also used for a smoker pulling on a tobacco pipe).
- *wirryal*, pull or snap something off where it is attached or held. For example: snatching something out of a person’s hand; breaking a branch off a tree; pulling a hank of hair out of someone’s head; tearing or ripping up a piece of cloth.
- *burrgal*, pull off. This verb refers to a more deliberate and controlled action than *wirryal*. For example: pulling leaves off a tree; picking fruit one by one; plucking a bird.

Yidiñ is a fairly agglutinative language with regular verb derivations and inflections. When I came to work on Jarawara it posed greater problems. Here I reserved an A4 sheet (sometimes several) for each verb and, every evening (or

whenever I could), copied onto it each sentence in which the verb appeared. I am fully aware that if all the data had been typed onto a computer, the machine would have done this for me. But the machine would not have learnt the language for me, let alone analysed it. As I copied out each sentence I was learning it, thinking about its structure and meaning, getting ideas—perhaps a hypothesis concerning grammatical structure that could be checked out on the morrow.

When the time came to publish the vocabulary of Yidiñ (Dixon 1991b), it was produced as a thesaurus, with alphabetical finder lists of Yidiñ forms and of English translations. It is surely more useful to place the three words of pulling all together, plus illustrative sentences, where they can be compared—and that which is most suitable for a particular circumstance can then be chosen—rather than to have them scattered, each a lonely alphabetical entry separated from its congeners. This is the approach I recommend for publishing the lexicon of any language.

People who have not themselves undertaken intensive fieldwork sometimes get ideas about what can or should be done. They may suggest using grammatical questionnaires—a set of sentences in English (or Spanish, or whatever) which they believe can be used to elicit information about the grammatical structure of every language in the world. If this were practicable, it should surely apply the other way round. That is, one should be able to devise questionnaires in Quechua or Igbo or Dyirbal, application of which would enable a linguist to determine the grammatical structure of English or French or Russian. Absolutely impracticable.

Universally applicable word lists are only a little less foolish. The Swadesh lists—in 100-word or 200-word form—have achieved wide circulation and not a little notoriety. They were oriented to the unsound idea that genetic relationships between languages could be discerned by comparing such a list of ‘core vocabulary’. (In fact, the lists include grammatical forms such as ‘I’, ‘this’, ‘at’, and ‘if’; as pointed out in §5.1, they are not lexical lists at all.) The Swadesh lists are set out alphabetically. I’ve heard of people who try to elicit words thus, to the bewilderment of speakers. The word for ‘foot’ is elicited, followed by ‘four’ and then ‘freeze’. ‘But what about “leg”?’ the speaker enquires. ‘You’ve asked for “foot”, surely “leg” is next?’ ‘Not yet’, comes the response, ‘I’ll ask that later, under “L”’. (In fact, some languages have a single lexeme covering both ‘foot’ and ‘leg’, which is a further reason for operating in terms of semantic fields rather than alphabetical order. And see §1.7.)

Once a linguist has worked on a given language for a fair while, and comes to know the lexicon and its cultural milieu, they may want to survey other dialects. A carefully chosen semantically oriented word list can be a useful tool

in this endeavour. Such a list has to be culturally specific, in terms of concepts which the language operates with, not based on words in some totally different language.

For example, a critical item for inclusion in a lexical list for use in Australia (but not for anywhere else in the world) is ‘boomerang’. This is *mayañ* in northern dialects of Dyirbal, *wajal* in central dialects, and *warrgiñ* in the southernmost dialect, Girramay. It is again *warrgiñ* in Warrgamay, Dyirbal’s southerly neighbour, and *wa:ŋal*—with a long vowel—in Nyawaygi, next language to the south.

### 8.3 Structure of the lexicon

There are a number of semantic types which characterize human language. Each individual language has lexemes in its Verb class referring to Motion, Rest, Giving, Affect (‘hit’, ‘burn’, etc.), Corporeal (‘eat’, ‘laugh’), Attention (‘see’, ‘hear’), and Speaking. There are always lexemes in the Noun class for referring to parts of the human body and parts of other things, and for fauna, flora, sun, moon and stars, water and fire, house, and artefacts.

Some lexical concepts vary in one of a number of ways. They may correspond to different word classes in different languages. ‘Rain’ and ‘thunder’ are coded as nouns in one language and as verbs in another, or—as in English—they may relate to both classes. And there are languages with no specific lexeme ‘rain’; one just has to say ‘water falls’.

A kinship term, such as ‘mother’, refers to a person in the way that a noun (such as ‘woman’) does, but it also indicates a transitive relation, ‘X is the mother of Y’. In some languages (see §1.3), kin terms are inalienably possessed; one cannot say just ‘mother’, only ‘my mother’, ‘your mother’, and so on. In a number of languages, at least some kin terms are transitive verbs, taking subject and object, literally ‘X mothers Y’. (This was mentioned in §1.8.)

Many languages have abstract nouns referring to qualities—sometimes nouns derived from adjectives or verbs (such as *sincerity*, *jealousy*, *hatred*, and *admiration* in English), and sometimes underived nouns (*glory*, *joy*). Other languages lack abstract nouns of this type. One just has to employ the appropriate adjective or verb; rather than ‘Mary is full of admiration for the way John sings’, one would say something like ‘Mary really admires the way John sings’. Some languages have another type of abstract noun—‘time’, ‘distance’, ‘size’, ‘colour’, and the like. Others lack these; rather than ‘What time is it?’ one might ask ‘Where is the sun in the sky?’; and instead of ‘What is its size?’, the appropriate question would be ‘Is it large or small?’

There are two kinds of generic noun. One is used when a specific noun cannot, at that moment, be remembered: ‘what’s-it-called?’; this is not likely

to be employed in prepared speech, or in writing. The other is simply a noun ‘thing’, which may be much used as a feature of the speech style of that language. In Jarawara *jama* ‘thing’ can be employed if one does not feel the need to be specific—within the context of situation, listeners can be expected to know what is being talked of. Or it can just be something to add a modifier to; for example ‘night’ is *jama soki*, literally ‘thing black’. In Jarawara, some foodstuffs are accorded masculine gender (manioc, maize, cashew fruit) and some are feminine (potato, pineapple, banana). One day I took two Jarawara to a restaurant and asked the gender for new types of food they had never before seen. ‘All we can do is call them *jama*, since we have no other names for them.’ And *jama* ‘thing’ is always feminine. In contrast, there are languages which lack anything resembling a generic noun ‘thing’.

Numbers are particularly fascinating. In some languages they fall into the verb class; in some they are adjectives and in others nouns. ‘In Semitic, the cardinal numbers for “one” and “two” are adjectives; those from “three” to “ten” are abstract nouns’ (Gray 1934: 68). In a number of languages it is appropriate to recognize ‘numerals’ as constituting a distinct word class (also including quantifiers such as ‘many’, ‘few’, ‘some of’, and ‘how many’); see §4.4.

Some concepts are universal—they are coded either by a lexeme or by a grammatical element in every language. But each language has, in addition, its own rather specific lexemes, which relate to beliefs, cultural practices, and social organization. Many speech communities have religious beliefs which include the concept of ‘praying’, making humble supplication to a god or gods, and there will be one or more lexemes expressing this. Other communities lack any notion of praying.

In Australia, a girl may be betrothed at a very young age to a slightly older boy. He can claim his wife once she attains puberty and he has proved himself as hunter and provider. During the intervening years, the girl may look with yearning at her future husband as he grows in strength and authority, and the boy likewise, as his wife-to-be matures into a comely maiden. Appropriately, there is in Dyrbal a transitive verb *ɲilbi-l* ‘look longingly at promised wife or husband, before actual marriage’.

The next three subsections discuss recurrent characteristics of noun, adjective, and verb lexemes. The term ‘adverb’ is commonly employed in grammar writing, but it is used in widely different ways. Typically, an adverb modifies a verb in similar fashion to an adjective modifying a noun. And it is not uncommon for at least some adverbs to be morphologically derived from adjectives. For example, English *serious-ly*, *quick-ly*, and Fijian adverbs *va’a-biibii* ‘seriously’, *va’a-totolo* ‘quickly’ derived from adjectives *biibii* ‘heavy, serious’ and *totolo* ‘quick’. But the label ‘adverb’ may be applied, by a variety of linguists, to sentence modifiers such as ‘not’ and ‘maybe’, and to much more besides.

All languages have some lexemes referring to location—‘far’, ‘outside’—and to time—‘yesterday’, ‘later’, etc. (see p. 121 above). These sometimes pattern as members of the noun or verb class; other times it is appropriate to recognize a separate class of location and/or of time words.

Many languages across the world—including, most notably, African languages—have a set of lexemes called ‘ideophones’ (or ‘expressives’ or ‘mimetics’). These typically have rather different phonology from the rest of the lexicon; they may involve inherent reduplication and are often onomatopoeic, relating to ‘manner, colour, sound, smell, action, state or intensity’ (Doke 1935: 118). In some languages ideophones are treated by linguists as a subclass of adverbials, in others they appear to be distributed over several word classes. The papers in Voeltz and Kilian-Hatz (2001) nicely portray the nature of lexemes called ideophones across a variety of languages. (See also Kulemeko 1996.)

### 8.3.1 Nouns

As noted in §1.11, every language has a large open class of nouns, with—at the least—several thousand members. There may be a number of grammatical parameters applying for the word class; *every* noun has to be specified for *each* of these. They may include gender or noun class, or the range of classifiers which a given noun may take. And there may be limited number specification. For example, in English most nouns may take plural suffix *-s* (these are ‘count nouns’). Some—such as *mud* and *knowledge*—do not, and should be identified in the lexicon as ‘mass nouns’.

There are typically a number of generic terms, although their scope can vary both within and between languages. For example, in English *animal* can be used for (a) any living organism of the Kingdom Animalia (as opposed to Plantia), or (b) any of these excluding humans; or (c) just any mammal (members of the Class Mammalia), or (d) any mammal other than a human.

A generic term may have fixed reference within a language, but might not correspond to a similar generic term in another language. For example, in Dyirbal *wadam* covers all snakes except pythons, and *jabu* refers to all fishes except stone fish and toad fish, two fishes which are harmful to people (see §1.9). In Jarawara, *awa* covers all trees except for pines (which are non-flowering).

For nouns referring to individual types of fauna and flora, a fieldworker must always attempt to provide as specific a description as possible. Ideally, zoological and botanical genus and species should be given. For these to be stated accurately, the linguist needs to work with a biologist. This is a long-term goal, not always easy to achieve. In the short term, speakers may know the name for a species in the lingua franca, which is a fair beginning. What is

totally unsatisfactory is for ‘type of monkey’ to be given for each of a dozen distinctive lexemes. Or ‘type of basket’. It is not a hard task to describe each named variety of basket (and maybe provide a drawing or photograph of it).

In many languages, there is a subclass of nouns for which a possessor must be stated. For instance, in Fijian *ulu-* ‘head’ requires a possessive suffix; one must say *ulu-qu* ‘my head’, or *ulu-mu* ‘your head’ or *ulu-na* ‘his/her head’, not just \**ulu*. This property should be marked in the lexicon, here achieved by including a hyphen at the end of the root, *ulu-*. Some nouns may be used either with or without a possessive suffix, there then being a difference of meaning. For instance, obligatorily possessed noun *yaca-* is ‘name’ but free noun *yaca* is ‘namesake’; *yaca-qu* is ‘my name’, with possessive suffix *-qu* ‘my’, whereas *qou yaca* is ‘my namesake’, with free possessive pronoun *qou* ‘my’.

Obligatorily possessed nouns typically cover body parts of humans and animals, also parts of trees or of artefacts. They may, in some languages, extend to concepts such as ‘name’, ‘smell’, ‘companion’. Body-part possessed nouns often have a secondary sense concerned with orientation. In Fijian (as in many Austronesian languages), *mata-* is ‘face’ and also ‘front of (any object)’.

As mentioned in §1.3, possessed nouns may constitute just a part of a semantic set. Kin terms referring to blood relations—such as ‘mother’ and ‘son’—may be inalienably possessed, but not those referring to relatives by marriage—‘husband’, ‘mother-in-law’, and so on. In Fijian, what we can call ‘primary body parts’—including *ulu-* ‘head’, *mata-* ‘face’, *liga-* ‘arm, hand’, *yava-* ‘leg, foot’—must take a possessor suffix, but this does not apply to ‘secondary body parts’. *Drau* ‘hair’ (also used for leaf of a tree and page of a book) falls into this set. For ‘my head hair’ one must say *drau ni ulu-qu*, associating the secondary body-part term *drau* with the primary body-part noun *ulu-*, which takes possessive suffix *-qu* ‘my’ (rather than saying \**drau-qu*). There is fairly detailed discussion of types of possessed items in §16.5.

The degree of specificity in a given semantic area varies. For example, Nyawaygi, an Australian language, does not have lexemes meaning just ‘moon’ or ‘sun’. One must specify either *ɲilgan* ‘full moon’ or *balanu* ‘new moon’, and either *jula* ‘hot sun (as at midday)’ or *bujira* ‘sun which is less hot (as in the early morning)’. Other languages deal with these concepts in a quite different way. Proto-Arawak did have lexemes ‘sun’ and ‘moon’, but in its modern descendant Tariana the original word for ‘moon’, *ke:ri*, has been extended to also cover ‘sun’. If there is need to distinguish these, one would say ‘*ke:ri* of the day’ for ‘sun’ and ‘*ke:ri* of the night’ for ‘moon’. (The original ‘sun’ lexeme, *kamu*, has shifted its meaning to ‘hot season’.)

The set of terms referring to kinship relations deserves particular study. There may be special grammatical processes for indicating kinship pairs—for example ‘grandfather and grandchild’ or ‘mother and son’. Some languages



have kin terms which relate to three individuals—how X refers to Y when talking to Z. For instance, if a Dyirbal person talks to their father about their younger brother, the person will refer to the younger brother as *mirray*. When talking to their mother, the younger brother is referred to as *ganajunu*. Talking to a grandparent or a sibling about the younger brother they use *malaJurru*. Talking to anyone else, a younger brother will be referred to as *yabuju*.

### 8.3.2 Adjectives

Adjective classes were discussed in §1.11, in §3.6, and in §4.9, and they are the topic of Chapter 12 (Volume 2). In summary, an adjective class can be large and open (with several hundred members) or small and closed (with from a handful to a hundred or so members). In some languages the adjective class has grammatical properties similar to those of nouns, in some languages it is similar to verbs, in some it is like both, and in a further set of languages it is different from both.

Small adjective classes include lexemes from the semantic types Dimension, Age, Colour, and Value. Slightly larger classes include some Physical Property lexemes. The Human Propensity type enters when an adjective class is reasonably large.

There is a tendency for the description of human propensities to be achieved by metaphorical-type compounds of a body-part term plus a Dimension or Physical Property adjective or a number lexeme. This is observable even in a language with a large open class of adjectives, such as English. For example, there are a number of compounds involving physical property adjectives *heavy* and *light*, *hard* and *soft*, plus body-part nouns *head*, *heart*, and *hand*. (Only very rough glosses are provided here.)

|                      |                            |                      |                               |
|----------------------|----------------------------|----------------------|-------------------------------|
| <i>heavy-headed</i>  | ‘feeling dull or drowsy’   | <i>light-headed</i>  | ‘dizzy, unsteady’             |
| <i>heavy-hearted</i> | ‘sorrowful, sad’           | <i>light-hearted</i> | ‘cheerful, carefree’          |
| <i>heavy-handed</i>  | ‘harsh, oppressive’        | <i>light-handed</i>  | ‘having a deft touch’         |
| <i>hard-headed</i>   | ‘practical, unsentimental’ | <i>soft-headed</i>   | ‘feeble, stupid’              |
| <i>hard-hearted</i>  | ‘cruel, unfeeling’         | <i>soft-hearted</i>  | ‘overly sympathetic and kind’ |

These compound adjectives in English are in addition to simple adjectives such as *sad*, *stupid*, and *cruel*, and derived forms such as *cheerful*, *carefree*, and *unsentimental*. In Tzotzil, a Mayan language from Mexico, most human propensity concepts can only be expressed by compounds based on ‘heart’ or ‘head’. For example, ‘happy’ is literally ‘one-hearted’, ‘sad’ is ‘two-hearted’, ‘cowardly’ is ‘small-hearted’, ‘offended’ is ‘heart is broken/ruined’, ‘furious’ is ‘head is gnawing’, ‘angry’ is ‘bad head’, and ‘deceitful’ is ‘much head’. (Note

that Dimension and number lexemes also enter into compound adjectives in English, such as *big-headed* and *two-faced*.) Different languages focus on different body parts for the description of emotions—‘heart’, ‘liver’ or ‘stomach’; ‘head’, ‘temple’ or ‘ear’; ‘blood’ or ‘bone’.

As with other word classes, each language is likely to focus on adjectival concepts which relate to its cultural priorities. Dravidian languages from South India, whose speakers are particularly attuned to the flavour of their foods, make much use of the four taste adjectives—‘sweet’, ‘sour’, ‘bitter’, and ‘salty’—and the languages of herding peoples from East Africa have many adjectives referring to plain and mottled colours, used in the identification of cattle.

### 8.3.3 Verbs

There are never as many simple verbs as there are nouns, but most languages have, at the least, between 500 and 1,000 monomorphemic verb lexemes. However, there are languages which involve only a few (a hundred or even less) verbs in the strictest sense of the term. If a language has morphological processes relating to tense, aspect, modality, and the like, these will apply only to members of the verb class. If there are few such verbs, then they are likely to combine with what are often called ‘coverbs’ (which most often are not subject to morphological processes) to form a large number of compound verbs; this was illustrated in §1.11. There was mention under (b) in §3.4 of ‘serial verb constructions’, where the predicate includes two or more verbs, each of which could make up a predicate on its own. Although serial verb constructions are found in languages which have large, open sets of verbs, they are particularly appropriate for forming complex predicates when the number of simple verb lexemes is restricted.

Just as nouns should be specified for gender and the like, so should the transitivity values of verbs be clearly indicated in the lexicon—whether strictly intransitive, extended intransitive, strictly transitive, extended transitive (or ditransitive), or ambitransitive of type  $S = O$  or of type  $S = A$ ; see §3.3 and Chapter 13 (Volume 2). For those verbs which take a complement clause in a core argument slot, the lexicon must specify which complement clause types apply and in which argument slot(s); see §3.10 and Chapter 18 (Volume 2). Any other grammatical properties with restricted occurrence should be stated in the entries of those verbs for which they are applicable. For example, suppletive forms relating to the number of a core argument, mentioned under (5) in §3.17. If most verbs may be passivized, then those which do not occur in passive constructions should be marked thus, in the lexicon. In §8.1, Malinowski was quoted as insisting that an entry in the lexicon should include ‘above all... a number of well-chosen examples’. This is important for nouns and adjectives. It is even more vital for verbs, since example sentences show not only the

range of senses of a verb, but also the kinds of arguments it typically occurs with.

Languages differ as to the generality or specificity of verb lexemes. English has verbs *wash* and *carry*. Some languages lack any general verb corresponding to these. There are, instead, sets of verbs with more specific reference. One must choose between ‘wash face’, ‘wash hands’, ‘wash rest of body’, and ‘wash clothes’; and between ‘carry over the shoulder’, ‘carry on the head’, ‘carry in the hand’, and ‘carry against the stomach’.

Types of eating may be specified in one of a variety of ways. The Girramay dialect of Dyirbal has three rather specific verbs, depending on *the nature of the foodstuff* that is being consumed:

- rubiñu* ‘eat fish’  
*burñjan* ‘eat meat’  
*nanban* ‘eat fruit and vegetables’

Jarawara also has a number of verbs of eating, but these describe *the nature of the activity*, not the type of object involved:

- kaba-* ‘eating where a lot of chewing is involved’ (this is used of meat, fish, maize, yams, manioc, biscuits, etc.)  
*jome -na-* ‘eating where little or no chewing is needed’ (eating an orange or a banana); also used for swallowing a pill  
*komo -na-* ‘eating which involves spitting out seeds’ (for example, *jifo*, a palm fruit)  
*bako -na-* ‘eating by sucking’ (for example, watermelon, cashew fruit)

For some foods there is a choice of verb available; for instance, eating a pineapple could be described by *jome -na-* or by *bako -na-*.

In Warlpiri, from Central Australia, a single verb *ngarni* covers both eating and drinking, while in Manambu, from New Guinea, the verb *kə* is used for eating, drinking, and also smoking—‘consumption of any substance (independently of its consistency) that involves swallowing or going down a person’s throat’. Which type of action is being referred to in any particular instance of use of *ngarni* in Warlpiri or of *kə* in Manambu can be inferred from the nature of the O NP.

Some verbal concepts are found in all languages—‘cut’, ‘tell’, ‘fall’, and many more—but others are not. For example, Dyirbal has a rather specific verb *walgarrañu* ‘want to do something to satisfy a persistent emotional worry or desire’ and also *garrgiñu* ‘want to go to a place’, but it does not have any general verb similar to *want* in English. As mentioned in §1.6, this language also lacks a verb ‘know’ (although there are two adjectives ‘don’t know’—*ñandu* ‘be unfamiliar with (a person or place)’ and *juru* ‘haven’t seen, never been told’).

‘Think’ is sometimes rendered by a simple verb, but often has more complex realization. In Manambu the verb *wukə-* ‘hear, listen’ has secondary meanings ‘understand’ and ‘think about’. In Dyirbal ‘think’ is, literally, ‘listen to oneself’. In Jarawara the way to say ‘think’ is, literally, ‘my inside language talks’; in Telugu it is ‘say to oneself’.

All languages have verbs of speaking which may frame quotations of direct speech, as in English *‘I’m hungry,’ said John*. In some languages the direct speech functions as O argument for a transitive verb of speaking such as ‘say’. In others the direct speech is in apposition to the framing verb and does not function as a syntactic argument for it.

As mentioned in §6.5, all languages have a set of posture verbs, typically ‘sit’, ‘stand’, and ‘lie’ (in some languages the set is extended to include other verbs, such as ‘hang’). Often, one of the posture verbs also has a general sense ‘stay, live at a place’; this is ‘sit’ in some languages – for instance, Dyirbal and Fijian—and ‘lie’ in others—for example, Dyirbal’s northerly neighbour Yidiñ. In the Western Desert language of Australia it is ‘stand’ which has a generic meaning ‘be somewhere’. There are languages in which existence is indicated using a posture verb, with each being earmarked for use with a particular semantic set of nouns—women ‘sit’, men ‘stand’, rivers ‘lie’. And these verbs are particularly liable to be grammaticalized, becoming copulas that simply indicate a relation between copula subject and copula complement arguments (with the original postural meaning having been lost).

Of course, every language has lexemes with specific and rather unusual meanings. One of my favourites is *jugarrba-n* in Yidiñ. This is used to describe being unable to sleep at night when one’s mind isn’t able to relax but keeps moving around. It is also used to describe someone who, in the middle of the day, is unable to think straight, can’t focus their mind on something, and (as Dick Moses explained it to me) ‘has a fizzy head’. The overarching meaning of the verb appears to be ‘have unsettled mind’.

This chapter has provided some hints on how to do fieldwork; the next chapter goes into these matters in greater detail.

## Sources and notes

Examples of how the concept ‘want’ may require different grammatical frames for ‘same subject’ and for ‘different subject’ are in Dixon (2006a: 37; 2009).

8.1. The *Cobuild dictionary* (Sinclair 2001), makes some attempt to define difficult words in terms of simpler ones, and thus has fewer vicious circles than other dictionaries (but it still sports a goodly number).

Wierzbicka (1972, 1980, and many subsequent publications) espouses the idea of always defining words with a specific meaning in terms of words with a more general meaning. She adopts a small set of semantic primitives (the number and identity of which has been many times revised) and attempts to define all other words in terms of these. Unfortunately, this often requires long and involved definitions which are not at all easy to comprehend.

Description here of the field technique employed for Jalnguy is largely repeated from Dixon (1973: 449–51; 1982: 84–7).

8.2. The scheme of thesaurus organization outlined here can be seen in use for each of the grammatical sketches in the five volumes of the *Handbook of Australian languages* (Dixon and Blake 1979–2000).

8.3. An example of a language for which numbers have been recognized as a separate word class is Fijian; see Dixon (1988: 141–5).

8.3.2. Information on Tzotzil from John Haviland (personal communication); repeated here from Dixon (1977c: 53; 1982: 43–4).

8.3.3. Information on Manambu from chapter 21 of Aikhenvald (2008). ‘Think’ in Telugu from Krishnamurti (2003: 13–14).

Jarawara is an example of a language in which direct speech functions as O argument for the verb of speaking which frames the discourse; see Dixon (2004a: 394–5; 2006b: 112–13). Dyirbal is an example of a language in which direct speech is in apposition to the framing verb (and is not an argument of it); see Dixon (1972).

Useful discussion of ‘sit’, ‘stand’, and ‘lie’ is in Newman (2002). See particularly the chapter by Alan Rumsey on Papuan languages (pp. 179–211) and that on the Western Desert language by Cliff Goddard and Jean Harkins (pp. 213–38).

The Yidiñ verb *jugarrba-n* is from Dixon (1991b: 271).

## Field Linguistics

Fieldwork is the most important and most exciting part of linguistics. But there are many misconceptions concerning what it is and how to do it. This is a short introduction, based on my experience of fieldwork in Australia, Fiji, and Brazil (commencing in 1963 and extending until today), providing a personal view of the discipline. (§2.1, ‘Writing a grammar’, could well be reread at the end of this chapter.)

### 9.1 The fundamentals

#### 9.1.1 *What is linguistic fieldwork?*

Going into a community where a language is spoken, collecting data from fluent native speakers, analysing the data, and providing a comprehensive description, consisting of grammar, texts, and lexicon.

#### 9.1.2 *Why do it?*

There are two main reasons for undertaking fieldwork:

- (a) *To learn linguistics.* The only way to learn any discipline is to get out there and do it. One first learns the theoretical concepts that make up the science of linguistics. But one can only fully understand these concepts through using them—through using them to describe a language.

A medical student learns about the human body and about the techniques of modern surgery. They must then perform actual operations in order to become a surgeon. At a later stage they may write a book about the principles of surgery. In similar fashion, a linguist must actually work on describing a language (preferably a language that has not been fully described before) in order to understand the principles of linguistic analysis.

- (b) *Because one enjoys the intellectual excitement of working on a ‘new language’.* A linguist must take pleasure in what they are doing to be able

to do it well and produce worthwhile and exciting results. If they take pleasure in doing fieldwork and writing the resulting grammar then others will draw pleasure from reading the grammar.

### 9.1.3 *Poor reasons for doing fieldwork*

If someone undertakes fieldwork for any reason other than for its own sake, they are doing it for the wrong reason. For example:

- (i) Just now there is a fad interest in ‘endangered languages’. Many people have taken up this cause and there is lots of talk about it—lots of talk but limited action (and a fair amount of that misdirected). It is not apparent that there are significantly more people undertaking fieldwork now than there were before the fad became current. A warning is in order. No one should undertake fieldwork simply because they feel a social responsibility to document some language before it disappears. Only because they want to learn linguistics and love the challenge of working on a language.
- (ii) Many missionaries feel a call to translate parts of the Christian Bible into some new language. Some of these also have reasons (a) and (b) above, and these people generally produce a fairly good grammar and then a good translation. Others don’t really like linguistics and try to avoid writing a grammar. If they do produce any language description it is invariably bad and their translations are generally so poor that native speakers scarcely recognize them as documents in the intended language.
- (iii) To try to test or prove some theoretical point. Such fieldwork is likely to focus just on one part of a language, ignoring the rest. This is unsound methodology since the part exists with respect to the whole and can only be properly understood when considered within the context of the whole. Fieldwork of this kind is also likely to consist almost exclusively of elicitation; as pointed out in §9.5, this is not the way to really understand what a language is like.

A linguist should work in terms of basic linguistic theory, the cumulative theoretical framework which underlies almost all grammar writing and typological generalization, as set out in the book you hold in your hand.

A fieldworker should be a good all-round linguist. Every part of a language description is equally important and each part interrelates with the others. Someone who says (to others, or just to themselves) anything like ‘I’m basically a phonologist’ or ‘I’m primarily a syntactician’ is not likely to produce a good overall description.

## 9.2 Getting started

### 9.2.1 *The ethics of fieldwork*

Of course, one should only go where one is wanted. A fieldworker will only go into a community which welcomes them and the work they are doing.

Two basic types of field situation can be distinguished. (These are polar extremes; there are intermediate cases.)

- (i) *When the language is still spoken, as first language, by everyone in the community.* Some (or most, or all) people will also have a degree of competence in a lingua franca of the region. My experience (and other people's) is that such a community is likely to welcome a stranger who wishes to learn their language and to provide whatever linguistic feedback the community requests (e.g. a practical dictionary, a volume of traditional texts).

Communities of this type are often located off the beaten track (for example, in the jungles of New Guinea or Amazonia) and tend to have minimal contact with the mainstream of 'civilization'.

- (ii) *When the language is only spoken (or perhaps just remembered) by some older people.* A community of this sort tends to be situated within the mainstream of a nation; working in it can be a delicate political exercise. Some of the younger members, who do not speak the language themselves, may resent an outside linguist learning it. People will be aware of past injustices (which have been a contributing factor towards the loss of language and other aspects of traditional culture) and as a consequence may resent the attentions of a member of that ethnic group which has oppressed them in the past.

In such situations the actual speakers of the language are generally eager to work with a linguist who will record their heritage for posterity, but other members of the community may raise objections. Each situation has its own characteristics, and the linguist must (if they can) negotiate an agreement that is acceptable to all parties. The community may wish to monitor everything that the linguist does and to be consulted concerning what the linguist publishes.

Many linguists have worked happily and fruitfully in this type of situation. However, just occasionally, some members of the community may impose such restrictions that productive fieldwork is not possible; the linguist is then best advised to seek to work elsewhere.

### 9.2.2 *How to choose a language*

All languages have special points of interest and all should—in an ideal world—be accorded a comprehensive description. However, there is no way



that every language with, say, less than a thousand speakers can be described before these pass into extinction. Priorities have to be set, from the global perspective of having materials on a representative selection of languages. A higher priority should be given to working on a previously undescribed 'isolate' language, than to a language from a large and relatively homogeneous subgroup where there are already several good grammars for other languages in the subgroup.

Without in any way maligning any language currently spoken in the world today (all of which are important and interesting), one should recognize the following facts:

1. Some languages are more interesting than others, in terms of their typological characteristics. If a language shows a previously unreported value for a certain category (or, better still, a previously unreported category) then it is of particular interest. For example, a language with a seven-term evidentiality system (larger than any so far known), or a language with more types of conditioning factors for an ergativity split than have so far been reported.
2. Some languages are more difficult to learn, to analyse, and to describe than others. For some languages a comprehensive grammar can be achieved in 400 pages; for others a grammar of similar depth may require 700 pages.

How difficult a language is to learn does, of course, depend in part on the languages already known to the learner. Someone who speaks Tamil will find it easier to learn Telugu than French, and someone who already speaks French will find it easier to learn Italian than Tamil. Someone who already knows a tone language from Africa or Asia may find it easier to learn a tone language from Mexico than someone who has no experience with tone languages.

In a similar way, a linguist who has already described a highly synthetic language will face less of a challenge in describing a newly discovered highly synthetic language than would a linguist whose previous experience has been in describing analytic languages.

Of course, this is not a reason to avoid a particular language. A good, dedicated linguist can learn and describe any language. It is just that in certain cases a little more effort may be needed than in other cases.

When deciding on a language to study a linguist should take into account: (i) the sort of language that would be of particular interest to them; and (ii) the level of difficulty likely to be involved. Taking these points one at a time.

- (i) *Type of language.* Suppose a linguistics student is faced with a choice between which of two languages they should work on. Suppose that

language A is known (or thought) to have a large vowel inventory (including nasalized vowels), ejective consonants, a basically agglutinative morphology, a rich system of cases, and a complex system of TAM, while language B has just five vowels and twelve consonants, a fusional verb structure with three sets of bound pronouns, but just a two-term tense system. Everything else being equal the linguist will choose that language whose points of complexity interest (or excite) them the most.

- (ii) *Level of difficulty.* Take a specific example. In the Solomon Islands there are two kinds of languages, some belonging to the Oceanic subgroup of Austronesian and some non-Austronesian (the cover term 'Papuan' is used for non-Austronesian languages spoken on and in the vicinity of New Guinea). There are about 500 Oceanic languages and good descriptions have been provided for several score of them. Each language has its own characteristics but there are pervasive similarities between them. Describing a new Oceanic language is like filling in the spaces on an established linguistic grid, and adding a few unexpected spaces for language-specific features. In contrast, the Papuan languages of the Solomons are not known to be related to any already-described languages (not enough is yet known to tell whether the Papuan languages of the Solomons are related to each other). Describing a new Papuan language is like establishing a new linguistic grid, and then filling in the spaces on it.

Suppose you want to work on a language from the Solomon Islands for a PhD dissertation (which has to be completed within three or four years). The safest option would be to work on an Oceanic language (and this would yield an interesting and valuable dissertation). The more challenging option would be to work on a Papuan language.

There is one other comment that should be made at this point. People sometimes remark on how lucky it is that the best linguists seem to choose the most difficult languages (those which really require a brilliant linguist, to do justice to them). Similar comments are sometimes made about anthropologists and communities. For example, people sometimes say how lucky it was that the great anthropologist Evans-Pritchard chose to work on Nuer society, from the Sudan. His analytic skills were able to reveal the manifold complexities of Nuer life. This is in fact little comment on the Nuer, simply on Evans-Pritchard. Such was his excellence that, whatever society he had chosen to study, he would have made it seem fascinating.

Similarly for linguists and languages. I shall avoid citing specific cases but there are many examples of two linguists, A and B, working on two nearby languages, X and Y. Linguist A writes a wonderfully long and detailed grammar

of language X which is much admired and quoted. Linguist B writes a short grammar of language Y, which has few features of interest. People comment on how lucky it is that A, who is one of the top people in the field, chose to work on X, which is a really complex and interesting language, rather than on Y, which is a rather simple and dull language. The truth often is that X and Y are equally complex and interesting, if analysed in the right way. It is just that A is a good linguist and B is a rather poor one. For any language A worked on, this linguist would reveal its complexities and make it interesting. For any language that B worked on that linguist would make it seem simple and dull.

There is a warning here. Some grammars of languages are very short and make the language appear unusually simple and uninteresting. In very many cases this is not a valid comment on the language, but on the lack of training and sophistication of the linguist involved. I have seen a certain language worked on by a poor linguist (of type B) and then—some years later—the same language is worked on by a good linguist (of type A). The second linguist describes and explains all sorts of morphophonological patterns, construction types, and so on that the first linguist never dreamed existed.

### 9.2.3 *Fieldwork locations*

Starting from the fifteenth century, white-skinned ethnic groups (speaking dominant languages) began a series of economic exploitations that by the middle of the twentieth century had taken over most of the globe. The ethnic groups that were displaced have mostly disappeared, together with their languages. This applies, for example, to the eastern seaboard of the USA and Brazil, and the major areas of settlement in southern Australia.

By and large, those languages which are still spoken, and which have not yet been described, are in inaccessible places. The languages and their speakers remain, simply because the territory they live in was not thought worthy of economic development. There are, it is true, some endangered languages, spoken on idyllic islands in the South Seas. But the great majority of languages in need of study are deep in a jungle or high on a mountain range, in a place that is not only difficult to get to but may also be difficult to live in. The largest number of languages in need of study are in New Guinea and Amazonia where there may be extreme heat and humidity, plenty of annoying insects, and lots of diseases (malaria, dengue fever, yellow fever, hepatitis, and so on).

These are not easy places to live in, which is why so few people have worked in them in the past. But, by taking sensible precautions (having the relevant inoculations and pills, and a reserve supply of antibiotics, using a mosquito net, purifying drinking water, and the like) living conditions can be made quite

bearable. And any physical discomfort is more than compensated for by the joyous intellectual exhilaration of linguistic fieldwork on a new and exciting language, as well as by the fascination of living as part of a community so different from one's own.

By far the best place to study a language is in a community in which it is actively spoken, observing how the speakers live and how they use the language. Sometimes this is not practicable; for example, if there is a war going on in that area. It is possible to do good work with native speakers in non-native locations. One should still (see §9.3.3) concentrate on recording and analysing texts. However, in these circumstances participant observation—(a) under §9.3.3 below—will scarcely be available. If a period of fieldwork is undertaken with one or a few speakers away from their home territory, this should if at all possible be backed up with a later period of fieldwork in a native community.

## 9.3 Working in the field

### 9.3.1 *How to do it*

My experience is that it is not a good thing to look upon fieldwork as a business arrangement between linguist and language consultant. People have always worked with me as a gesture of friendship, because they are interested in what I am doing and wish to help me to learn their language and to fully document it.

Tribal communities—which have nearly always had their language denigrated by outsiders—are generally glad to have a linguist come to learn and record it. As long as the linguist shows some aptitude for the task (for example, can pronounce it properly) and learns from what they are told, people will be eager to teach them and to answer their questions. Speakers are almost always glad to record texts and help to build up a comprehensive lexicon. There will be some speakers (not necessarily those who recorded the texts) who will be interested in the lengthy and demanding task of transcribing and translating texts.

Of course, one does provide concrete recompense to the language consultants that one works with, either in money or in kind (whichever they request). This should not be regarded as payment for a job done, but rather as a gift in respect of services rendered (by one member of the community to another). One also provides any other kind of help and advice that is appropriate. For example, early in the 1970s I was able to interest the Australian government in buying a substantial block of land and providing houses for the Dyirbal people (who were at that time scattered, each family living under sufferance on a different white farmer's property).

One should, on a daily basis, provide any help that is requested in the community. For example, writing letters for people, providing medicines, and giving technical assistance in areas in which one has some competence (for instance, repairing or cleaning machines).

Each field situation is different. If a consultant has a regular job but arranges to have time off to work with a linguist, then they should be paid at the same rate that they would have received on their regular job (or perhaps at a slightly higher rate).

One builds up a relationship with each of a small coterie of intelligent, reliable, interested, and willing language consultants. The consultants will get on 'the same wavelength' as the linguist, understanding what the linguist is trying to do and sometimes even anticipating what the linguist is about to ask. This is a priceless intellectual partnership.

A responsible linguist will continue for the rest of their life a relationship with a language community among which they have worked, always being available—as needed—with advice and assistance. An academic may find ways of drawing the attention of government bodies to the needs of indigenous groups which the groups themselves would not be able to essay.

A linguist is accepted in different sorts of ways in different societies. In many places some member of the community will act as mentor and 'adopt' the linguist as a certain type of relative. For my fieldwork in North Queensland, Australia, Bessie Jerry adopted me as her *gaya* 'mother's younger brother' (a relation that a woman can have close comradeship with). In Fiji, Elia Waqa adopted me as his *luve-* 'son'. These are both classificatory kinship systems and once I was assigned a relationship with one person, then I also had a relationship (according to an implicit algorithm that everybody knows) with each other person in the community. In contrast, I have done extensive fieldwork among the Jarawara people of southern Amazonia and they do not place strangers in their classificatory kinship system.

There are advantages and disadvantages to each scenario. If you are assigned a close relationship to a certain set of people in the community you will have obligations towards them as they will have towards you. This can be a help to the linguist, as in Fiji when members of my kin group assisted me in building a house. But being aligned with one group may create difficulties; for instance, members of other groups—who have critical specialized knowledge—may not be keen to work with you. And once you are a part of the kinship system you may automatically be in a taboo relationship with certain people (who might be people you would have wished to have contact with). However, there is seldom any choice involved. If a community has a custom of including outsiders (who they respect and value) in the kinship system, then this is what you should aim for. If they lack this custom then that is the way things are.

A linguist can achieve a degree of acceptance into a community (and this is irrespective of whether or not they are taken into the kinship system—for instance, I have as warm a relationship with the Jarawara as I had in Fiji or Australia) but it is important to realize that it is only that, *a degree of acceptance*. There is no way in which a linguist or anthropologist can be fully assimilated into any community (that is, short of marrying into the community and living exactly the same life as everyone else, cutting off all regular contact with their previous life). A linguist will probably have a different skin colour and type of hair; they may help with some community activities but they will spend a great deal of their time writing. They won't be used to living in that environment and are likely to have to take special medicines, and probably to purify drinking water, in order not to fall seriously ill. They are unlikely to have the stamina to perform tasks that members of the community regard as routine (for instance, climbing a high tree in search of honey). It is a mistake to try too hard to become 'one of the community'. You are different, and will always be perceived as different. You will be respected for what you are, not for what you try to be.

The ideal place to work also varies, depending on local conditions. When I began fieldwork in Australia, in 1963, the pioneer anthropologist A. P. Elkin advised: 'always go to their house, don't make them come to where you are staying.' This was sound advice for fieldwork in state-run or mission-run Aboriginal communities in Australia at that time. Each settlement was strictly divided into an area of housing for the white staff and an area for the Aboriginal people. Elkin was right; Aboriginal people were most uneasy about working in the white guest house where I had to stay; I always worked in their part of the settlement, either squatting under a tree or else (if it rained) in the language consultant's house.

Things are generally quite different outside the confines of such a prescriptive settlement. In Fiji and in Brazil I had my own house in the village. People would drop in to see me all the time and we would work together, or else I would go to their house and we would work there. At a later period in Australia (outside the confines of a government settlement or mission) I would stay for a few nights in a consultant's house or they would stay in mine. The point to bear in mind is that people may have preferences as to where to work, and the linguist must be sensitive to these.

Equipment should be kept to a minimum. Too much flashy machinery may alienate the linguist from members of the community and will make it more difficult to achieve success in immersion fieldwork. And the more machinery one takes into the field (and the more complicated it is) the more there is to go wrong. A good-quality robust recorder (of either the cassette or the mini-disc

type) is essential. In case this might fail, there should be a back-up recorder of the same type.

There is nowadays a fashion to talk of ‘documentation’ which involves use of a video camera. The experience of some—but by no means all—practised fieldworkers is that to introduce a video camera into a fieldwork situation gravely disturbs the chance of establishing a close relationship between linguist and speech community. This ‘documentation’ may severely jeopardize the likelihood of success for a standard linguistic description (grammar, texts, and lexicon).

Similar remarks apply for computers. In addition, most fieldwork situations have no electricity, or else an occasional and unreliable supply. If a linguist takes into the field a reliance on computers, it is highly likely that they will, from time to time, become frustrated, with their productivity being impaired. As mentioned in §8.2, energy which has to be spent on computational matters is far better directed towards learning and analysing the language.

Each fieldwork situation is subtly different and culturally specific. It is generally not sensible to attempt to employ questionnaires which were not constructed with this language community in view. Video clips from another society (and materials like ‘Pear’ and ‘Frog’ stories) are often confusing and are unlikely to produce reliable data in the way that immersion fieldwork will. Such extraneous ‘aids’ should only be employed with great care.

### 9.3.2 *What to get*

The aim of a linguistic fieldworker should be to produce and publish:

- (i) *A comprehensive reference grammar of the language (written in terms of basic linguistic theory)*. Alongside this should go a study of dialect differences and social styles of language use (in rituals, ceremonies, and just in varying aspects of everyday life). The language of songs is a further topic for study; in many instances this poses difficulties, and is better left for a later stage (after the grammar is completed).

Before embarking on fieldwork on a previously undescribed language, one should not have any preconceptions about what features its grammar will include. But once a certain feature is recognized it should be fully investigated, examining all the standard parameters of variation as set out in the appropriate typological literature. Many of the chapters in Shopen (1985, 2007) provide a useful overview of grammatical categories and construction types. Appendix 3 of Aikhenvald (2000: 447–51), ‘Fieldworker’s guide to classifier languages’, discusses the types of questions that need to be dealt with for languages with classifiers, noun classes, or genders. And at the end of Aikhenvald

(2004: 385–90) there is a ‘Fieldworker’s guide: How to gather materials on evidentiality systems’. In similar vein, each chapter in Volume 2 of the present work includes a section ‘What to investigate’.

- (ii) *A series of texts with interlinear glosses and full translation, with notes on the social context of the texts and on points of particular grammatical and social interest.* Ideally, the texts should be of varying types—traditional legends, accounts of historical events, autobiographical reminiscences, instructions on how to garden or fish or hunt game or manufacture artefacts, and so on. Some should be monologues but some should involve more than one person—perhaps a recording of a village meeting or just an everyday conversation. The texts should be from a variety of speakers, from different age groups. Ideally they should be from speakers of both sexes; however, in some communities a linguist would not be encouraged to work with someone of a different sex from them.
- (iii) *A reasonably full lexicon.* As described in §8.2, this is best produced in the form of a thesaurus, by semantic fields. In this way, all of a certain type of animal or plant will be listed together, and their meanings compared. Similarly, all adjectives of colour will be in one place, and all verbs of motion, rather than scattered through an alphabetically arranged lexicon. There should be two alphabetical finder lists (keyed to the thesaurus entries), one on the language under description, and one on the lingua franca used for the description (e.g. English, Spanish).

One should also produce such literacy and other materials as the community requests, for its own use. These may include primers, storybooks of various kinds, and practical dictionaries.

There are some situations in which a language is teetering towards extinction and the work of a linguist can help it revive and continue, at least for a while. But there are many field situations where the language is only spoken by a few old people and is past the stage at which it could be revived. (People may not want to believe this to be the case, but the linguist will be able to see that this is so.) In such a circumstance the linguist should devote some time to producing materials for use in school and so on; these can be an aid to increasing ethnic pride in the cultural past of the community. But in such a situation it may be not the best use of limited resources for the linguist to devote too much time to preparing teaching materials (which, if produced, would never be likely to be fully utilized) at the expense of the main task, which is to produce a full and scientific grammar, plus text collection and lexicon.

There is no directionality to producing a language description. One must—right from the start—be working simultaneously on grammar, texts, and



lexicon. Within the grammar there is again no directionality; the linguist must be working simultaneously on phonology, morphology, and syntax, gradually refining each as the results from one feed into the others.

A warning is in order. It is not advisable to publish on one particular aspect of the structure of a language until one has a thorough understanding of the whole system. You may think that you understand the rules for stress assignment, say, at the end of the first period of fieldwork. But when later on you delve deep into morphological structure, you may uncover new factors (say, concerning different kinds of morphological boundaries) which lead to a radical reassessment and restatement of the stress rules. This could be rather embarrassing if you have already published the stress rules, as you stated them before having achieved a full understanding of the grammar.

### 9.3.3 *What to do*

The most important thing of all is to be fully organized. You need to decide on a standard set of notebooks that you will use (I always use A4 or foolscap size spiral bound books, since one can turn the page back and easily use the book on one's knee, or even when standing). Each notebook should be given a code letter and each page numbered. Whenever you write anything in a field notebook, note the date and location and the name of the consultant(s) who supply the information. When you are drafting a grammar sketch you should refer to where each example comes from—for example, a particular instance of an aversive construction was offered by Makabi at Viidawa on 29 February 2004 (and taken down in field notebook C on page 29).

Throughout all fieldwork a linguist should be working simultaneously on three fronts:

- (a) *Becoming a part of the community* (to the extent that one can—see comments above) *and beginning to speak the language*. This involves one's presence within the community being accepted as a normal thing, so that people come and see you (just for a chat, or to look over what you have in your house, or to ask to borrow or be given something) and you go to see them (for similar reasons). They may invite you to come along to some of their communal activities—it could be a meeting or a party or a fishing expedition or just a trip to a nearby town. During all this you will be trying to learn the language. People will generally be keen to teach you the names of things and tell you how to describe what you are doing and what other people are doing. You will of course begin by communicating with them in the lingua franca of the region (English in Australia, Tok Pisin in Papua New Guinea, Spanish in Peru, and so on) but you will gradually use more and more of the local

language. Just keep working steadily at this; don't try to rush it, and don't get disheartened if your progress seems to you to be slow. Every week you will acquire a little more fluency. The important thing is to listen to what people tell you, to encourage people to correct all your mistakes, and to learn from these.

Some linguists insist that one doesn't have to develop any ability at speaking a language—or at learning to understand it (at least to some degree) as it is used in everyday interaction—in order to write an acceptable grammar. While a linguist may produce a grammar without developing any facility for using the language themselves, the grammar will be greatly improved if they are able to include participant observation among the techniques for data gathering. Franz Boas (1911: 60), one of the pioneers of linguistic fieldwork, insisted that 'a command of the language is an indispensable means of obtaining accurate and thorough knowledge, because much information can be gained by listening to conversations of the natives and by taking part in their daily life, which, to the observer who has no command of the language, will remain entirely inaccessible'.

- (b) *Compiling a lexicon.* One should steadily build up a lexicon of every word encountered in texts and conversation, also employing systematic eliciting for certain semantic fields, as described in §§8.2–3.
- (c) *Recording and analysing texts.* Texts are the lifeblood of linguistic fieldwork. The only way to understand the grammatical structure of a language is to analyse recorded texts in that language (*not* by asking how to translate sentences from the lingua franca). One should start gathering and working on texts right from the beginning—if not in the first week of fieldwork then certainly in the second week. Tell consultants that you'd like them to tell a short story in the language (maybe a short account of who they are, where they were born, and so on) so that you can write it down and in this way learn to understand and speak the language.

The first texts you get are likely to be short (it may be a good thing if they are). You then need to transcribe them, with the help of a consultant, try to divide them up into words and then the words into morphemes, and work out the meaning of each sentence, of each word, and (eventually) of each morpheme. As you get to understand the language a little better—and as consultants come to know you better and to respect the work you are doing—they will be likely to proffer longer and more challenging texts. Some topics may take a while to come forth; for instance, it wasn't until my third field trip among the Jarawara in Brazil that they felt they knew me well

enough to record texts about their traditional religion and the spirit world.

Ideally, the person who tells a story (which may take twenty or forty minutes) and the person who helps you transcribe and analyse the story (which may take ten or twenty hours) would be the same. Often this is not practicable. The best storytellers may be old people who don't have the patience to assist with transcribing. But there should be no lack of younger people who will be willing to work with you on transcribing an older relative's story.

Part of a fieldworker's duties should be to encourage and assist native speakers to write down their own language. It is then possible to get them to transcribe texts on their own, which the linguist can use. I don't do this myself simply because I find every stage of text collection, transcription, and analysis to be really important to me in learning the language and understanding its structure. But other linguists find it useful to pay people to transcribe texts for them.

All texts should be transcribed in the field. In the early stages the linguist should not attempt transcription without a native speaker by their side. As the linguist gets to know the language better they can attempt an initial transcription on their own, but should always then go over it in detail with a consultant. In the case of a language I had been working on for more than thirty years (Dyirbal) I could get a new text 95 per cent right, but there were always a few points that I missed the proper meaning or full significance of, and had to have them pointed out by an expert consultant. One thing one should *never* do is just record texts in the field and try to transcribe them later on, back at base (whether one is working on phonetics, phonology, grammar, discourse, or whatever); this is a sure recipe for an incompetent analysis.

Texts are the most important part of a field linguist's database but they can never be the full story. They must be supplemented by what the linguist hears around them—by what people say and by what they tell the linguist to say (this is 'immersion fieldwork'). There are likely to be some construction types which come up frequently in conversation but are seldom (or never) encountered in the more formal milieu of recorded texts. To rely solely on texts is to miss an important data source (and is almost as bad as not using texts at all, but just using sentence elicitation as the basis for a grammar).

You will note that while I have mentioned lexical elicitation, in compiling a vocabulary, I have not mentioned grammatical elicitation—going through a battery of sentences in the lingua franca and asking for their translation

into the native language as a way of getting the verbal paradigm, or relative clause structure, or whatever. Such elicitation should play *no role whatsoever* in linguistic fieldwork.

At first one learns and listens, records and transcribes texts, and tries to analyse them. Suppose that the verb appears to have many possible suffixes. The thing to do is to see what combinations of suffixes co-occur. Suppose one can get suffix 1 followed by suffix 2, and 3 followed by 2 and 4 followed by 2, but that there are no examples of 1 plus 3 or 1 plus 4 or 3 plus 4. The likely hypothesis is that 1, 3, and 4 make up a grammatical system from which only one member may be chosen, and that the 1/3/4 slot is followed by the 2 slot. The linguist should put forward preliminary generalizations on language-internal grounds; that is, by analysing texts in the language.

At a post-initial stage of fieldwork (say, after a couple of months) one does bring a fourth strand into the work:

- (d) *Using elicitation in the language under study to check generalizations and fill in gaps in paradigms.* When working in Brazil I have almost never given a sentence in Portuguese (which most of the men in the community have some knowledge of) and asked how it would be said in Jarawara. What I do is make up Jarawara sentences (that are generated by the grammatical rules I am positing) and ask if these are bona fide utterances. I will seldom just give a sentence, but first describe a context (using either Portuguese or Jarawara for this) and then give a short dialogue, ending with the sentence that I want to test. Or else I will quote some sentence that I know is alright (because I have heard it in a text or conversation) and ask about variants on it, perhaps changing the verb to a similar one ('cough' in place of 'laugh') or adding or subtracting an affix or a word. Consultants get the idea of what I am trying to do and either confirm that my made-up sentence is correct, or else offer an appropriate correction.

When asking whether a sentence is alright it is not sufficient to have the consultant say 'yes'; you must get the consultant to actually say the sentence. People often say 'yes' to a sentence which is correct in one respect but erroneous in some other way. When you get them to say it themselves the sentence may come out differently from the way you had it. (My experience has been that consultants are generally happy to respond to my 'you say it!', and get used to doing so. If a consultant won't do this, then they are simply not a good person to use for this phase of the work.)

It is important to bear in mind that there is a difference between (a) what people think they ought to say, (b) what they think they do say, and (c) what

they actually say. Sometimes people will judge a certain type of sentence as unacceptable although in fact it crops up often in texts.

One is able to work out many parts of the grammar from what occurs in texts. But there may be a few gaps in a paradigm. If some fairly obscure form (for example, the locative form of 2nd person dual pronoun) has not turned up after you have analysed a good few texts then it is sensible to try to elicit it. You may try to predict what the form is, on the basis of the paradigm already gathered. Or you may give a sentence with locative of the 2nd person plural pronoun and then ask what it should be if there were only two people involved. (Sometimes there really is a gap in a paradigm with no form at all. This is unusual, but it does happen.)

What you should *not do* is try to elicit a paradigm from scratch, going inexorably through the parameters (for example: 1st person, 2nd person, 3rd person; singular, plural; past, present, future). Get what you can from texts, and just use elicitation (not all at once, but a bit here and a bit there) to fill in the gaps. For instance, at one time I needed to check two transitive construction types in Jarawara for four kinds of subject (1st or 2nd person singular, 1st or 2nd person non-singular, 3rd person singular, 3rd person non-singular), four kinds of object (the same four possibilities), and four kinds of predicate (with just mood specified, with just tense specified, with both, and with neither). This gave sixty-four possibilities. Checking through texts I found that I had examples of about 80 per cent of these. I directed elicitation towards the remainder, just one or two each day (some of the missing combinations came up in the new texts I was going through during the two weeks it took to complete the elicitation).

A very small number of linguists just analyse texts and scorn any sort of elicitation, even that solely in the language under study. They publish a grammar with gaps in the paradigms, simply because certain forms didn't happen to occur in their particular sample of texts. Using nothing but texts is almost as bad as not using texts at all. To provide a comprehensive grammar of a language—and this should be the goal of every fieldworker—one should base the study on texts and on participant observation, but this must always be augmented by judicious elicitation in the language, to fill in gaps and also to check generalizations.

I said earlier that I almost never asked for the Jarawara translation of a sentence in Portuguese, to investigate some grammatical point. In fact I did this just once. It seemed to me (and was confirmed by consultants) that a certain Jarawara sentence could mean either 'Who hit John?' or 'John hit who?' (i.e. 'Who did John hit?'). I asked how to distinguish the two happenings, using Portuguese for this. (The answer was that one has to employ a two-clause sentence.)

It is beneficial to follow a pragmatic approach in every aspect of linguistic fieldwork. I generally have available a piece of string, a small stick, a nail, a model canoe, and so on. When discussing the various verbs of pulling I offer the string to a consultant who demonstrates that one lexeme means ‘pulling with steady pressure’, another is ‘pull with a sudden jerk’, a third is ‘pull up, hand over hand’, and so on. It is often easier for people to demonstrate things than to describe them.

When trying to ascertain the meaning of a verb I imitate what I think it might mean. ‘Does the verb mean squat like this?’, and I squat down on my toes, heels in the air, bottom off the ground. ‘No’, I am told, and the consultant squats in a different fashion, to show the type of action that the verb refers to. People who may at first be rather shy soon join in with my acting out simple things, and show me the meaning of a word in the best possible manner, by pragmatic demonstration.

#### 9.3.4 *Other fieldwork situations*

The last subsection presented an account of what to do when a language is still in daily use throughout a community, with the linguist able to live in the community for the period of fieldwork and to become a part of it. This is what has been called ‘immersion fieldwork’.

However, many undescribed languages have moved some way along the path towards extinction. A language may be used in limited circumstances just by a number of elderly people, in a community where the language of daily discourse is a lingua franca (for example, English or Indonesian or Spanish or Swahili). Or it may not be in regular use at all, being just remembered—or partly remembered—by a few old people. Sometimes the rememberers are able to give texts, but there are situations where the last speakers (or semi-speakers) of a language are unable to provide texts, just translate short sentences from the lingua franca.

Most linguistic fieldwork situations in North America and in Australia, for instance, fall short of the ideal described in §9.3.3, where there is a community in which the language is in daily use and where the linguist can live and hear the language spoken around them. There is no longer in existence an active language milieu in which the fieldworker can immerse themselves. In such cases the participant observation strand—(a) in §9.3.3—is scarcely available.

One simply does what one can in the circumstances that prevail, through ‘interview fieldwork’. When the only speakers available are unable to give texts, one has to begin by asking for translation of sentences from the lingua franca. (The linguist should still—in due course—pursue the other kind of elicitation, generating sentences in the language under study and asking whether they are

correct, in order to check grammatical hypotheses.) Such a study will yield only a partial grammar, in comparison with the results of an ideal fieldwork situation. But it will still be a useful contribution to knowledge.

## 9.4 Making sure you have it right

A crucial part of making sure that your data is accurate and correct involves obtaining correction from native speakers. One way of doing this is to say (a) something that you know to be correct, and then (b) something that you believe to be incorrect, according to your understanding of the structure of the language. If you *are* correct, the consultant will accept (a)—and repeat it back to you—but reject (b).

The great phonetician Daniel Jones emphasized the importance of this technique, with respect to the sound component of language (Jones and Plaatje 1916: 12):

When you have doubts about one of your sounds, pronounce the word you are practising making slight deviations from what you believe to be the correct sound. You mispronounce on purpose, and notice whether your native [speaker] is just as well satisfied with your intentional mispronunciation as he is with your attempt at the right sound. If he is just as well satisfied, you may be sure that your attempt at the right sound is still very wide of the mark. If your attempt at the right sound is a really good one, the native [speaker] will certainly prefer it to your intentional mispronunciation.

One can use a variation on this method in working out what phones belong to a single phoneme, and which relate to different phonemes; see the discussion of Dyirbal in §7.1. The same technique applies for morphology and syntax. When the linguist has formulated a hypothesis concerning the structure of complex words, for example, they make up a series of what should be acceptable words (if the hypothesis is correct) for ratification by consultants. The linguist should, in addition, concoct a number of words which should be ungrammatical, according to the hypothesis, and check that these are not acceptable to the consultant. Similarly for complex sentence constructions. And so on.

## 9.5 What not to do

These points were mentioned above, but since they appear in the literature, they need to be emphasized.

- (a) *Transcribe after you leave the field.* U. Canger (1994) states in an encyclopedia entry: ‘before the tape recorder, all texts were dictated and simultaneously subjected to a basic analysis.’ This is simply untrue; the

analysis would of course have to follow some way behind the dictation. The article continues: ‘in the early 1990s the linguist will tape-record much material in the field, which he [*sic*] will transcribe and analyse only after he has returned to his desk.’ One can carry on analysing after one has left the field (as one could before the advent of tape-recorders) but to transcribe without the help of a language consultant is likely to be a foolish exercise.

- (b) *Pursue ‘controlled elicitation’*. In an entry for another encyclopedia, Judith L. Aissen (1992) states: ‘while nothing in generative linguistics excludes text collection, direct elicitation is unavoidable.’ She goes on to say: ‘in practice, most fieldwork uses both highly controlled elicitation and more open-ended dialog, combining these with text collection and, *perhaps* [my italics], participant observation.’ Not only is this bad technique (trying to study just some part of a language, to feed in to some part of a formal theory), it is unlikely to obtain the desired results. I have known linguists try to study relative clauses in language X by asking how to translate English sentences including relative clauses into X. Not a single relative clause was used in the translations obtained. Yet the language does have a rich set of relative clause constructions, which can be observed if one studies texts. In another instance the linguist did get some ‘relative clauses’ by this technique but they were simply calques (literally, word-for-word translations) from the lingua franca, and quite different from the relative clauses that are encountered in texts.

If one simply poses a battery of questions in the lingua franca and asks ‘how do you say this in language X?’ one tends to—at best—learn ways in which this language is similar to the lingua franca. Construction types that are unlike anything in the lingua franca are rather unlikely to come up in ‘highly controlled elicitation’.

Thomas E. Payne, in his book *Describing morphosyntax: A guide for field linguists* (1997), emphasizes the value of gathering texts. Yet on page 370 he suggests: ‘perhaps a rule of thumb would be to begin with 90 percent elicited data, and 10 percent text data, then move gradually to 90 percent text data and 10 percent elicited data some time in the second year.’ By ‘elicited data’ he undoubtedly means elicitation by asking for translation of sentences in the lingua franca, something I would recommend doing only very occasionally and certainly not for 90 per cent of the first year of fieldwork.

Some linguists, who are trying to prove a particular point, get frustrated when a speaker—who is in fact a totally reliable consultant—gives what they consider (from their viewpoint) to be inconsistent data.



They maintain that the speaker is making mistakes. In my experience this is almost never so. What the consultant is saying is correct, and can be seen to be so if analysed in its own terms. But the linguist is trying to project onto the language a theoretical model which is not appropriate for it.

Speakers do of course have different levels of competence, and different kinds of judgement (see Leonard Bloomfield's illuminating comments on Menomini speakers in his classic 1927 paper 'Literate and illiterate speech'). Some people tend to have generous judgements of what one can say, while others are meaner in this regard. The linguist will strike up a rapport with a number of consultants (it should always be several, rather than just one) who get to understand what the linguist is trying to do and will guide them along the path towards a full understanding of the language.

Early in fieldwork one writes down everything one is told. But a little later one realizes that what X says is always reliable, that Y sometimes will muddle in material from another language which he also speaks, and that Z will tend to accept some things that other speakers find dubious. By looking back over old notes, the linguist can identify that something which puzzled them was actually from Z (before the linguist stopped using Z as a source of data) and is not upheld by other speakers (such as X). This emphasizes the value of always noting who said what and when and in what context.

## Appendix 1 Describing the fieldwork situation

The anthropologist 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 the 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.

Similar comments are relevant for linguistic reports of work on little-known languages. First of all, was the data gathered from a speaker who has now settled in Los Angeles (or some other big city), or did the linguist actually undertake fieldwork and go to the community where the language is spoken on a daily basis? How many consultants were used? Was the data gathered mainly by elicitation or by analysis of texts? Where actual fieldwork was involved, the points mentioned by Maybury-Lewis should be addressed.

## Appendix 2 Planning a fieldwork PhD

I have had several dozen students each write a grammar of a previously undescribed language for their PhD dissertation (within the 3- to 3½-year time frame allowed for PhD programmes in Australian universities). The normal schedule is:

- 3–6 months: Preliminary library research and planning for fieldwork (including getting appropriate visas and permissions).
- 9–12 months: First (long) field trip.
- 9–12 months: Back at the university, writing up a first draft of every chapter of the dissertation. This must be completed before:
- 2–3 months: Second (short) field trip—checking generalizations and hypotheses, filling in gaps in paradigms, etc.
- 9–12 months: Back at the university, revising the grammar for submission.

## Appendix 3 Field methods courses

Some good linguistics departments offer their students courses in ‘field methods’. These can be most valuable (whether or not a student plans to go on to undertake work in the field themselves) since a student is then faced with a real language, with all its inconsistencies and untidy ends, rather than the neat ‘doctored’ sets of data they will have been asked to ‘solve’ in courses on phonology, morphology, syntax, and the like.

Just as every linguist has their own way of doing fieldwork, so every linguistics teacher seems to have their own way of teaching field methods classes. Different techniques can be equally effective. Let me here just add some general comments, partly concerning the way in which I conduct such courses myself.

- (a) A field methods course should be taught by a linguist who has done real fieldwork, and published a grammar based on it. It should involve a native speaker of a language (preferably one for which there is no grammar readily available).
- (b) In such a course one cannot really approximate an actual field situation. Quite a lot of the time must be spent on elicitation from the *lingua franca*, something that I seldom do in the field. However, I also get students to record short texts, and to transcribe them with the consultant, so that the joint venture has some textual input (and so that the students learn a little of how to handle texts).
- (c) The teacher should have one lecture at the beginning on the ethics and politics of fieldwork, reinforcing this throughout the course by further comments on what to do in the field (and what life is like in a field situation).
- (d) It is a good thing to choose for the field methods course a language which is not too hard in phonology or morphology. If the language has a tough phonology, too much of the course will be spent trying to figure this out. Ideally, working out the basics of phonology should take no more than a couple of weeks, and the essentials of morphology should then be do-able, so that some time is left at the end of the course for topics in syntax.

- (e) In a field methods course of limited duration (typically somewhere between twelve and twenty weeks) one cannot expect to produce a correct description of any aspect of a language. The aim is to teach the students how to undertake fieldwork, not to produce a finished product.

In 1990 I taught a field methods class on Motuna, a Papuan language—of complex structure—from Bougainville in Papua New Guinea. One of the students in the class, Masayuki Onishi, then went on to complete a fine PhD thesis on the language; see Onishi (1994). When he was finished I asked: ‘Those analyses the students had at the end of the field methods course don’t bear too much relation to what the language is like, now that you’ve understood it pretty well, do they?’ Masa simply smiled in response.

- (f) I always choose for a field methods class a language that I know nothing about, and for which there is no published description. I work with the consultant for just an hour or so before the course commences to make sure that everything will be alright. I take little active part in the class myself. Almost all the time is taken up by the students taking it in turn to work with the consultant. I listen to what is happening and offer suggestions and comments on how to proceed.

Some teachers spend most of the course working with the consultant themselves, with the students sitting there taking notes. They are showing the students how to do it. I get the students to *do it themselves*, to learn by experience.

- (g) It follows from (e) and (f) that a teacher should never aim to publish anything at the end of the field methods course. Certainly not a collection of students’ assignments. If the teacher publishes a paper or grammar sketch under their own name, on the basis of a field methods course, then—in my opinion—they have approached the course in the wrong way, as if it were for their benefit rather than for the benefit of the students.

I know of instances of a linguist publishing a paper on some aspect of a language solely on the basis of a field methods course they had taught. When someone else undertook extensive fieldwork (in a field location) on the language the ‘results’ of the field methods paper were shown to be absolutely erroneous. In at least once instance some crucial aspects of a formal theory were revised on the basis of such a field methods course paper—see Lawler (1975, 1977). Later, another linguist did extensive fieldwork on the language and pointed out manifold errors in work which had been based on elicitation in the context of a field linguistics course—see Durie (1988).

## Note

This chapter is a revision of Dixon (2007a).

# Glossary

Definitions are provided for a number of technical terms which recur in these volumes. For some entries there is reference to the chapter or section in which they are discussed. Note that Chapters 1–9 are in Volume 1 while Chapters 10–18 comprise Volume 2. Complementary terms are cross-references by ‘Compl.’

**ABLATIVE:** marker indicating movement away from the referent of the noun phrase to which it is attached.

**ABSOLUTIVE:** case inflection marking intransitive subject (S) and transitive object (O). Compl. ergative. §3.9, §13.2, §13.5.4.

**ACCUSATIVE:** case inflection marking transitive object (O). Compl. nominative. §3.9, §13.2, §13.5.4.

**ACTIVE/STATIVE:** label covering split-S and fluid-S systems.

**ADJECTIVE:** class of words which typically refer to properties and have two main roles: (a) make a statement that something has a certain property through functioning in intransitive predicate slot or copula complement slot; and (b) help to specify the referent of the head noun in an NP by functioning as modifier to it. §3.6, §4.5, §6.1, §8.3.2, Chapter 12.

**ADPOSITION:** a marker of a (predominantly peripheral) grammatical relation which is realized as a separate phonological word or as a clitic, not as an affix. §5.4.

**AFFINAL:** kinship relation which involves a link by marriage. §1.3, §16.1.

**AFFIX:** a bound form added to a root or stem. §5.4.

**AFFIXATION:** morphological process which involves adding an affix to a root or stem. §3.13.

**AGGLUTINATIVE:** a type of language whose words are readily segmentable into a sequence of morphemes, each of which typically conveys one piece of information. §5.5.

**AGREEMENT:** when two words (for example, noun and modifying adjective within an NP) are marked for the same grammatical category. §5.6.

**AIRSTREAM MECHANISM:** a system for initiating a flow of air which will facilitate speech; see pulmonic, glottalic. §7.2.

**ALIENABLE POSSESSION:** when the possessed does not have an inherent connection with the possessor. §1.3, §16.5.

**ALLATIVE:** marker indicating movement towards the referent of the noun phrase to which it is attached.

**ALLOMORPH:** one of several alternative forms of a morpheme. §5.2.

**ALLOPHONE:** one possible pronunciation of a phoneme. §7.1.

- AMBITRANSITIVE:** verb which can function in both a transitive and an intransitive clause; of type  $S = A$  or  $S = O$ . §3.3, §13.3.
- ANALYTIC:** language whose words generally each have a small number of grammatical components. Compl. synthetic. §5.5.
- ANAPHORA:** a pronoun or demonstrative referring to something which was explicitly stated earlier in the discourse, such as *he* in *John came in and he sat down*. §15.3.
- ANTIPASSIVE:** valency-reducing derivation which puts underlying A argument into derived S function, and places underlying O argument in a peripheral function. §3.20.
- APPLICATIVE:** valency-increasing derivation which prototypically operates on an intransitive clause, putting underlying S argument into A function and introducing a new O argument (which may have been in peripheral function in the underlying clause). §3.20.
- ARCHIPHONEME:** unit resulting from the neutralization of a phonological contrast in a certain environment. §7.2.
- ARGUMENT, CORE:** an obligatory argument for a specific verb, which must be either stated or understood from the context. §3.2, §3.9, §5.6, §13.2.
- ARGUMENT, PERIPHERAL:** non-core argument, which is optional; typically includes instrument, accompaniment, recipient, beneficiary, time, place, manner. §3.9, §5.6.
- ARTICLE:** a type of determiner, whose prototypical role is to mark an NP as definite or indefinite. The label is used in special ways for particular languages; for instance the tradition in Fijian linguistics is to use ‘article’ for the first word of an NP, which is *a* or *na* if the NP head is a common noun and *o* if the head is a proper name or pronoun. §3.4, §3.18.
- ARTICULATORS:** an active articulator (for example, tongue tip) is brought into contact with—or into approximation with—a passive articulator (for example, the teeth). §7.2.
- ASPECT:** term used for composition (perfective/imperfective), sometimes also for boundedness, completion, etc. §3.15.
- ASSIMILATION:** a process by which one sound changes to become more similar to a neighbouring sound, for example *-nb-* becoming *-mb-*.
- ATELIC:** an event which is unbounded and has no definite end-point. Compl. telic. §3.15.
- AUGMENTED:** pronoun paradigm in which one or more further participants are added to each term in a minimal paradigm. Compl. minimal. §3.7, §15.1.2.
- AUXILIARY:** a grammatical form (sometimes called an auxiliary verb) which occurs together with a lexical verb. It typically inflects for some non-spatial setting categories, in place of the verb inflecting for these categories.
- AVERSIVE:** case which is added to a noun or pronoun referring to something for fear of which the action described by the verb of the clause takes place or should take place. For example, ‘Come away from the fire for fear of the flying sparks.’
- BENEFICIARY:** peripheral argument referring to someone who will benefit from an action, as in *John wrote the letter [for Mary]*<sub>BENEFICIARY</sub>.

- BOUNDEDNESS** (or telicity): grammatical category indicating whether or not an activity has a definite end-point; see telic, atelic. §3.15.
- BOUND FORM**: form which cannot occur alone but must be attached to some other form, e.g. *un-* in English. Compl. free form. §5.2.
- CASE**: a system of nominal inflections, marking the syntactic function of an NP in its clause. §1.5, §1.10, §13.2.
- CATAPHORA**: a pronoun or demonstrative referring to something which is explicitly stated earlier in the discourse, such as *he* in *After he stopped smoking, John lived to a ripe old age*. §15.3.
- CAUSAL**: peripheral argument whose referent is responsible for a state or activity, as in *John is sick [from eating rotten meat]<sub>CAUSAL</sub>*.
- CAUSATIVE**: valency-increasing derivation which prototypically operates on an intransitive clause, putting underlying S argument into O function and introducing a 'causer' as A argument. §3.20.
- CIRCUMFIX**: a type of affix made up of one part which precedes the root or stem (like a prefix) and one part which follows (like a suffix). §5.2.
- CLASSIFIERS**: a set of (free or bound) forms which serve to categorize most of the nouns of a language, typically in terms of shape, composition, arrangement, or function/use. §3.16.
- CLAUSE**: the description of some activity, state or property. Consists of an obligatory predicate which requires certain core arguments and may also have peripheral arguments. §3.2.
- CLITIC**: a surface element part-way between a word and an affix in its properties. It is typically a separate grammatical word which is attached to a contiguous phonological word. §5.4, §10.5.
- COGNATES**: forms which are historically related; that is, go back to a single original form.
- COMITATIVE**: an affix (generally derivational, sometimes inflectional) added to a form with reference X, giving the meaning 'with (accompanied by) an X' or 'having an X'. Compl. privative.
- COMMON ARGUMENT**: an argument shared, in their underlying structures, by main clause and relative clause within a relative clause construction. Chapter 17.
- COMPARATIVE CONSTRUCTION**: typically involves comparing two participants (the comparee and the standard) in terms of some property (the parameter) this being marked by an index. §3.23.
- COMPLEMENTARY DISTRIBUTION**: the occurrence of each of two or more items (sounds or forms) in mutually exclusive environments.
- COMPLEMENT CLAUSE**: clause which fills a (normally core) argument slot in a higher clause. §1.9, §3.10, Chapter 18.
- COMPLEMENTIZER**: grammatical form which marks a complement clause. Chapter 18.
- COMPLEMENT-TAKING VERB**: a verb which may have a complement clause filling one of its (generally, core) argument slots. Chapter 18.
- COMPLETION**: grammatical category covering perfect and imperfect. §3.15.
- COMPOSITION**: grammatical category covering perfective and imperfective. §3.15.

- COMPOUNDING:** morphological process which joins two roots to form one stem. §3.13
- CONCORD:** when two words (for example, noun and modifying adjective within an NP) are marked for the same grammatical category. §5.6.
- CONJUGATION:** a class of verbs all of which take the same inflectional allomorphs.
- CONJUNCT:** grammatical element showing that the subject is 1st person in a statement and 2nd person in a question. Compl. disjunct. §15.1.10.
- CONSANGUINEAL:** kinship relation which does not involve marriage but is entirely through descent (a 'blood relation'). §1.3, §16.1.
- CONSTITUENT:** anything which fills a slot in a syntactic structure. §5.6.
- CONSTITUENT ORDER:** the order in which phrasal constituents occur within a clause (often mis-termed 'word order'). §2.4, §5.6.
- CONSTRUCTION:** type of clause (or, sometimes, phrase) with specified properties. §5.6.
- CONTENT INTERROGATIVE:** question which enquires concerning a core or peripheral argument (including time, place, and manner), or predicate, or some action or state or property. A word defining such a question. §3.7.
- CONTINUOUS:** see durative.
- COPULA CLAUSE:** indicating a relational meaning between CS (copula subject) and CC (copula complement) functions. §3.2, Chapter 14.
- COPULA COMPLEMENT (CC):** the argument in a copula clause which is shown to be in a specified relation to the copula subject (typically, may be realized as a plain NP, an NP marked with a preposition, a possessive clause, an adjective, or a complement clause). Chapter 14.
- COPULA SUBJECT (CS):** that argument in a copula clause which is topic for the discourse in which it occurs (generally realized by an NP or a complement clause). Chapter 14.
- CORE ARGUMENT:** an obligatory argument for a specific verb, which must be either stated or understood from the context. §3.2, §3.9, §5.6, §13.2.
- COVERB:** word (generally non-inflecting) which may be combined with an inflecting verb to form a complex verbal lexeme. §1.11.
- DATIVE:** a case which typically marks the beneficiary of 'give', the addressee of 'tell', and the person to whom something is shown for 'show'.
- DECLARATIVE:** choice from a mood system used in a statement. §3.2.
- DEICTIC REFERENCE:** pointing to some participant, activity, or place within the context of speaking. §15.2.
- DEMONSTRATIVE:** grammatical element whose primary function is to point to an object in the situation of discourse; may also have anaphoric and/or cataphoric functions. §3.7, §§15.2–3.
- DERIVATION:** optional morphological process which applies to a root or stem and derives a stem; may or may not change word class. §3.13–14, §5.3.
- DETERMINER:** grammatical modifier within an NP, typically including demonstratives and articles.
- DIPHTHONG:** vowel phoneme which has two or more phonetic components. §4.9.
- DIRECT SPEECH:** verbatim quotation of what was said.
- DISJUNCT:** grammatical element showing that the subject is not 1st person in a statement and not 2nd person in a question. Compl. conjunct. §15.1.10.

- DISSIMILATION:** change by which one sound becomes more dissimilar to some neighbouring sound.
- DURATIVE** (also called continuous or progressive): an event seen as unfolding over a period of time. Compl. punctual. §3.15.
- ENCLITIC:** clitic which is attached to the end of a word. §5.4, §10.5.
- ERGATIVE:** case inflection marking transitive subject (A). Compl. absolutive. §3.9, §13.2, §13.5.4.
- EVIDENTIALITY:** grammatical system providing information about the evidence on which a report is based. §1.5, §3.15.
- EXCLUSIVE:** non-singular 1st person pronoun, referring to speaker and one or more other people who do not include the addressee. Compl. inclusive. §15.1.2.
- EXTENDED INTRANSITIVE:** clause type with two core arguments, in S (intransitive subject) and E (extension to core) functions. Verb which occurs in the predicate of such a clause. §3.2, §13.1.
- EXTENDED TRANSITIVE** (or ditransitive): clause type with three core arguments, in A (transitive subject), O (transitive object), and E (extension to core) functions. Verb which occurs in the predicate of such a clause. §3.2, §13.1.
- EXTENT:** grammatical category covering punctual and durative. §3.15.
- FLUID-S:** system where some verbs may have their S argument marked like A (Sa) or like O (So) with a difference in meaning. §3.9, §13.2, §13.5.4.
- FOCAL CLAUSE:** that clause in a linking construction which carries the mood of the sentence. §3.11.
- FOCUS:** an argument accorded prominence within a clause. §3.21.
- FORMAL MARKEDNESS:** if a term in a grammatical system has zero realization (or a zero allomorph) it is said to be formally unmarked. Other terms in the system are formally marked. §5.7.
- FREE FORM:** a form which constitutes a grammatical word without any morphological processes having to be applied. §5.2.
- FUNCTIONAL LOAD** of a contrast: the extent to which that contrast is utilized within that language.
- FUNCTIONAL MARKEDNESS:** a term in a grammatical system which is employed in neutral or unspecified circumstances (or when a contrast is neutralized) is said to be functionally unmarked. Other terms in the system are functionally marked. §5.7.
- FUSIONAL:** a type of language whose words involve a number of grammatical elements fused together (that is, not segmentable in surface structure). §5.5.
- GENDER:** small system of noun classes one of whose semantic distinctions is masculine/feminine. §1.5, §1.10, §3.16.
- GENITIVE:** marker of an intra-NP possessive relation, which is added to the possessor item. Compl. pensive. §1.10, §16.2.
- GLOTTALIC AIRSTREAM MECHANISM:** air movement initiated at the glottis. §7.2.
- GRAMMATICAL WORD:** a unit on the hierarchy of grammatical units (just below phrase) defined on grammatical criteria. Generally (but not necessarily always) coinciding with phonological word. Chapter 10.



- HEAD:** obligatory nucleus of a phrase which determines the grammatical profile of the whole phrase (for example, gender of a noun phrase). §3.4, §5.6, §16.8, §17.2.
- HETERORGANIC:** sequence of sounds which have different place of articulation, for example *-nb-*.
- HOMORGANIC:** sequence of sounds which have the same place of articulation, for example *-mb-*.
- IDEOPHONE:** word class which often has special phonology (often involving inherent reduplication and onomatopoeia). Typically relating to manner, colour, sound, smell, action, state, or intensity. §8.3.
- IMPERATIVE:** choice from a mood system used in a direct command. §1.5, §3.2.
- IMPERFECT:** something which began in the past and is still continuing. Compl. perfect. §3.15.
- IMPERFECTIVE:** focusing on the temporal make-up of an event. Compl. perfective. §3.15.
- INALIENABLE POSSESSION:** when the possessed has an inherent connection with the possessor, and cannot be given away. §1.3, §16.5.
- INCLUSIVE:** non-singular 1st person pronoun, referring to speaker and one or more other people who do include the addressee. Compl. exclusive. §15.1.2.
- INDIRECT SPEECH:** a report of what someone else has said (often cast into the reporter's own words).
- INFLECTION:** morphological process which obligatorily applies to a root or derived stem of a certain word class, producing a grammatical word. §3.13, §5.3.
- INSTRUMENTAL:** case inflection marking the referent of the NP to which it is attached as weapon, tool, or material used in the activity described by the verb. §4.3, §13.2.1.
- INTERJECTION:** a conventionalized cry, typically indicating the speaker's emotional response to something that has happened to them, or something which they have observed or become aware of. §10.7.
- INTERNAL CHANGE:** morphological process which involves changing a vowel (or, less frequently, a consonant) in the middle of a word, for instance, from *take* /teik/ to *took* /tuk/ in English. §3.13.
- INTERROGATIVE:** choice from a mood system used in a (content or polar) question. A content interrogative word. §3.2, §3.7.
- INTONATION:** type of prosody realized by pitch, generally applying over clause or sentence. §7.6.
- INTRANSITIVE:** clause type with one core argument, in S (intransitive subject) function. Verb which occurs in the predicate of such a clause. §3.2, §5.6, Chapter 13.
- IRREALIS:** referring to something that didn't happen (but could have happened) or which might happen. Compl. realis. §3.15.
- ISOLATING:** a type of language most of whose words consist of one morpheme. §5.5.
- LABILE:** older name for ambitransitive.
- LANGUAGE:** in the technical sense of linguists, a number of forms of speech are said to constitute a single language if they are mutually intelligible.
- LENITION:** the replacement of a sound by another sound that has the same place of articulation but a weaker manner of articulation (involving less muscular tension).

- LEXEME (OR LEXICAL ITEM):** a root or underlying form. §10.2.
- LOCATIVE:** marker indicating position of rest at, on, or near the referent of the noun phrase to which it is attached.
- LOGOPHORIC PRONOUN:** used in a complement clause, this refers back to the subject of the matrix clause. §15.3.4.
- MARKEDNESS:** see formal markedness, functional markedness. §5.7.
- MINIMAL:** pronoun paradigm in which ‘me and you’ is a term on a par with 1st person singular and 2nd person singular (and, in some languages, 3rd person singular). Compl. augmented. §3.7, §15.1.2.
- MODALITY:** one of a number of choices (within irrealis) referring to some aspect of the future. §3.15.
- MODAL VERB:** a verb which indicates a modality.
- MOOD:** grammatical system indicating the pragmatic function of a sentence, covering indicative (for a statement), interrogative (for a question), and imperative (for a command). §3.2.
- MORA:** unit between phoneme and syllable, variously defined. §7.6.
- MORPHEME:** the minimum meaningful unit of speech. §5.2.
- MORPHOLOGICAL PROCESS:** process which applies to a root, forming a stem. §3.13.
- MORPHOLOGY:** that part of grammar which studies the structure of words. Compl. syntax. §3.13, §5.2.
- NEUTRALIZATION:** when a certain grammatical or phonological contrast may not apply in a certain environment, it is then said to be neutralized. §5.7, §7.2, §15.1.3.
- NOMINAL HIERARCHY:** hierarchy of items which can be head of an NP, according to how likely they are to be in A rather than in O function. §3.9; §13.5.4.
- NOMINALIZATION:** morphological derivation which forms a noun stem from a verb or adjective root or stem. §3.14.
- NOMINATIVE:** case inflection marking intransitive subject (S) and transitive subject (A). Compl. accusative. §3.9, §13.2, §13.5.4.
- NON-CANONICAL MARKING OF CORE ARGUMENTS:** when most of the instances of a core argument receive a certain marking, but there are a minority of instances which attract a different marking, this is termed non-canonical. §13.6.
- NON-SPATIAL SETTING:** covers the range of parameters which describe the setting for an activity or state other than those referring to spatial location. It typically includes evidentiality, reality, degree of certainty, phase of activity, completion, boundedness, extent, composition (some of the last three, and more besides, may be called aspect), and tense. §3.15.
- NOUN:** word class whose primary function is as head of an NP; many of its members refer to concrete objects. §3.3, §8.3.1, Chapter 11.
- NOUN CLASSES:** grouping of all the nouns of a language into a number of small classes which comprise a small closed grammatical system. Noun class membership must be marked somewhere outside the noun itself. Also see gender. §1.9, §3.16.
- NOUN INCORPORATION:** the incorporation of a noun (generally in underlying S or O function) into a verb to create a compound stem.

**NOUN PHRASE (NP):** a constituent which can fill an argument slot in clause structure.

It has a noun or pronoun or demonstrative, etc. as head. §3.4, §5.6, §11.4.

**NP:** see noun phrase.

**NUMBER:** grammatical system one of whose terms is singular. There will be one or more further terms. §1.4, §13.7.

**PASSIVE:** valency-reducing syntactic derivation which puts underlying O argument into derived S function and places underlying A argument in a peripheral function. §3.20.

**PAUCAL NUMBER:** referring to 'a few' (more than two).

**PERFECT:** a past action which is completed but still has present relevance. Compl. imperfect; §3.15.

**PERFECTIVE:** an event regarded as a whole, without regard for its temporal constituency. Compl. imperfective; §3.15.

**PERIPHERAL ARGUMENT:** a non-core argument, which is optional. Typically includes instrument, accompaniment, recipient, beneficiary, time, place, manner. §3.9, §5.6.

**PERIPHERAL PLACE OF ARTICULATION:** cover term for sounds made at the front or back of the mouth, covering bilabial and dorso-velar.

**PERSON:** speech act participants; always including 1st person (speaker) and 2nd person (addressee), and sometimes also 3rd person (neither speaker nor addressee). §15.1.1.

**PERTENSIVE:** marker of an intra-NP possessive relation, which is added to the possessed item. Compl. genitive. §16.2.

**PHASE OF ACTIVITY:** whether beginning, continuing, ending, etc. §3.15.

**PHONEME:** the minimum segmentable unit of phonology. §7.1.

**PHONETICS:** articulatory and/or acoustic study of the sounds of speech.

**PHONOLOGICAL WORD:** a unit on the hierarchy of phonological units (just above syllable) defined on phonological criteria. Generally (but not necessarily always) coinciding with grammatical word. Chapter 10.

**PHONOLOGY:** description of the phonetic contrasts which are used to distinguish between distinct words in a given language. Chapter 7.

**PHONOTACTICS:** statement of which consonants and vowels may correspond to each structural slot in syllable (and word) structure. §7.4.

**PHRASE:** a constituent which can fill a slot in clause structure—noun phrase in an argument slot and verb phrase in predicate slot. §3.4.

**PIVOT:** a topic which is recognizable as such by its grammatical properties. §3.21.

**POLAR QUESTION:** question enquiring whether or not a proffered statement is correct. Can be answered by 'yes' or 'no' in languages which have such words (not all do). §3.2.

**POLARITY:** grammatical system whose terms are positive and negative. §3.12.

**POLYSYNTHETIC:** highly synthetic. §5.5.

**POSSESSIVE PHRASE:** a type of NP which is included within a larger NP and indicates the possessor with respect to the head of the larger NP, which is the possessed. §3.4, Chapter 16.

**POSTPOSITION:** an adposition which follows the constituent for which it provides grammatical marking. §5.4.

- PRAGMATICS:** the practical consequences of the use of a given portion of language.
- PREDICATE:** the central (and obligatory) structural element of a clause, generally realized by a verb phrase (with verb as head). It determines the number and type of core arguments required in the clause. §2.5, §3.2, §11.5.
- PREFIX:** an affix which precedes a root or stem.
- PREPOSITION:** an adposition which precedes the constituent for which it provides grammatical marking. §5.4.
- PRIMARY VERBS:** referring directly to an activity or state. Compl. secondary verbs. §1.11, §18.5.
- PRIVATIVE:** an affix (generally derivational, sometimes inflectional) added to form with referent X, giving the meaning 'without an X'. Compl. comitative.
- PROCLITIC:** clitic which is attached to the beginning of a word: §5.4, §10.5.
- PROGRESSIVE:** see durative.
- PRONOUN:** small closed class of grammatical items which relate to person (and usually also to number). Can be free forms or bound forms. §3.7; §15.1.
- PROSODY:** a system of phonological contrasts which has scope over a sequence of segments. §7.5.
- PROTO-LANGUAGE:** putative single ancestor language for a group of modern languages that are held to be genetically related, each having developed by regular changes from the proto-language.
- PULMONIC AIRSTREAM MECHANISM:** air movement initiated in the lungs. §7.2.
- PUNCTUAL:** an event which happens more or less instantaneously. Compl. durative. §3.15.
- REALIS:** referring to something that has happened or is happening. Compl. irrealis. §3.15.
- REALITY:** grammatical category covering realis and irrealis. §3.15.
- RECIPROCAL:** clause describing several instances of an activity such that what is A argument in one instance is O argument in another. §3.22.
- REDUPLICATION:** morphological process which involves repeating all or part of a root (or stem or full word) either before, after, or in the middle of it. §3.13.
- REFLEXIVE:** clause in which underlying A and O arguments have the same reference. §3.22.
- RELATIVE CLAUSE:** clause which modifies the head of an NP. Relative clause and main clause share, in their underlying structures, a common argument. Chapter 17.
- ROOT:** unanalysable lexical element.
- S = A AMBITRANSITIVE:** the S argument, when the verb is used intransitively, corresponds to the A argument, when it is used transitively. §3.3, §13.3.
- S = O AMBITRANSITIVE:** the S argument, when the verb is used intransitively, corresponds to the O argument, when it is used transitively. §3.3, §13.3.
- SECONDARY CONCEPTS:** provide modification for a primary verb. May be realized as an affix or as a verb (a secondary verb). Compl. primary verb. §1.11; §18.5.
- SEMANTIC ROLES:** the types of participant involved with verbs of a certain semantic type. §1.9, §3.3, §13.5.1.

- SEMANTICS:** study of the meaning relations conveyed by the grammatical systems and lexical contrasts of a language.
- SEMANTIC TYPE:** a set of words with similar meanings and grammatical properties. §1.9, §1.11, §3.3, §8.3, §12.4, §13.5.12, §18.5.
- SENTENCE:** no simple definition is feasible—see §3.11.
- SERIAL VERB CONSTRUCTION:** has a predicate consisting of two (or more) verbs, each of which could make up a predicate on its own, and whose combination is conceived of as describing a single action; there must be a single subject applying to the whole. §18.6.1.
- SHIFTER:** grammatical item whose reference changes depending on who is speaking (pronouns) or what the place or time is. §3.7.
- SPLIT-S:** system where the S argument for some verbs is marked like A (Sa) and for other verbs S is marked like O (So); also called active/stative. §3.9, §13.2, §13.5.4.
- STATIVE/ACTIVE:** label covering split-S and fluid-S systems.
- STEM:** the nucleus of a word, to which an inflectional process applies, forming a word.
- STRESS (or accent):** a contrastive prosody generally having scope over a word, characterized by some or all of: loudness, vowel quality, pitch, and length. §7.6.
- SUBGROUP:** set of languages within a language family which descend from a single ancestor language, this being itself a descendant of the proto-language for the whole language family.
- SUBTRACTION:** morphological process which involves deleting something from a root. §3.13.
- SUFFIX:** an affix which follows a root or stem.
- SUPPLETION:** when a lexeme has two forms which are not cognate (as *go* and *went* in English).
- SUPPORTING CLAUSE:** that clause in a linking construction which does not carry the mood of the sentence. §3.11.
- SYLLABLE:** a phonological unit centred on a nucleus (typically a vowel) which may be preceded and/or followed by one or more consonants. §1.4, §6.3, §7.4.
- SYNCHRONIC DESCRIPTION:** description of a language system at one point in time, without taking account of historical changes.
- SYNTAX:** study of the organization and interrelation of the components of a grammar above the level of word.
- SYNTHETIC:** language whose words generally each have a large number of grammatical components. Compl. analytic. §5.5.
- TELIC:** an event which is bounded and has a definite end-point. Compl. atelic. §3.15.
- TENSE:** grammatical category, with shifting reference, which refers to time. §1.5, §1.7, §1.10, §3.15.
- TOPIC:** an argument which occurs in a succession of clauses in a discourse and binds them together. §3.21.
- TRANSITIVE:** clause type with two core arguments, in A (transitive subject) and O (transitive object) functions. Verb which occurs in the predicate of such a clause. §3.2, §5.6, Chapter 13.

- TRIPARITE MARKING: when each of transitive subject (A), intransitive subject (S), and transitive object (O) receives a distinct surface marking. §3.9, §13.2.
- UNMARKED: see formal markedness, functional markedness.
- VALENCY: the number of core arguments a verb requires.
- VALENCY-CHANGING: derivations which may increase valency (causative, applicative) or decrease it (passive, antipassive, some varieties of reflexive and reciprocal, etc.). §3.20.
- VERB: word class whose primary function is as head of a predicate. Most of its members refer to actions and states. §3.3, §8.3.3, Chapter 11.
- VERBALIZATION: morphological derivation which forms a verb stem from a noun or adjective root or stem. §3.14.
- VERBLESS CLAUSE: similar to a copula clause but with the predicate slot left blank. It indicates a relational meaning between verbless clause subject and verbless clause complement. Chapter 14.
- VERBLESS CLAUSE COMPLEMENT (VCC): the argument in a verbless clause which is shown to be in a specified relation to the verbless clause subject (typically, may be realized as a plain NP, an NP marked with a preposition, a possessive clause, an adjective, or a complement clause). Chapter 14.
- VERBLESS CLAUSE SUBJECT (VCS): that argument in a verbless clause which is topic for the discourse in which it occurs (generally realized by an NP or a complement clause). Chapter 14.
- VERB PHRASE: a constituent which can fill the predicate slot within a clause. Typically has a verb as its head. §3.4, §5.6.
- VOWEL HARMONY: prosody applying over a phonological stretch (typically, a phonological word) whereby all vowels within that stretch agrees in some feature, e.g. front/back.
- WORD: the result of applying optional derivational processes to a root, and then any obligatory inflectional process to the resulting stem. Subtypes: phonological word, grammatical word. Unit at the intersection of morphology and syntax. §3.1, Chapter 10.
- WORD ORDER: the order in which words must or may occur in a phrase, in a clause, or in a sentence. (This label is often misleadingly used for (phrasal) constituent order.) §2.4, §5.6.
- YES/NO QUESTION: see polar question.
- ZERO: when one term in a grammatical system has no explicit marking it is said to have zero realization ( $\emptyset$ ). For example, in English a noun with singular number reference receives zero marking (for instance *horse- $\emptyset$*  whereas one with plural reference is marked by orthographic *-s* (*horse-s*). §3.13, §5.3.
- ZERO ANAPHORA: when anaphora is shown simply by leaving a gap. Compare anaphoric *he* in *John came in and he sat down* with anaphoric  $\emptyset$  in *John came in and  $\emptyset$  sat down*. §15.3.
- ZERO DERIVATION: a word-class-changing derivation with zero marking. Compare noun *hospital* and verbalization *hospital-ize*, marked by *-ize*, with noun *market* and verbalization *market- $\emptyset$* , with zero marking. §3.5, §3.13, §11.3.

# References

- Abbott, Miriam. 1991. 'Macushi', pp. 23–160 of *Handbook of Amazonian languages*, Vol. 3, edited by Desmond C. Derbyshire and Geoffrey K. Pullum. Berlin: Mouton de Gruyter.
- Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Aikhenvald, Alexandra Y. 1998. 'Warekena', pp. 215–429 of *Handbook of Amazonian languages*, Vol. 4, edited by Desmond C. Derbyshire and Geoffrey K. Pullum. Berlin: Mouton de Gruyter.
- 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- 2002a. *Dicionário Tariana–Português e Português–Tariana (Boletim do Museu Paraense Emílio Goeldi*, Vol. 17, Part 2). Belém: Editora do Museu Paraense Emílio Goeldi.
- 2002b. 'Typological parameters for the study of clitics, with special reference to Tariana', pp. 42–78 of Dixon and Aikhenvald 2002.
- 2003. *A grammar of Tariana, from northwest Amazonia*. Cambridge: Cambridge University Press.
- 2004. *Evidentiality*. Oxford: Oxford University Press.
- 2007. 'Typological distinctions in word formation', pp. 1–65 of Shopen 2007, Vol. 3.
- 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.
- Forthcoming. *Imperatives and commands*. Oxford: Oxford University Press.
- and Dixon, R. M. W. 1998. 'Dependencies between grammatical systems', *Language* 74: 56–80.
- ——— 2006. Editors of *Serial verb constructions: A cross-linguistic typology*. Oxford: Oxford University Press.
- ——— and Onishi, Masayuki. 2001. Editors of *Non-canonical marking of subject and objects*. Amsterdam: John Benjamins.
- Aissen, Judith L. 1992. 'Fieldwork and linguistic theory', pp. 9–11 of *International encyclopedia of linguistics*, Vol. 2, edited by William Bright. New York: Oxford University Press.
- Allen, W. S. 1951. 'Some prosodic aspects of retroflexion and aspiration in Sanskrit', *Bulletin of the School of Oriental and African Studies* 13: 939–46. [Reprinted as pp. 82–90 of Palmer 1970.]
- Amberber, Mengistu. 1995. 'Amharic grammar summary'. RCLT internal document.
- Ameka, Felix. 1996. 'Body parts in Ewe grammar', pp. 780–840 of *The grammar of inalienability: A typological perspective on body part terms and the part-whole relation*, edited by Hilary Chappell and William McGregor. Berlin: Mouton de Gruyter.

- Anderson, Victoria, and Otsuka, Yuko. 2006. 'The phonetics and phonology of "defin-itive accent" in Tongan', *Oceanic Linguistics* 45: 21–42.
- Andrade, Manuel J. 1933. 'Quileute', pp. 149–292 of *Handbook of American Indian languages*, Vol. 3, edited by Franz Boas. New York: Columbia University Press.
- Andrews, Edna. 1990. *Markedness theory: The union of asymmetry and semiosis in language*. Durham, NC: Duke University Press.
- Ashton, E. O. 1947. *Swahili grammar (including intonation)*, 2nd edition. London: Longmans.
- Aspinion, Robert. 1953. *Apprenons le berbère: Initiation aux dialectes chleuhs*. Paris: Félix Moncho.
- Bakaev, C. H. 1966. 'Kurdskij jazyk' ['Kurdish language'], pp. 257–80 of *Jazyki narodov SSSR Indoevropskie Jazyki [Languages of the peoples of the USSR: Indo-European languages]*, edited by V. V. Vinogradov. Moscow: Nauka.
- Bani, Ephraim, and Klokeid, Terry J. 1971. 'An outline of Mabuig morphology'. MS. Australian Institute of Aboriginal Studies.
- Barnes, Janet. 1979. 'Los imperativos en tuyuca', *Artículos en lingüística y campos afines* 6: 87–94.
- 1984. 'Evidentials in the Tuyuca verb', *International Journal of American Linguistics* 50: 255–71.
- Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.
- and Huddleston, Rodney. 2002. '19. Lexical word-formation', pp. 1621–721 of *The Cambridge grammar of the English language*, by R. D. Huddleston and G. K. Pullum (chief authors). Cambridge: Cambridge University Press.
- Bazell, C. E. 1958. *Linguistic typology*. London: School of Oriental and African Studies.
- Bazin, Louis. 1983. 'La réforme linguistique en Turquie', pp. 155–77 of *Language reform: History and future*, edited by István Fodor and Claude Hagège, Vol. 1. Hamburg: Buske.
- Benveniste, Emile. 1971. *Problems in general linguistics*. Coral Gables: University of Florida Press. [Translated by Mary Elizabeth Meek from the French original *Problèmes de linguistique générale*. Paris: Gallimard, 1966.]
- Bergslund, Knut, and Vogt, Hans. 1962. 'On the validity of glottochronology', *Current Anthropology* 3: 115–58.
- Berlin, Brent. 1992. *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton: Princeton University Press.
- and Kay, Paul. 1969. *Basic color terms: Their universality and evolution*. Berkeley and Los Angeles: University of California Press.
- Bloch, Bernard. 1941. 'Phonemic overlapping', *American Speech* 16: 278–84.
- Bloomfield, Leonard. 1927. 'Literate and illiterate speech', *American Speech* 2: 432–9. [Reprinted as pp. 391–6 of Hymes 1964.]
- 1933. *Language*. New York: Holt. [Reissued in 1984 by University of Chicago Press.]
- Boas, Franz. 1911. 'Introduction', pp. 1–83 of *Handbook of American Indian languages*, Part 1, edited by Franz Boas. Washington, DC: Smithsonian Institution.



- Boas, Franz and Deloria, Ella. 1941. *Dakota grammar*. Washington, DC: US Government Printing Office.
- Bolinger, Dwight L. 1950. 'Rime, assonance and morpheme analysis', *Word* 6: 117–36.
- Brown, D. Richard. 1994. 'Kresh', pp. 163–89 of *Typological studies in negation*, edited by Peter Kahrel and René van den Berg. Amsterdam: John Benjamins.
- Burridge, Kate. 2002. 'Changes within Pennsylvania German grammar', pp. 207–32 of *Ethnosyntax: Explorations in grammar and culture*, edited by N. J. Enfield. Oxford: Oxford University Press.
- Burzio, Luigi. 1981. 'Intransitive verbs and Italian auxiliaries'. Ph.D. dissertation, MIT.
- Bybee, Joan, Perkins, Revere, and Pagliuca, William. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Canger, U. 1994. 'Fieldwork and field methods', pp. 1219–21 of *The encyclopedia of language and linguistics*, Vol. 3, edited by R. E. Asher. Oxford: Pergamon Press.
- Carlson, Robert. 1994. *A grammar of Supyire*. Berlin: Mouton de Gruyter.
- Cawdrey, Robert. 1613. *A table alphabeticall, contayning and teaching the true writing and understanding of hard vsuall English words, borrowed from the Hebrew, Greeke, Latine, or French, &c . . .* Set forth by R.C. [i.e. Robert Cawdrey] corrected and much enlarged . . . The 3. edition. London: Edmund Weauer.
- Chalker, Sylvia, and Weiner, Edmund. 1994. *The Oxford dictionary of English grammar*. Oxford: Oxford University Press.
- Chao, Yuen-Ren. 1934. 'The non-uniqueness of phonemic solutions of phonetic systems', *Bulletin of the Institute of History of Philology, Academica Sinica* 4: 363–97. [Reprinted as pp. 38–54 of Joos 1958.]
- Charney, Jean Ormsbee. 1991. *A grammar of Comanche*. Lincoln: University of Nebraska Press.
- Churchill, Winston S. 1951. *The Second World War*, Vol. 4: *The hinge of fate*. London: Cassell.
- Churchward, C. Maxwell. 1940. *Rotuman grammar and dictionary*. Methodist Church of Australasia.
- 1953. *Tongan grammar*. London: Oxford University Press.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- 1989. *Language universals and linguistic theory*, 2nd edition. Oxford: Basil Blackwell.
- Cook, Eung-Do. 1984. *A Sarcee grammar*. Vancouver: University of British Columbia Press.
- 2004. *A grammar of Dëne Suñiné (Chipewyan)*. Winnipeg: Algonquian and Iroquoian Linguistics.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- 2000. *Number*. Cambridge: Cambridge University Press.
- Craig, Colette Grinevald. 1977. *The structure of Jacaltec*. Austin: University of Texas Press.
- 1986. 'Jacaltec noun classifiers: A study in language and culture', pp. 263–93 of *Noun classes and categorization*, edited by Colette Craig. Amsterdam: John Benjamins.

- Croft, William. 1990. *Typology and universals*. Cambridge: Cambridge University Press.
- Cytowic, Richard E. 1989. *Synesthesia: A union of the senses*. New York: Springer.
- Dahl, Otto Christian. 1976. *Proto-Austronesian*. Lund: Studentlitteratur and London: Croom Helm.
- Day, Christopher. 1973. *The Jacalteco language (Language Science Monographs, Vol. 12)*. Bloomington: Indiana University Press and The Hague: Mouton.
- Dench, Alan Charles. 1995. *Martuthunira: A language of the Pilbara region of Western Australia*. Canberra: Pacific Linguistics.
- Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. Amsterdam: John Benjamins.
- Diffloth, Gerard. 1994. 'i: big, a: small', pp. 107–114 of Hinton, Nichols, and Ohala 1994.
- Dirr, A. 1912. 'Rutulskij jazyk' ['The Rutul language'], *Sbornik Materialov dlya Opisanija Plemen Kavkaza* (Tbilisi) 42, 3: 1–204.
- Dixon, R. M. W. 1968. 'The Dyirbal language of North Queensland'. PhD thesis, University of London.
- 1972. *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.
- 1973. 'A method of semantic description', pp. 436–71 of *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*, edited by D. D. Steinberg and L. A. Jakobovits. Cambridge: Cambridge University Press.
- 1977a. *A grammar of Yidij*. Cambridge: Cambridge University Press.
- 1977b. 'Some phonological rules in Yidin?', *Linguistic Inquiry* 8: 1–34.
- 1977c. 'Where have all the adjectives gone?', *Studies in Language* 1: 19–80.
- 1979. 'A note on Dyirbal ergativity', pp. 90–1 of *Papers from the Fifteenth Regional Meeting of the Chicago Linguistic Society*, edited by Paul R. Clyne, William F. Hanks, and Carol L. Hofbauer. Chicago: Chicago Linguistic Society, University of Chicago.
- 1980. *The languages of Australia*. Cambridge: Cambridge University Press.
- 1981. 'Wargamay', pp. 1–144 of Dixon and Blake 1981.
- 1982. *Where have all the adjectives gone? And other papers in semantics and syntax*. Berlin: Mouton.
- 1983. 'Nyawaygi', pp. 430–525 of Dixon and Blake 1983.
- 1984. *Searching for Aboriginal languages: Memoirs of a field worker*. St Lucia: Queensland University Press. [Reprinted in 1989 by University of Chicago Press.]
- 1988. *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press.
- 1989a. 'Subject and object in universal grammar', pp. 91–118 of *Essays on grammatical theory and universal grammar*, edited by Doug Arnold, Martin Atkinson, Jacques Durand, Claire Grover, and Louise Sadler. Oxford: Clarendon Press.
- 1989b. 'The Dyirbal kinship system', *Oceania* 59: 245–68.
- 1991a. *A new approach to English grammar, on semantic principles*. Oxford: Oxford University Press.
- 1991b. *Words of our country: Stories, place names and vocabulary in Yidiny, the Aboriginal language of the Cairns/Yarrabah region*. St Lucia: University of Queensland Press.
- 1991c. 'Mbabaram', pp. 348–402 of Dixon and Blake 1991.
- 1994. *Ergativity*. Cambridge: Cambridge University Press.

- Dixon, R. M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.
- 2000. 'A typology of causatives: form, syntax and meaning', pp. 30–83 of *Changing valency: Case studies in transitivity*, edited by R. M. W. Dixon and Alexandra Y. Aikhenvald. Cambridge: Cambridge University Press.
- 2002. *Australian languages: Their nature and development*. Cambridge: Cambridge University Press.
- 2004a. *The Jarawara language of southern Amazonia*. Oxford: Oxford University Press.
- 2004b. 'Adjective classes in typological perspective', pp. 1–49 of Dixon and Aikhenvald 2004.
- 2004c. 'Proto-Arawá phonology', *Anthropological Linguistics* 46: 1–83.
- 2005a. *A semantic approach to English grammar*. Oxford: Oxford University Press.
- 2005b. 'Comparative constructions in English', *Studia Anglica Posnaniensia* 41: 5–27.
- 2006a. 'Complement clauses and complementation strategies in typological perspective', pp. 1–48 of Dixon and Aikhenvald 2006.
- 2006b. 'Complement clause type and complementation strategy in Jarawara', pp. 93–114 of Dixon and Aikhenvald 2006.
- 2007a. 'Field linguistics: A minor manual', *Linguistic fieldwork* (edited by Alexandra Y. Aikhenvald), special issue of *Language Typology and Universals* 60: 12–31.
- 2007b. 'Clitics in English'. *English Studies* 88: 574–600.
- 2008. 'Comparative constructions: a cross-linguistic typology', *Studies in Language* 32: 787–817.
- 2009. 'The semantics of clause linking in typological perspective', Chapter 1 of *The semantic of clause linking: A cross-linguistic typology*, edited by R. M. W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press.
- and Aikhenvald, Alexandra Y. 2000. 'Introduction', pp. 1–29 of *Changing valency: Case studies in transitivity*, edited by R. M. W. Dixon and Alexandra Y. Aikhenvald. Cambridge: Cambridge University Press.
- — 2002. Editors of *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press.
- — 2004. Editors of *Adjective classes: A cross-linguistic typology*. Oxford: Oxford University Press.
- — 2006. Editors of *Complementation: A cross-linguistic typology*. Oxford: Oxford University Press.
- and Blake, Barry J. 1979–2000. Editors of *Handbook of Australian languages*, Vol. 1, 1979; Vol. 2, 1981; Vol. 3, 1983, Canberra: Australian National University Press and Amsterdam: John Benjamins; Vol. 4, 1991; Vol. 5, 2000, Melbourne: Oxford University Press.
- Doke, C. M. 1935. *Bantu linguistic terminology*. London: Longmans, Green.
- Donaldson, Tamsin. 1980. *Ngiyambaa: The language of the Wangaaybuwan*. Cambridge: Cambridge University Press.
- Douglas, W. H. 1964. *An introduction to the Western Desert language*, revised edition (Oceania Linguistic Monographs, No. 4). Sydney: Oceania Publications.

- Du Bois, John W. 1987. 'The discourse basis of ergativity', *Language* 63: 805–55.
- Dunn, Michael John. 1999. 'A grammar of Chukchi'. PhD thesis. Australian National University.
- Duponceau, Peter S. 1819. 'Report of the Corresponding Secretary to the committee of his progress in the investigation committed to him of the General Character and Forms of the Languages of the American Indians—Read, 12th Jan, 1819', *Transactions of the Historical and Literary Committee of the American Philosophical Society, held at Philadelphia, for Promoting Useful Knowledge*. Vol. 1, pp. xvii–xlvi. Philadelphia: Abraham Small.
- Durie, Mark. 1988. 'The so-called passives of Acehnese', *Language* 64: 104–13.
- Eades, Diana. 1979. 'Gumbaynggir', pp. 244–361 of Dixon and Blake 1979.
- Eckert, Paul, and Hudson, Joyce. 1988. *Wangka Wiru: A handbook for the Pitjantjatjara language learner*. Underdale: University of South Australia.
- Enfield, N. J. 2007. *A grammar of Lao*. Berlin: Mouton de Gruyter.
- England, Nora C. 1983. *A grammar of Mam, a Mayan language*. Austin: University of Texas Press.
- Evans, Nicholas D. 1995. *A grammar of Kayardild, with historical-comparative notes on Tangkic*. Berlin: Mouton de Gruyter.
- Fleck, David W. 2008. 'Coreferential fourth-person pronouns in Matses', *International Journal of American Linguistics* 74: 279–311.
- Fodor, István. 1983. 'Hungarian: Evolution—stagnation—reform—further development', pp. 49–84 of *Language reform, history and future*, edited by István Fodor and Claude Hagège, Vol. 2. Hamburg: Buske.
- Foley, William A. 1991. *The Yimas language of New Guinea*. Stanford, Calif.: Stanford University Press.
- Frajzyngier, Zygmunt. 2002. *A grammar of Hdi*. Berlin: Mouton de Gruyter.
- Freeland, L. S. 1951. *Language of the Sierra Miwok*. Indiana University Publications in Anthropology and Linguistics, Memoir 6. (Supplement to the *International Journal of American Linguistics*, Vol. 17, No. 1.)
- Fries, Charles C., and Pike, Kenneth L. 1949. 'Coexistent phonemic systems', *Language* 25: 29–50.
- Glass, Ameer, and Hackett, Dorothy. 1970. *Pitjantjatjara grammar: A tagmemic view of the Ngaanyatjara (Warburton Ranges) dialect*. Canberra: Australian Institute of Aboriginal Studies.
- Goddard, Cliff. 1985. *A grammar of Yankunytjatjara*. Alice Springs: Institute for Aboriginal Development. [Publication of 1983 ANU PhD thesis, 'A semantically-oriented grammar of the Yankunytjatjara dialect of the Western Desert language'.]
- Gordon, Lynn. 1986a. 'The development of evidentials in Maricopa', pp. 75–88 of *Evidentiality: The linguistic coding of epistemology*, edited by Wallace Chafe and Johanna Nichols. Norwood, NJ: Ablex.
- 1986b. *Maricopa morphology and syntax*. Berkeley and Los Angeles: University of California Press.
- Gordon, Raymond G., Jr. 2005. Editor of *Ethnologue: Languages of the world*, 15th edition. Dallas: SIL International.
- Greenberg, Joseph H. 1957. *Essays in linguistics*. Chicago: University of Chicago Press.

- Greenberg, Joseph H. 1963. 'Some universals of grammar with particular reference to the order of meaningful elements', pp. 73–113 of *Universals of language*, edited by Joseph H. Greenberg. Cambridge, Mass.: MIT Press.
- 1966. *Language universals*. The Hague: Mouton.
- Grey, Louis H. 1934. *Introduction to Semitic comparative linguistics*. New York: Columbia University Press.
- Grimes, Barbara. 2000. Editor of *Ethnologue*, Vol. 1: *Languages of the world*. Dallas: SIL International.
- Guirardello, Raquel. 1992. 'Aspectos da morfosintaxe da língua Trumai (isolada) e de seu sistema de marcação de caso'. MA thesis, Universidade Estadual de Campinas.
- 1999. 'A reference grammar of Trumai'. PhD dissertation, Rice University.
- Haas, Mary R. 1941. *Tunica*. New York: J. J. Augustin.
- Hajaú, Péter. 1963. *The Samoyed peoples and languages*. Bloomington: Indiana University Press and The Hague: Mouton.
- Hale, Horatio. 1846. 'The languages of Australia', pp. 366–424 of *Ethnography and philology* (Vol. 6 of *Reports of the United States Exploring Expedition, under the command of Charles Wilkes*). New York: Lea and Blanchard.
- Hale, Kenneth L. 1966. 'Kinship reflections on syntax: Some Australian languages', *Word* 22: 318–24.
- Halle, Morris. 1970. 'Is Kabardian a vowel-less language?' *Foundations of Language* 6: 95–103.
- Halpern, A. M. 1942. 'Yuma kinship terms', *American Anthropologist* 44: 425–41.
- Hansen, K. C., and L. E. 1978. *The core of Pintupi grammar*. Darwin: Summer Institute of Linguistics, Australian Aborigines Branch.
- Haspelmath, Martin. 1993. *A grammar of Lezgian*. Berlin: Mouton de Gruyter.
- 2006. 'Against markedness (and what to replace it with)', *Journal of Linguistics* 42: 25–70.
- Dryer, Matthew S., Gill David, and Comrie, Bernard. 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- Heath, Jeffrey. 1978. *Ngandi grammar, texts and dictionary*. Canberra: Australian Institute of Aboriginal Studies and New Jersey: Humanities Press.
- Heine, Bernd. 1982. 'African noun class systems', pp. 189–216 of *Apprehension: Das sprachliche Erfassen von Gegenständen*, Vol. 1: *Bereich und Ordnung der Phänomene*, edited by Hansjakob Seiler and Christian Lehmann. Tübingen: Gunter Narr.
- Henderson, E. J. A. 1951. 'The phonology of loanwords in some South-East Asian languages', *Transactions of the Philological Society* 1951: 131–58. [Reprinted as pp. 54–81 of Palmer 1970.]
- Hill, Archibald A. 1952. 'A note on primitive languages', *International Journal of American Linguistics* 18: 172–7. [Reprinted in abridged form on pp. 86–8 of Hymes 1964.]
- Hill, Deborah. 1992. 'Longgu grammar'. PhD thesis, Australian National University.
- Hinton, Leanne, Nichols, Johanna, and Ohala, John J. 1994. Editors of *Sound symbolism*. Cambridge: Cambridge University Press.
- Hock, Hans H. 1991. *Principles of historical linguistics*, 2nd edition. Berlin: Mouton de Gruyter.

- Hockett, Charles F. 1954. 'Two models of grammatical description', *Word* 10: 210–31. [Reprinted as pp. 386–99 of Joos 1958.]
- Holisky, Dee Ann. 1987. 'The case of the intransitive subject in Tsova-Tush (Batsbi)', *Lingua* 71: 103–32.
- Hopper, Paul J., and Thompson, Sandra A. 1980. 'Transitivity in grammar and discourse', *Language* 56: 251–99.
- Horne, Kibble M. 1966. *Language typology: 19th and 20th century views*. Washington, DC: Georgetown University Press.
- Hosokawa, K. 1991. 'The Yawuru language of West Kimberley: A meaning-based description'. PhD thesis, Australian National University.
- Hurch, Bernhard. 2005. Editor of *Studies on reduplication*. Berlin: Mouton de Gruyter.
- Hyman, Larry M. 1975. *Phonology: Theory and analysis*. New York: Holt, Rinehart and Winston.
- Hymes, Dell. 1964. Editor of *Language in culture and society: A reader in linguistics and anthropology*. New York: Harper and Row.
- Ikoru, Suanu. 1997. 'Akan: Grammar summary'. RCLT internal document.
- Jacobsen, William, Jr. 1979. 'Noun and verb in Nootka', pp. 83–155 of *The Victorian conference on North-Western languages*, edited by Barbara S. Efrat. British Columbia Provincial Museum, Heritage Record No. 4.
- Jake, Janice. 1978a. 'Why Dyrbal isn't ergative at all', pp. 179–92 of *Papers from the Fourteenth Regional Meeting of the Chicago Linguistic Society*, edited by Donka Farkas, Wesley M. Jacobson, and Karol W. Todrys. Chicago: Chicago Linguistic Society, University of Chicago.
- 1978b. 'Why Dyrbal isn't ergative at all', *Studies in the Linguistic Sciences* 8, 1: 97–110.
- Jakobson, Roman. 1962. 'Phoneme and phonology', pp. 231–3 of *Roman Jakobson: Selected writings*, Vol. 1: *Phonological studies*. The Hague: Mouton. [First published in the Second Supplementary volume of the *Czech Encyclopedia*, Prague, 1932.]
- 1990. *On language*, edited by Linda R. Waugh and Monique Monville-Burston. Cambridge, Mass.: Harvard University Press.
- Jespersen, Otto. 1909–49. *A modern English grammar, on historical principles*, Parts I–VII. Heidelberg: Winter (Parts I–IV), Copenhagen: Munksgaard (Parts V–VII).
- 1924. *The philosophy of grammar*. London: Allen and Unwin.
- Jones, Daniel, and Plaatje, Solomon T. 1916. *A Sechuana reader*. London: London University Press.
- Joos, Martin. 1958. Editor of *Readings in linguistics: The development of descriptive linguistics in America since 1925*, 2nd edition. New York: American Council of Learned Societies.
- Keenan, Edward L., and Comrie, Bernard. 1977. 'Noun phrase accessibility and universal grammar', *Linguistic Inquiry* 8: 63–99.
- — 1979. 'Data on the noun phrase accessibility hierarchy', *Language* 55: 333–51.
- and Ochs, Elinor. 1979. 'Becoming a competent speaker of Malagasy', pp. 113–58 of *Languages and their speakers*, edited by Timothy Shopen. Cambridge, Mass.: Winthrop.

- Keesing, Roger M. 1975. *Kin groups and social structures*. New York: Holt, Rinehart and Winston.
- Kennedy, Benjamin Hall. 1962. *The revised Latin primer*, edited and further revised by James Mountford. London: Longmans.
- Key, Harold. 1965. 'Some semantic functions of reduplication in various languages', *Anthropological Linguistics* 7, 3: 88–102.
- Khaidakov, S. M. 1963. 'Principy raspredelenija imen suščestvidel'nyx po grammatičeskam klassam v lakskom jazyke' ['The principles of the distribution of nouns across grammatical classes in Lak'], *Studia Caucasica* 1: 48–55.
- 1966. 'The dialect-divison of Lak', *Studia Caucasica* 2: 9–18.
- Kibrik, Aleksandr E., Kodzasov, Sandro V., and Muravyova, Irina A. 2004. *Language and folklore of the Alutor people*. Kyoto: Endangered Languages of the Pacific Rim.
- Kimball, Geoffrey D. 1991. *Koasati grammar*. Lincoln: University of Nebraska Press.
- Klamer, Marian. 1998. *A grammar of Kambara*. Berlin: Mouton de Gruyter.
- Krishnamurti, Bhadriraju. 2003. *The Dravidian languages*. Cambridge: Cambridge University Press.
- Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press.
- Kuipers, Aert H. 1960. *Phoneme and morpheme in Kabardian (Eastern Adyghe)*. The Hague: Mouton.
- Kulemeka, Andrew T. 1996. 'Determining the grammatical category of Chichewa ideophones', *Linguistic Analysis* 26: 84–116.
- Kumar, Ann, and Rose, Phil. 2000. 'Lexical evidence for early contact between Indonesian language and Japanese', *Oceanic Linguistics* 39: 219–55.
- Laanest, Arvo. 1975. *Sissejuhatus Läänemeresoome keeltesse* [Introduction to Balto-Finnic languages]. Tallinn: Eesti NSV Teaduste Akadeemia [Estonian Academy of Sciences]. [An abridged version published as pp. 5–121 of *Pribaltijsko-finskije jazyki. Osnovy Finno-Ugorskogo jazykoznanij: Pribaltijsko-finskije, Saamskij and Mordovskije jazyki* [Foundations of Finno-Ugric linguistics: Balto-Finnic, Saami, and Mordvin languages]. Moscow: Nauka, 1975.]
- Ladefoged, Peter. 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- and Maddieson, Ian. 1996. *The sounds of the world's languages*. Oxford: Blackwell.
- Lakoff, George. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Landar, Herbert. 1966. *Language and culture*. New York: Oxford University Press.
- Lang, Adrienne. 1975. *The semantics of classificatory verbs in Enga and other Papua New Guinea languages*. Canberra: Pacific Linguistics.
- Langacker, Ronald W. 1972. *Fundamentals of linguistic analysis*. New York: Harcourt Brace Jovanovich.
- Lawler, John M. 1975. 'On coming to terms in Acehese: The function of verbal disagreement', pp. 398–408 of *Papers from the parasession on functionalism*, edited by Robert E. Grossman et al. Chicago: Chicago Linguistic Society.
- 1977. 'A agrees with B in Acehese: A problem in relational grammar', pp. 219–48 of *Grammatical relations (Syntax and semantics, Vol. 8)*, edited by Peter Cole and Jerrold M. Sadock. New York: Academic Press.

- Lee, Jennifer. 1987. *Tiwi today: A study of language change in a contact situation*. Canberra: Pacific Linguistics.
- Leslau, Wolf. 1995. *Reference grammar of Amharic*. Wiesbaden: Harrassowitz.
- Lindau, Mona. 1985. 'The story of r', pp. 157–68 of *Phonetic linguistics*, edited by Victoria A. Fromkin. Orlando, Fla.: Academic Press.
- Loograman, Alfons. 1965. *Swahili grammar and syntax*. Pittsburgh: Duquesne University Press.
- Lorimer, D. L. R. 1938. *The Burushaski language*, Vol. 3: *Vocabularies and index*. Oslo: Aschehoug.
- Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- 1977. *Semantics*, 2 vols. Cambridge: Cambridge University Press.
- McGuckin, Catherine. 2002. 'Gapapaiwa', pp. 297–321 of *The Oceanic languages*, edited by John Lynch, Malcolm Ross, and Terry Crowley. Richmond, Surrey: Curzon.
- Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
- Majid, Asifa, Enfield, N. J., and van Staden, Miriam. 2006. Editors of *Parts of the body: Cross-linguistic classification*. Double issue of *Language Sciences*, 28, 2–3.
- Malinowski, Bronisław. 1935. *Coral gardens and their magic*, Vol. II: *The language of magic and gardening*. London: George Allen and Unwin.
- Marsh, James Lewis. 1976. 'The grammar of Mantjiltjara'. MA thesis, Arizona State University.
- Martin, Laura. 1986. '“Eskimo words for snow”: A case study in the genesis and decay of an anthropological example', *American Anthropologist* 88: 418–23.
- Matisoff, James A. 1973. *The grammar of Lahu*. Berkeley and Los Angeles: University of California Press.
- Matthews, P. H. 1981. *Syntax*. Cambridge: Cambridge University Press.
- 1991. *Morphology*, 2nd edition. Cambridge: Cambridge University Press.
- 1997. *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.
- Maybury-Lewis, David. 1968. *Akweà-Shavente society*. Oxford: Clarendon Press.
- Meillet, Antoine. 1925. *La Méthode comparative en linguistique historique*. Oslo: Instituttet for Sammenlignende Kulturforskning. [English translation by Gordon B. Ford, Jr., *The comparative method in historical linguistics*. Paris: Champion, 1967.]
- 1926. *Linguistique historique et linguistique générale*. Paris: Champion.
- Merlan, Francesca C. 1994. *A grammar of Wardaman, a language of the Northern Territory of Australia*. Berlin: Mouton de Gruyter.
- Miller, Jim. 1995. 'Does spoken language have sentences?', pp. 136–47 of *Grammar and meaning: Essays in honour of Sir John Lyons*, edited by F. R. Palmer. Cambridge: Cambridge University Press.
- Mitchell, T. F. 1962. *Colloquial Arabic, the living language of Egypt* (Teach Yourself Books). London: English Universities Press.
- Mithun, Marianne. 1984. 'The evolution of noun incorporation', *Language* 60: 847–94.
- 1987. 'Is basic word order universal?', pp. 281–328 of *Coherence and grounding in discourse*, edited by Russell S. Tomlin. Amsterdam: John Benjamins.
- Morphy, Frances. 1983. 'Djapu, a Yolngu dialect', pp. 1–188 of Dixon and Blake 1983.



- Murray, L. 1795. *English grammar*. York: Wilson, Spence and Mawman. [Fourth edition in 1798.]
- Newman, John. 2002. Editor of *The linguistics of sitting, standing and lying*. Amsterdam: John Benjamins.
- Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven: Yale University Press.
- Nida, Eugene A. 1946. *Morphology: The descriptive analysis of words*. Ann Arbor: University of Michigan Press.
- 1948. 'The identification of morphemes', *Language* 24: 414–41.
- 1949. *Morphology: The descriptive analysis of words*, 2nd edition. Ann Arbor: University of Michigan Press.
- Noonan, Michael. 1985. 'Complementation', pp. 42–140 of *Language typology and syntactic description*, Vol. 2: *Complex constructions*. Edited by Timothy Shopen. Cambridge: Cambridge University Press
- 1992. *A grammar of Lango*. Berlin: Mouton de Gruyter.
- Ohala, John. 1984. 'An ethnological perspective on common cross-language utilization of F<sub>0</sub> of voice', *Phonetica* 41: 1–16.
- Onishi, Masayuki. 1994. 'A grammar of Motuna (Bougainville, Papua New Guinea)'. Ph.D. thesis, Australian National University.
- 1997. 'Bengali: Grammar summary'. RCLT internal document.
- Osborne, C. R. 1974. *The Tiwi language*. Canberra: Australian Institute of Aboriginal Studies.
- Palmer, F. R. 1970. Editor of *Prosodic analysis*. London: Oxford University Press.
- 1986. *Mood and modality*. Cambridge: Cambridge University Press.
- Palmer, H. E. 1924. *A grammar of spoken English on a strictly phonetic basis*. Cambridge: Heffer. [Second edition in 1939.]
- and Blandford, F. G. 1969. *A grammar of spoken English*, 3rd edition, revised and rewritten by Roger Kingdon. Cambridge: Heffer.
- Payne, David L. 1981. *Phonology and morphology of Axininca Campa*. Dallas: SIL and University of Texas at Arlington.
- Payne, Doris. 1990. *The pragmatics of word order: Typological dimensions of verb initial languages*. Berlin: Mouton de Gruyter.
- Payne, Thomas E. 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.
- Pesetsky, David. 1982. 'Paths and categories'. Ph.D. dissertation, MIT.
- Pike, Kenneth L. 1947a. 'Grammatical prerequisites to phonemic analysis', *Word* 3: 155–72.
- 1947b. *Phonemics: A technique for reducing languages to writing*. Ann Arbor: University of Michigan Press.
- 1948. *Tone languages: A technique for determining the number and types of pitch contrasts in a language, with studies in tonemic substitution and fusion*. Ann Arbor: University of Michigan Press.
- 1952. 'More on grammatical prerequisites', *Word* 8: 108–21.

- Plank, Frans. 1995. '(Re-)introducing Suffixaufnahme', pp. 3–110 of *Double case: Agreement by Suffixaufnahme*, edited by Frans Plank. New York: Oxford University Press.
- Platt, John T. 1972. *An outline grammar of the Gugada dialect, South Australia*. Canberra: Australian Institute of Aboriginal Studies.
- Pullum, Geoffrey K. 1988. 'Topic . . . comment: Citation etiquette beyond thunderdome', *Natural Language and Linguistic Theory* 6: 579–88. [Reprinted as pp. 147–58 of Pullum 1991.]
- 1989. 'Topic . . . comment: The great Eskimo vocabulary hoax', *Natural Language and Linguistic Theory* 7: 75–81. [Reprinted as pp. 159–71 of Pullum 1991.]
- 1991. *The Great Eskimo Vocabulary Hoax, and other irreverent essays on the study of language*. Chicago: University of Chicago Press.
- Rankin, Robert, Boyle, John, Graczyk, Randolph, and Koontz, John. 2002. 'Synchronic and diachronic perspectives on "Word" in Siouan', pp. 180–204 of Dixon and Aikhenvald 2002.
- Refsing, Kirsten. 1986. *The Ainu language: The morphology and syntax of the Shizunai dialect*. Aarhus: Aarhus University Press.
- Reh, Mechthild. 1996. *Anywa language: Description and internal reconstruction*. Cologne: Rüdiger Köffe.
- Reichard, Gladys A., Jakobson, Roman, and Werth, Elizabeth. 1949. 'Language and synesthesia', *Word* 5: 224–33.
- Reichmann, W. J. 1964. *The use and abuse of statistics*. Harmondsworth: Penguin.
- Roberts, John R. 1987. *Amele* (Croom Helm Descriptive Grammars). London: Croom Helm.
- Robins, R. H. 1958. *The Yurok language: Grammar, texts, lexicon*. Berkeley and Los Angeles: University of California Press.
- Roth, Walter E. 1897. *Ethnological studies among the north-west-central Queensland Aborigines*. Brisbane: Government Printer and London: Queensland Agent-General's Office.
- Rubino, Carl. 2005. 'Reduplication: Forms, functions and distribution', pp. 11–29 of Hutch 2005.
- Rumsey, Alan. 1982. *An intra-sentence grammar of Ungarinjin, north-western Australia*. Canberra: Pacific Linguistics.
- 2000. 'Bunuba', pp. 34–152 of Dixon and Blake 2000.
- Sapir, Edward. 1917. Review of 'Het identificeerend karakter der possessieve flexie in talen van Noord-Amerika' by C. C. Uhlenbeck, *International Journal of American Linguistics*, 1: 86–90. [Reprinted as pp. 75–80 of *The collected works of Edward Sapir*, Vol. 5: *American Indian Languages* 1, volume editor William Bright. Berlin: Mouton de Gruyter, 1990.]
- 1921. *Language*. New York: Harcourt Brace.
- 1922. 'The Takelma language of south-western Oregon', pp. 1–296 of *Handbook of American Indian languages*, Vol. 2. Bulletin 40, Bureau of American Ethnology, Smithsonian Institution, Washington.
- 1925. 'Sound patterns in language', *Language* 1: 37–51. [Reprinted as pp. 33–45 of Sapir 1949, and as pp. 19–25 of Joos 1958.]

- Sapir, Edward. 1930. 'Southern Paiute: A Shoshonean language', *Proceedings of the American Academy for Arts and Sciences*, 65, 1: 1–296.
- 1933. 'La Réalité psychologique des phonèmes', *Journal de psychologie normale et pathologique* 30: 247–65. [English version, 'The psychological reality of phonemes', pp. 46–60 of Sapir 1949.]
- 1949. *Selected writings of Edward Sapir in language, culture and personality*, edited by David G. Mandelbaum. Berkeley and Los Angeles: University of California Press.
- Saussure, Ferdinand de. 1916. *Cours de linguistique générale*, publié par Charles Bally . . . et Albert Sechehaye . . . avec la collaboration de Albert Riedlinger. Lausanne. [Available in two English translations, as *Course in general linguistics*.]
- Shopen, Timothy. 1985. Editor of *Language typology and syntactic description*, 3 vols. Cambridge: Cambridge University Press.
- 2007. Editor of *Language typology and syntactic description*, 3 vols., 2nd edition. Cambridge: Cambridge University Press.
- Siewierska, Anna. 1984. *The passive: A comparative linguistic analysis*. London: Croom Helm.
- Silverstein, Michael. 1976. 'Hierarchy of features and ergativity', pp. 112–71 of *Grammatical categories in Australian languages*, edited by R. M. W. Dixon. Canberra: Australian Institute of Aboriginal Studies and New Jersey: Humanities Press. [Republished as pp. 163–232 of *Features and projections*, edited by Pieter C. Muysken and Henk van Riemsdijk. Dordrecht: Foris, 1986.]
- Sinclair, John. 2001. Founder editor-in-chief of *Collins Cobuild English dictionary for advanced learners*. Glasgow: HarperCollins.
- Skorik, P. J. 1961. *Grammatika čukotskogo jazyka*. Tom 1 [*Grammar of the Chukchi language*, Vol. 1]. Moscow and Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Sneddon, James Neil. 1996. *Indonesian reference grammar*. Sydney: Allen and Unwin.
- Sohn, Ho-min. 1994. *Korean*. London: Routledge.
- Stankiewicz, Edward. 1972. *A Baudouin de Courtenay anthology*, translated and edited with an introduction by Edward Stankiewicz. Bloomington: Indiana University Press.
- Stenson, Nancy. 1981. *Studies in Irish syntax*. Tübingen: Gunter Narr.
- Swadesh, Morris. 1934. 'The phonemic principle', *Language* 10: 117–29. [Reprinted as pp. 32–7 of Joos 1958.]
- 1938. 'Nootka internal syntax', *International Journal of American Linguistics* 9: 77–102.
- and Voegelin, Charles F. 1939. 'A problem in phonological alternation', *Language* 15: 1–10.
- Sweet, Henry. 1891. *A new English grammar, logical and historical*, Vol. 1. Oxford: Clarendon Press.
- Tamura, Suzuko. 2000. *The Ainu language*. Tokyo: Sanseido.
- Taumoefolau, Malenaite. 2002. 'Stress in Tongan', *MIT Working Papers in Linguistics* 44: 341–54.
- Tepljashina, T. I., and Lytkin, V. I. 1976. 'Permskije jazyki' ['Permic languages'], pp. 97–228 of *Osnovy Finno-Ugorskogo jazykoznanija; Marijskij, Permskije i Ugorskije*

- jazyki* [Foundations of Finno-Ugric Linguistics; Mari, Permic, and Ugric languages]. Moscow: Nauka.
- Thomsen, Marie-Louise. 1984. *The Sumerian language: An introduction to its history and grammatical structure*. Copenhagen: Akademisk Forlag.
- Toulmin, Stephen E. 1984. 'Science, philosophy of', p. 382 of *Encyclopaedia Britannica* 1984 edition, Macropedia Volume 16.
- Trubetzkoy, Nicolai S. 1939. *Grundzüge der Phonologie* (Travaux du Cercle Linguistique de Prague, No. 7.) [English translation by Christine A. M. Baltaxe. *Principles of phonology*. Berkeley and Los Angeles: University of California Press, 1969.]
- Vászolyi, Eric G. 1979. *Teach yourself Wangkatja: An introduction to the Western Desert language (Cundeelee dialect)*. Perth: Mount Lawley College.
- Vitale, Anthony J. 1981. *Swahili syntax*. Dordrecht: Foris.
- Voeltz, F. K. Erhard, and Kilian-Hatz, Christa. 2001. Editors of *Ideophones*. Amsterdam: John Benjamins.
- Watters, David E. 2002. *A grammar of Kham*. Cambridge: Cambridge University Press.
- Westermann, Diedrich. 1947. 'Pluralbildung und Nominalklassen in einigen afrikanischen Sprachen', *Abhandlungen der Deutschen Akademie der Wissenschaften zu Berlin*. 1945/6. Phil.-hist. Klasse No 1. Berlin.
- Whorf, Benjamin Lee. 1938. 'Language: Plan and conception of arrangement'. MS. [Published as pp. 124–33 of *Language, thought and reality: Selected writings of Benjamin Lee Whorf*, edited by John B. Carroll. Cambridge: The Technology Press of Massachusetts Institute of Technology and New York: John Wiley, 1956.]
- Wierzbicka, Anna. 1972. *Semantic primitives*. Frankfurt: Athenäum.
- 1980. *Lingua mentalis*. New York: Academic Press.
- Wordick, F. J. F. 1982. *The Yindjibarndi language*. Canberra: Pacific Linguistics.
- Zeitoun, Elizabeth. 2007. *A grammar of Mantaoran (Rukai)*. Taipei: Institute of Linguistics, Academia Sinica.

# Author Index

The indices cover both volume 1 (shown by 1:) and volume 2 (shown by 2:), followed by page numbers.

- Abbott, Miriam 1: 91; 2: 93–4, 400  
Abercrombie, David 1: 287  
Abraham, R. C. 2: 114  
Adelaar, Willem 2: 181, 184, 204, 286  
Aikhenvald, Alexandra Y. 1: 87, 180–1, 236, 238, 241, 255; 2: 31, 406  
    on cases 1: 87; 2: 55, 120, 221  
    on clitics 1: 222; 2: 21, 23  
    on evidentiality 1: 56, 87, 180, 260, 319; 2: 113, 244, 259–60, 419  
    on nominal classification 1: 158, 180, 238, 263, 318; 2: 243, 246, 259, 277, 368  
    See also entries in the Index of languages on Baniwa of Içana, Bare, Manambu, Tariana, Tucano, Warekena and Zekkara.  
Aissen, Judith L. 1: 327  
Aitken, Percy 2: 271  
Akmajian, Adrian 2: 204  
Akpati, Elizabeth 2: 278  
Alford, Denny Keith 2: 278  
Allen, W. S. 1: 279; 2: 10, 12, 214, 259, 391, 305  
Alpher, Barry 2: 59, 97, 104  
Alvarez, José 2: 286  
Amberber, Mengistu 1: 180; 2: 344, 349  
Ameka, Felix 1: 54; 2: 36, 254, 280, 285, 288  
Andersen, Torben 2: 142  
Anderson, Stephen 2: 204, 213, 242–5  
Anderson, Victoria 1: 180  
Andrade, Manuel J. 1: 179; 2: 60  
Andrews, Avery D. 2: 368  
Andrews, Edna 1: 241  
Anschutz, A. 2: 312  
Arms, David G. 2: 108, 145  
Arnott, D. W. 2: 86  
Asher, R. E. 2: 91, 174, 180, 233, 251, 336, 357  
Ashton, E. O. 1: 85  
Aspinion, Robert 1: 54; 2: 89  
Atatürk, Kemal 1: 20, 55  
Austin, Peter 2: 157, 169, 181  
Avrorin, V. A. 2: 285  
Baker-Shenk, Charlotte 2: 339  
Ball, Martin J. 2: 12  
Bally, Charles 2: 5  
Bani, Ephraim 1: 55  
Barentsen, Adrian 2: 372  
Bargery, G. P. 2: 114  
Barnes, Janet 1: 55–6, 181  
Barshi, Immanuel 2: 311  
Baudouin de Courtenay, Jan 1: 180, 209  
Bauer, Laurie 1: 185, 241; 2: 61  
Bazell, C. E. 1: 241–2, 261; 2: 12  
Bazin, Louis 1: 55  
Beaumont, Clive H. 2: 80  
Beck, David 2: 84, 90  
Benjamin, Carmen 2: 176  
Benjamin, Geoffrey 2: 78  
Benveniste, Emile 1: 88, 241; 2: 174, 188, 203, 301, 312  
Bergslund, Knut 1: 241  
Berlin, Brent 1: 90, 256  
Besnier, Niko 2: 322, 349, 351  
Bhat, D. N. S. 2: 88, 200, 257, 260, 358, 368  
Bhatia, Tej K. 2: 158, 303, 358  
Bickel, Balthasar 2: 333, 349, 363  
Bing, Janet M. 2: 76  
Blackwell, Aleka A. 2: 76  
Blake, Barry J. 1: 308; 2: 368  
Blandford, F. G. 1: 90  
Bloch, Bernard 1: 288  
Bloomfield, Leonard 1: 85–88, 328; 2: 2, 10–11, 18–19, 32, 35, 37  
Boas, Franz 1: 84; 2: 64, 198, 243–4, 278–9, 281, 282, 288  
Bochner, H. 2: 17  
Bolinger, Dwight 1: 89–90; 2: 71  
Bontkes, Carolyn 2: 409  
Borer, Hagit 2: 325  
Borgman, Donald M. 2: 193  
Bowern, Claire 1: 241  
Boxwell, Maurice 2: 198  
Boyle, John 1: 353; 2: 456  
Breen, J. G. 2: 291  
Bright, William 2: 64  
Broschart, Jürgen 2: 54  
Brown, Cecil H. 1: 262

- Brown, D. Richard 1: 181  
 Brown, Penelope 2: 343, 364  
 Browne, W. 2: 336  
 Bruce, Les 2: 86  
 Bugenhagen, Robert D. 2: 231  
 Burgess, E. 2: 205  
 Burrige, Kate 1: 18, 55; 2: 393, 405  
 Burzio, Luigi 1: 91  
 Buse, J. E. 2: 70  
 Butt, John 2: 176  
 Bužarovska, Eleni 2: 361  
 Byarushengo, Ernest R. 2: 157  
 Bybee, Joan 1: 259, 262; 2: 301
- Canger, U. 1: 326  
 Carlin, Eithne B. 2: 197, 311  
 Carlson, Robert 1: 55; 2: 236, 251, 343  
 Carpenter, Kathie 2: 348  
 Cawdrey, Robert 1: 179  
 Ceria, Verónica G. 2: 243  
 Černý, Václav 2: 158  
 Chafe, Wallace 2: 205  
 Chalker, Sylvia 1: 91  
 Chao, Yuen-Ren 1: 88, 287; 2: 4, 30–1  
 Chappell, Hilary 2: 233, 310  
 Charney, Jean Ormsbee 1: 180  
 Charters, Helen 2: 330  
 Churchill, Winston S. 1: 64, 66  
 Churchward, C. Maxwell 1: 90, 179–80;  
 2: 25, 54, 233  
 Clark, Eve V. 2: 61  
 Clark, Herbert H. 2: 61  
 Clements, George N. 2: 253  
 Cokley, Dennis 2: 339  
 Cole, Peter 2: 85, 333  
 Comrie, Bernard 1: 55, 86; 2: 157–8, 170,  
 254, 259  
 on relative clauses 1: 89; 2: 320, 326, 330,  
 337, 352, 369, 368  
 Conklin, Harold C. 2: 196, 198, 258  
 Cook, Eung-Do 1: 179, 220–1; 2: 242  
 Cook, Kenneth W. 2: 126  
 Corbett Greville G. 1: 87, 180; 2: 85, 368  
 Corominas, Joan 2: 183  
 Coulmas, Florian 2: 231, 233  
 Craig, Colette Grinevald 1: 55, 91, 179;  
 2: 292, 304, 390, 404, 418  
 Crapo, Richley H. 2: 271  
 Crazzolaro, J. P. 2: 177  
 Creissels, Denis 2: 212, 259  
 Crisp, Simon 2: 158, 213  
 Cristofaro, Sonia 2: 369  
 Croft, William 1: 241  
 Crowell, Thomas H. 2: 81
- Crowley, Terry 2: 87, 363  
 Culy, Christopher 2: 253  
 Cummins, George M. III 2: 236  
 Curnow, Timothy J. 2: 87, 178, 231, 274,  
 300–1, 379, 403  
 Cysouw, Michael 2: 258  
 Cytowic, Richard E. 1: 90  
 Czaykowska-Higgins, Ewa 2: 72, 77
- Daguman, Josephine S. 2: 114, 165, 237  
 Dahl, Otto Christian 1: 54  
 Dahlstrom, Amy 2: 251  
 Dai, John Xiang-Ling 2: 35  
 Dallet, J. B. 2: 167  
 Davidson, Matthew 2: 61, 402  
 Davies, John 2: 331, 344, 353  
 Day, Christopher 1: 179  
 Deane, P. 2: 312  
 DeLancey, Scott 2: 334, 363  
 Deloria, Ella 1: 84; 2: 199, 282–3, 288  
 Demers, Richard A. 2: 38  
 Dench, Alan C. 1: 250, 267, 276; 2: 8, 157  
 Derbyshire, Desmond C. 2: 92, 271, 316,  
 341  
 Deutscher, Guy 2: 379, 385, 390, 393, 405–6,  
 408–12, 418, 420  
 Devitt, Dan 2: 188  
 Di Sciullo, Anna-Maria 2: 35  
 Dickens, Patrick 2: 260  
 Diessel, Holger 1: 90; 2: 233, 236, 238, 242,  
 245, 259–60  
 Diffloth, Gerard 1: 90  
 Dik, Simon C. 2: 418  
 Dimmendaal, Gerrit J. 2: 97, 419  
 Dirr, A. 1: 91  
 Doke, C. M. 1: 302  
 Donaldson, Tamsin 1: 180; 2: 213  
 Donohue, Mark 2: 61, 82  
 Douglas, Wilfrid H. 1: 262; 2: 291  
 Dowling, Bruce T. 2: 368  
 Dresher, Bezalel E. 2: 9  
 Dryer, Matthew S. 1: 348; 2: 157  
 Du Bois, John W. 1: 89  
 Dunn, Michael John 1: 181  
 Duponceau, Peter S. 1: 241, 260  
 Duranti, Alessandro 2: 157  
 Durie, Mark 1: 330; 2: 205, 258  
 Dutton, Tom E. 2: 136, 194
- Eades, Diana 1: 90; 2: 113  
 Eckert, Paul 1: 262  
 Elkin, A. P. 1: 317  
 Emenanjo, E. Nolue 2: 268, 340  
 Emeneau, M. B. 2: 250

- Enfield, N. J. 1: 55, 83; 2: 82, 113, 236, 259, 329, 338, 347  
 England, Nora C. 1: 82, 181; 2: 91  
 Englebretson, Robert 2: 420  
 Enrico, John 2: 140  
 Escalante, Fernando 2: 211  
 Evans, Nicholas D. 1: 55, 82, 179; 2: 36, 44–5, 49, 61, 208, 311  
  
 Faverey, Margot 2: 176  
 Feeling, Durbin 2: 77  
 Fenyvesi, Anna 2: 11, 346, 353  
 Ferrar, H. 2: 238  
 Ferris, Connor 2: 114  
 Filimonova, Elena 2: 258  
 Firth, J. R. 2: 35  
 Fleck, David W. 1: 262; 2: 204, 410, 419  
 Fodor, István 1: 55  
 Foley, William A. 1: 55, 63, 82, 181; 2: 9, 23–4, 169, 178, 199, 353, 393  
 Forchheimer, Paul 2: 258  
 Ford, Lysbeth J. 2: 87, 90, 215  
 Fortescue, Michael 2: 243, 275, 304  
 Fortune, Reo 2: 85  
 Fox, Barbara A. 2: 322  
 Frachtenberg, Leo J. 2: 37, 42, 64  
 Frajzyngier, Zygmunt 1: 179; 2: 79, 82–3, 169, 231, 330, 342  
 Franklin, Joice 2: 10  
 Franklin, Karl J. 2: 10  
 Freeland, L. S. 1: 84  
 Fries, Charles C. 1: 89, 283  
  
 Gak, V. G. 2: 35  
 Galloway, Brent D. 2: 61  
 Gamal-Eldin, Saad 2: 292, 353  
 Gardiner, Alan H. 2: 183  
 Gary, Judith Olmsted 2: 157, 292, 353  
 Genetti, Carol 2: 95, 165, 175, 258  
     on relative clauses 2: 319, 334, 342, 357, 363  
     on complementation 2: 390, 402, 419  
 George Madugu, Isaac S. 2: 74, 92  
 Gerzenstein, Ana 2: 286  
 Gill David 1: 348; 2: 445  
 Giridhar, P. P. 2: 42, 54, 267  
 Givón, Talmy 2: 63, 301, 338, 349, 369  
 Glasgow, Kathleen 2: 157  
 Glass, Ameer 1: 262  
 Goddard, Cliff 1: 262, 308  
 Gorbet, Larry 2: 334  
 Gordon, Lynn 1: 54, 181  
 Gordon, Raymond G, Jr. 1: xiii, 347  
  
 Graczyk, Randolph 1: 353; 2: 456  
 Gragg, Gene 2: 227  
 Gray, Louis H. 1: 301; 2: 9, 201  
 Green, Diana 2: 244  
 Greenbaum, Sidney 2: 245  
 Greenberg, Joseph H. 1: 65, 72–3, 90–1, 241, 254; 2: 258  
 Gregor, Douglas B. 2: 12  
 Grimes, Barbara 1: 90  
 Grimes, Charles E. 2: 270  
 Guedes, Marymarcia 2: 8  
 Guillaume, Antoine 2: 165, 181  
 Guirardello, Raquel 1: 179  
 Güldemann, Tom 2: 253, 419  
  
 Haag, Marcia 2: 61  
 Haas, Mary R. 1: 84; 2: 91, 201, 221–2, 258  
 Hackett, Dorothy 1: 262  
 Hagège, Claude 2: 81  
 Haiman, John 2: 84, 91, 199, 243, 273, 285–6, 293, 368  
 Hajdú, Péter 1: 110; 2: 42  
 Hale, Horatio 1: 197  
 Hale, Kenneth 1: 17, 55; 2: 358, 362  
 Hall, T. Alan 2: 27  
 Halle, Morris 1: 288  
 Halliday, M. A. K. 2: 260  
 Halpern, A. M. 1: 55  
 Hansen, K. C. 1: 262  
 Hansen, L. E. 1: 262  
 Harada, S. I. 2: 350  
 Harkins, Jean 1: 308  
 Harris, Alice C. 2: 168  
 Harris, Zellig S. 2: 4  
 Harrison, Sheldon P. 2: 101, 390  
 Hasan, Ruqaiya 2: 260  
 Haspelmath, Martin 1: 180, 241, 260, 262; 2: 148, 258, 347  
 Haviland, John 1: 90, 308; 2: 158, 260, 391  
 Hayward, Richard J. 2: 170, 237, 243  
 Hazlewood, David 2: 25  
 Heath, Jeffrey 1: 181; 2: 89, 201, 209  
 Hébert, Yvonne M. 2: 38, 61  
 Heffernan, J. A. 2: 211  
 Heine, Bernd 1: 180; 2: 188, 301–2, 310, 369, 421  
 Hellwig, Birgit 2: 419  
 Henadeerage, Kumara 2: 326  
 Henderson, E. J. A. 1: 283  
 Henderson, John 2: 9, 23, 25  
 Hengeveld, Kees 2: 188, 418  
 Hercus, Luise A. 2: 182, 194  
 Hess, Thom 2: 38, 51, 53, 56

- Hewitt, B. G. 2: 326, 344  
 Hildebrandt, Kristine 2: 95  
 Hill, Archibald A. 1: 65, 90  
 Hill, Deborah 1: 55; 2: 252, 386  
 Himmelmann, Nikolaus P. 2: 238  
 Hinds, John 2: 259  
 Hinton, Leanne 1: 90  
 Hock, Hans H. 1: 90; 2: 137  
 Hockett, Charles F. 1: 146, 180; 2: 4, 69  
 Hoffmann, Carl 2: 197  
 Hoffmann, J. 2: 43–4  
 Hoiijer, Harry 2: 60, 215  
 Holisky, Dee Ann 1: 91; 2: 121, 141, 157  
 Hopper, Paul J. 1: 89; 2: 143, 149, 156, 409  
 Horne, Kibble M. 1: 241  
 Hosokawa, K. 1: 56  
 Hovdhaugen, Even 2: 45  
 Huang, Chenglong 2: 79, 82–3, 114, 165, 173  
 Huber, Brigitte 2: 334  
 Huddleston, Rodney 1: 185; 2: 61, 360  
 Hudson, Joyce 1: 262; 2: 8  
 Hurch, Bernhard 1: 180, 262  
 Hyman, Larry M. 1: 288; 2: 157, 254, 278,  
     311  
 Hymes, Dell 1: 343, 348–9  
 Hyslop, Catriona 2: 45, 82, 113, 406, 408  
  
 Ikoro, Suanu 1: 55; 2: 136, 178, 192, 254, 299  
 Imedadze, Natela 2: 251  
 Ingram, David 2: 258  
 Isačenko, Alexander V. 2: 299, 301  
  
 Jacob, Judith M. 2: 231  
 Jacobsen, William, Jr. 1: 179; 2: 43, 51, 53,  
     258  
 Jaggard, Philip J. 2: 352  
 Jake, Janice 1: 66, 90  
 Jakobson, Roman 1: 70, 88, 90, 241, 287  
 Jany, Carmen 2: 405, 410  
 Jarkey, Nerida 2: 379, 385–6, 393, 398, 404  
 Jastrow, Otto 2: 11  
 Jauncey, Dorothy 2: 217, 299  
 Jefferson, Gail 2: 36  
 Jelinek, Eloise 2: 38  
 Jerry, Bessie 1: 20  
 Jespersen, Otto 1: 87–8; 2: 68, 114, 250, 360  
 Jeyapul, V. Y. 2: 269  
 Johanson, Lars 2: 88  
 Johansson, Stig 2: 71, 312  
 Johns, Brenda 2: 176  
 Johnson, Steve 2: 213  
 Jones, Daniel 1: 273, 326; 2: 28  
 Jones, Morris 2: 302  
  
 Joos, Martin 1: 344, 349, 353–4  
 Joseph, Brian D. 2: 10, 26, 178, 201, 390  
 Jungbluth, Konstanze 2: 260  
  
 Kachru, Yamuna 2: 174, 177  
 Kautzsch, E. 2: 369  
 Kay, Paul 1: 256  
 Keenan, Edward L. 1: 55; 2: 157, 204, 242–5  
     on relative clauses 1: 89; 2: 320, 324, 326,  
     330, 336, 350, 361  
 Keesing, Roger M. 1: 262  
 Kenesei, István 2: 11, 346, 353  
 Key, Harold 1: 180, 262  
 Khaidakov, S. M. 1: 55; 2: 242  
 Kibrik, Aleksandr E. 1: 262  
 Kilian-Hatz, Christa 1: 362  
 Kilby, David 2: 158  
 Kimball, Geoffrey D. 1: 82, 180; 2: 83, 200,  
     233, 316  
 Kinkade, M. Dale. 2: 38  
 Kiparsky, Paul 2: 158  
 Kita, Sotar 2: 260  
 Klamer, Marian 1: 55; 2: 352  
 Kleinhenz, Ursula 2: 27  
 Kleinschmidt, S. 2: 44  
 Klokeid, Terry 1: 55  
 Kockelman, Paul 2: 386  
 Kodzasov, Sandro V. 1: 262  
 Koehn, Edward 2: 274  
 Koehn, Sally 2: 274  
 König, Christa 2: 120  
 Koontz, John 1: 353; 2: 456  
 Kornfilt, Jaklin 2: 271  
 Koshal, Sanyukta 2: 291  
 Krámský, Jiří 2: 34  
 Kratochvíl, František 2: 240  
 Krishnamurti, Bh. 1: 308; 2: 324  
 Kroeber, Paul D. 2: 61, 224  
 Kroeger, Paul 2: 56  
 Kruspe, Nicole 1: 82; 2: 187  
 Kruszewski, Nikolaï 1: 180:  
 Kuipers, Aert H. 1: 288; 2: 61, 401  
 Kulemeka, Andrew T. 1: 302  
 Kullavanijaya, Pranee 2: 364  
 Kumar, Ann 1: 90  
 Kumar, Rajesh 2: 317  
 Kumari, T. C. 2: 91, 336, 357  
 Kuno, Susumu 2: 326, 350, 364  
 Kuroda, S.-Y. 2: 337, 360  
 Kuteva, Tania 2: 188, 301, 369, 421  
  
 Laanest, Arvo 1: 181  
 Ladefoged, Peter 1: 87, 274, 258



- Lakoff, George 1: 55  
 Landar, Herbert 1: 240  
 Lang, Adrienne 1: 262  
 Langacker, Ronald W. 1: 56  
 Langendoen, D. Terence 2: 178  
 LaPolla, Randy J. 2: 79, 82–3, 114, 165, 258  
 Launey, Michel 2: 61  
 Lawler, John M. 1: 330  
 Lee, Jennifer 1: 55, 205; 2: 199  
 Leech, Geoffrey 2: 456  
 Leeding, Velma J. 2: 291  
 Leer, Jeff 2: 271  
 Lehiste, Ilse 2: 9  
 Lehmann, Christian 2: 368  
 Lehmann, Thomas 2: 357  
 Leslau, Wolf 1: 181; 2: 364  
 Levi, Judith N. 2: 455  
 Levin, Beth 2: 158  
 Levinson, Stephen C. 2: 249, 261  
 Levy, Paulette 2: 77, 80, 84, 88  
 Lewis, Geoffrey 2: 64  
 Li Xinyin 2: 245  
 Li, Charles N. 2: 69, 183, 268, 331, 337, 342  
 Li, Feng 2: 183  
 Lichtenberk, Frantisek 2: 205, 259  
 Liddell, Scott K. 2: 339  
 Liddicoat, Tony 2: 273  
 Lindau, Mona 1: 288  
 Lindström, Eva 2: 238  
 Longacre, Robert E. 2: 34  
 Loogman, Alfons 1: 85  
 Lorimer, D. L. R. 1: 90  
 Lynch, John 1: 351; 2: 277  
 Lyons, John 1: 86, 234; 2: 5, 18, 69, 190, 229, 246, 261  
 Lysvåg, Per 2: 71  
 Lytkin, V. I. 1: 181  
  
 Maddieson, Ian 1: 87, 262, 274, 288  
 Mahajan, Anoop 2: 358  
 Mahootian, Shahrzad 2: 352, 360  
 Majid, Asifa 1: 55  
 Maling, J. 2: 148, 158  
 Malinowski, Bronislaw 1: 291, 298  
 Mallinson, Graham 2: 201, 353, 360, 368  
 Malotki, Eckehart 2: 340  
 Mamiani, Luis Vincencio 2: 277  
 Maring, Joel M. 2: 217, 258  
 Marsh, James Lewis 1: 262  
 Martin, Jack 2: 280, 287  
  
 Martin, Laura 1: 65, 90  
 Masica, Colin 2: 357  
 Matisoff, James A. 1: 64, 81; 2: 243, 363  
 Matsumoto, Yoshiko 2: 349  
 Matthews, P. H. 1: 86, 241; 2: 1–2, 7, 16, 19, 160, 369  
 Matthews, Stephen 2: 180, 192  
 Maybury-Lewis, David 1: 328  
 Mazaudon, Martine 2: 334, 363  
 McCarthy, John J. 2: 18  
 McConvell, Patrick 2: 197  
 McGregor, William 2: 310  
 McGuckin, Catherine 1: 54; 2: 281–2, 290  
 McKay, Graham R. 2: 196, 293–4, 359  
 Meillet, Antoine 1: 25, 88; 2: 9  
 Meinhof, Carl 2: 259  
 Meira, Sérgio 2: 92, 311  
 Mejlanova, U. A. 2: 158  
 Merlan, Francesca C. 1: 82, 277; 2: 198, 291  
 Meyerstein, Zlata P. 2: 236  
 Michael, Ian 2: 247  
 Migeod, F. W. H. 2: 114  
 Mikkelsen, Line 2: 187  
 Miller, Amy 2: 324  
 Miller, Jim 1: 91  
 Miller, Wick R. 2: 244, 258  
 Milner, G. B. 2: 24, 37, 65, 108, 112  
 Mitchell, T. F. 1: 180  
 Mithun, Marianne 1: 89, 91; 2: 60, 140, 198, 204–5, 257, 284, 311  
     on demonstratives 2: 239, 245  
     on Tuscarora 2: 239, 293, 300  
 Moliner, Maris 2: 183  
 Monserrat, Ruth 2: 268  
 Morgan, William, Sr. 2: 343  
 Morphy, Frances 1: 288  
 Mosel, Ulrike 2: 45, 61  
 Mous, Maarten 2: 200, 216, 221, 242  
 Mpaayei, J. T. O. 2: 86  
 Munro, Pamela 2: 167, 333, 337, 350, 352  
 Muravyova, Irina A. 1: 262  
 Murray, L. 1: 80  
  
 Nababan, P. W. J. 2: 82  
 Nagaraja, K. S. 2: 42, 50  
 Nakayama, Toshihide 2: 51–2  
 Nater, H. F. 2: 61  
 Nedjalkov, Igor V. 2: 16, 303  
 Nesper, Marina 2: 11, 23, 27, 34–5  
 Newman, John 1: 308; 2: 188  
 Newman, Paul 1: 60; 2: 227, 352

- Newman, Stanley 2: 2, 22, 103  
 Ng, Eve 2: 188  
 Nguyễn, Đinh-Hòa 2: 78  
 Nichols, Johanna 1: 90; 2: 88, 96, 148, 158, 285, 310–11  
 Nida, Eugene A. 1: 81, 85, 179  
 Nikiforidou, Kiki 2: 263, 265  
 Noonan, Michael 1: 54, 91, 288; 2: 125, 231, 280, 286, 288, 303, 418, 420  
 Nordlinger, Rachel 2: 53, 217, 221
- Obata, Kazuko 2: 382  
 Ochs, Elinor 1: 55  
 O'Connor, M. 2: 369  
 Ohala, John J. 1: 90  
 Okell, John 2: 340  
 Oksefjell, S. 2: 312  
 Olawsky, Knut J. 2: 92, 178, 182, 299  
 Ömçürüyi, Thomas O. 2: 109–10  
 Ōnishi, Masayuki 1: 55, 128, 330; 2: 149, 158, 178, 195, 244, 292, 300, 379  
 Osada, Toshiki 2: 44–5, 49, 61  
 Osam, E. Kweku 2: 114  
 Osborne, C. R. 1: 205  
 Otanes, Fe T. 2: 52–3, 56  
 Otsuka, Yuko 1: 180  
 Overall, Simon 2: 251  
 Owens, Jonathan 2: 167, 299
- Packard, Jerome L. 2: 31, 35  
 Pagliuca, William 1: 259; 2: 301  
 Palmer, F. R. 1: 56, 288; 2: 84, 88  
 Palmer, H. E. 1: 90  
 Pāṇini 1: 45; 2: 68  
 Paris, Catherine 2: 158  
 Parsons, F. W. 2: 107  
 Pascual, Joan 2: 183  
 Pastika, Wayan 2: 386  
 Pawley, Andrew 2: 277  
 Payne, David L. 1: 180  
 Payne, Doris L. 1: 241; 2: 311  
   on Panare 2: 246, 279, 288, 292, 329, 347, 379  
   on Yagua 2: 200, 241, 272, 292  
 Payne, Thomas E. 1: 327  
   on Panare 2: 246, 279, 288, 292, 329, 347, 379  
   on Yagua 2: 200, 241, 272, 292  
 Pensalfini, Robert J. 2: 167  
 Peranteau, Paul M. 2: 448  
 Perkins, Revere 1: 259; 2: 301  
 Perlmutter, David M. 2: 155  
 Pesetsky, David 1: 91
- Phares, Gloria C. 2: 448  
 Philippaki-Warburton, Irene 2: 10, 26, 178, 201, 390  
 Piau, Julie 2: 200  
 Pike, Kenneth L. 1: 88–9, 201, 283, 287–8; 2: 25  
 Plaattje, Solomon T. 1: 326  
 Plank, Frans 1: 55  
 Platero, Paul R. 2: 327, 332–3  
 Platt, John T. 1: 262  
 Pope, M. K. 2: 227  
 Poppe, Nikolai N. 2: 85, 88  
 Post, Mark 2: 65  
 Poulos, George 2: 87  
 Press, Margaret L. 2: 81  
 Prista, Alexander da R. 2: 22  
 Protagoras 1: 155  
 Pullum, Geoffrey K. 1: 81, 94, 106; 2: 35, 342, 360–1  
 Pustet, Regina 2: 187, 198, 288, 300
- Quesada, J. Diego 2: 90  
 Quirk, Randolph 2: 245, 361
- Ramirez, Henri 2: 205  
 Ramstedt, G. J. 2: 83  
 Randal, Scott 2: 167, 172  
 Rangan, K. 2: 203  
 Rankin, Robert 1: 180; 2: 35, 243  
 Ransom, Evelyn 2: 418  
 Rappaport Hovav, Malka 2: 158  
 Refsing, Kirsten 1: 84; 2: 226  
 Reh, Mechthild 1: 180; 2: 178  
 Rehg, Kenneth L. 2: 200, 231  
 Reichard, Gladys A. 1: 90  
 Reichmann, W. J. 1: 257–9, 262  
 Rennison, John R. 2: 178  
 Reuse, Willem P. de 2: 291  
 Ribeiro, Eduardo R. 2: 340, 373  
 Rice, Keren 2: 12, 199, 279, 287  
 Roberts, John R. 1: 53, 56, 181; 2: 84, 199, 279, 288, 293, 352  
 Robins, R. H. 1: 90; 2: 10  
 Robinson, Stuart P. 2: 391  
 Rodrigues, Aryon D. 2: 277  
 Rose, Françoise 2: 114  
 Rose, Phil 1: 90  
 Rosenbaum, Peter S. 2: 419  
 Rosetti, Alexandru 2: 31, 34  
 Roth, Walter E. 1: 181  
 Routenaa, Judy 2: 114  
 Rowan, O. 2: 205  
 Rubino, Carl 1: 262; 2: 321

- Rumsey, Alan 1: 262, 288, 308  
 Rusell, Kevin 2: 18
- Sacks, Harvey 2: 36  
 Sadiqi, Fatima 2: 89  
 Sadler, Louisa 2: 53, 221  
 Sadock, Jerrold M. 2: 44, 242  
 Saeed, John I. 2: 66, 114, 200  
 Saltarelli, Mario 2: 84, 203, 213, 251  
 Sândalo, Filomena 2: 243  
 Sandler, Wendy 2: 12  
 Sands, Kristina 2: 392  
 Sankoff, Gillian 2: 343, 364  
 Sapir, Edward 1: 84–8, 226, 265–6  
   on morphological processes 1: 138–9, 146  
   on possession 1: 54; 2: 310  
   on Southern Paiute 1: 84, 265; 2: 99, 197  
   on Takelma 1: 84; 2: 89, 218  
   on word 2: 1, 31, 35  
   on word classes 2: 37–8, 54  
 Sari, Nirmala 2: 134  
 Sasse, Hans-Jürgen 2: 217, 251  
 Saussure, Ferdinand de 1: 25, 87  
 Schachter, Paul 2: 37, 52–3, 56, 62, 69, 81, 321  
 Schebeck, Bernhard 2: 196  
 Schegloff, Emanuel A. 2: 36  
 Schleicher, August 1: 180  
 Schmidt, Ruth L. 2: 347  
 Schuh, Russell G. 2: 201, 360  
 Schultze-Berndt, Eva 2: 217  
 Schütz, Albert J. 2: 108, 112  
 Seiler, Hansjakob 2: 310–11  
 Seki, Lucy 2: 70, 78, 80, 213, 342  
 Shibatani, Masayoshi 2: 98  
 Shirjaev, E. N. 2: 187  
 Shopen, Timothy 1: 86, 318  
 Siewierska, Anna 1: 246; 2: 258  
 Silverstein, Michael 1: 179  
 Sinclair, John 1: 307  
 Skorik, P. Ja. 1: 181  
 Slater, Keith W. 2: 179, 329  
 Smeets, Ineke 2: 229, 242  
 Smith, Ian 2: 213  
 Smith, Norval 2: 259, 337, 360  
 Sneddon, James Neil 1: 181, 2: 181, 231, 241, 268, 292  
 Sohn, Ho-min 1: 55, 77, 82, 103; 2: 333, 338, 352  
 Sokolov, S. N. 2: 85  
 Sprigg, R. K. 2: 10  
 Staden, Miriam van 1: 55  
 Stankiewicz, Edward 1: 180, 209  
 Stassen, Leon 2: 187  
 Stebbins, Tonya N. 2: 114, 120, 157, 217  
 Stenson, Nancy 1: 128; 2: 373, 386, 393, 402  
 Stumme, Hans 2: 246  
 Subbarao, Karumuri V. 2: 317  
 Sumbuk, Kenneth K. 2: 114  
 Sutton, P. 2: 218  
 Svartvik, Jan 2: 456  
 Svorou, Soteria 2: 311  
 Swadesh, Morris 1: 88, 179, 215, 240, 287, 299; 2: 37, 43, 56, 90  
 Swanton, John R. 2: 64  
 Sweet, Henry 1: 135; 2: 1–2, 19, 36, 61, 227, 315, 359  
 Szemerényi, Oswald J. L. 2: 364
- Takeuchi, Lone 2: 94  
 Tamura, Suzuko 1: 179; 2: 166, 205, 226, 240, 290  
 Taumoeofalou, Malenaite 1: 179–80  
 Teeter, Karl V. 2: 286–7  
 Tepljashina, T. I. 1: 181  
 Terrill, Angela 2: 261  
 Thalbitzer, William 2: 44  
 Thiesen, Wesley 2: 201  
 Thomas of Erfurt 2: 68  
 Thomas, Alan R. 2: 302  
 Thomas, David 2: 258, 302  
 Thompson, Chad 2: 281–2, 301, 310  
 Thompson, Sandra A. 1: 89; 2: 69, 143, 149, 156, 183, 268, 331, 337, 342  
 Thomsen, Marie-Louise 1: 84  
 Thornell, Christina 2: 183  
 Thráinsson, H. 2: 148  
 Timberlake, Alan 2: 170  
 Topping, Donald M. 2: 78  
 Tosco, Mauro 2: 299, 301, 344, 352  
 Toulmin, Stephen E. 1: 1  
 Traugott, Elizabeth C. 2: 409  
 Travis, Catherine 2: 168, 175, 183  
 Trubetzkoy, Nicolai S. 1: 70, 85–8, 235–6, 272, 287; 2: 8–11;  
 Tucker, A. N. 2: 86  
 Tuite, Kevin 2: 251
- Ullmann, Stephen 2: 34
- Vago, Robert M. 2: 11, 346, 352  
 Valentine, J. Randolph 2: 181, 321, 384  
 Valenzuela, Pilar 2: 165, 322, 333  
 van den Berg, René 2: 244, 252  
 van Driem, George 2: 187  
 van Eijk, Jan P. 2: 38, 51, 53, 56, 61, 241, 244

- van Wyk, E. B. 2: 6, 17  
 Variar, K. M. 2: 91, 183  
 Vázolyi, Eric G. 1: 262  
 Velazquez-Castillo, Maura 2: 302, 310–11  
 Vendryes, Joseph 2: 6–7, 10  
 Verhaar, John 2: 187  
 Veselinova, Ljuba 2: 187  
 Vesper, Don R. 2: 178  
 Vincennes, St Louis de 2: 187  
 Vitale, Anthony J. 1: 85  
 Voegelin, Charles F. 1: 88  
 Voeltz, F. K. Erhard 1: 302  
 Vogel, Irene 2: 11, 23, 27, 34–5  
 Vogt, Hans 1: 241; 2: 53, 61, 346  
 von Roncador, Manfred 2: 419
- Walsh, Michael J. 2: 332  
 Waltke, Bruce K. 2: 369  
 Watanabe, Honoré 2: 61  
 Waterson, Natalie 2: 11  
 Watkins, Calvert 2: 364  
 Watkins, Laurel J. 2: 217, 283, 381  
 Watters, David E. 1: 82–3; 2: 63, 125, 222,  
 319, 408  
 Waugh, Linda R. 2: 114  
 Weber, David J. 2: 251, 384, 409  
 Weiner, Edmund 1: 91  
 Weinreich, Uriel 2: 10, 34  
 Wells, Rulon S. 2: 2  
 Welmers, Beatrice F. 2: 81, 84, 114  
 Welmers, William E. 2: 81, 84, 114  
 Werth, Elizabeth 1: 90  
 Westermann, Diedrich 1: 180
- Wetzler, Harrie 2: 114  
 Whorf, Benjamin Lee 1: 180  
 Wiesemann, Ursula 2: 201  
 Wierzbicka, Anna 1: 308  
 Wilkins, David 2: 259–60  
 Wilkinson, Melanie 2: 412  
 Willett, Thomas L. 2: 243  
 Williams, Corrinne J. 2: 180  
 Williams, Edwin 2: 35  
 Williamson, Janis S. 2: 344  
 Williamson, Kay 2: 84  
 Willie, Mary Ann 2: 327, 332  
 Wonderly, William L. 2: 8  
 Wongkhomthong, Preya 2: 364  
 Woodbury, Anthony C. 2: 19, 21, 23, 222  
 Wordick, F. J. F. 1: 288  
 Wouk, Fay 2: 176
- Xu Weiyaun 2: 69, 78–9, 81, 83
- Yip, Virginia 2: 180, 192  
 Yoshimura, Chikao 2: 272  
 Young, Robert W. 2: 343
- Zaenen, A. 2: 148  
 Zandvoort, R. W. 2: 245  
 Zeitoun, Elizabeth 1: 83; 2: 382, 387, 405  
 Zeshan, Ulrike 2: 12, 339  
 Zhu Deni 1: 179  
 Žirmunskij, V. M. 2: 1  
 Zuckermann, Ghil'ad 2: 373, 379, 385, 393,  
 399, 403, 405  
 Zwicky, Arnold M. 2: 35

# Language Index

- Abaza 2: 213, 305, 326, 345  
Abkhaz 2: 135, 326, 345  
Abui 2: 240  
Acehnese 1: 347, 359; 2: 135, 205, 268, 272  
Acoma 2: 217–18, 258  
Adjnyamathanha 2: 213  
African languages 1: 155, 302; 2: 29, 74, 256, 259, 301  
Aguaruna 2: 251  
Ainu 1: 84; 2: 96, 205, 207, 225, 379  
    copula clauses 1: 101, 179; 2: 166–8  
    demonstratives 2: 240–1, 250  
    possession 2: 290, 293, 301  
Akan 1: 10–11, 55; 2: 75–6, 114, 191, 277  
Akkadian 2: 379, 385, 390, 393, 405, 408–12, 420  
Alamblak 2: 86  
Aleut 2: 271, 292, 305  
Algonquian languages 1: 227; 2: 96, 204  
Alutor 1: 253, 262  
Amele 2: 84, 199, 352  
    possession 1: 5–6, 11–12, 54; 2: 278–9, 283, 285, 288, 293, 298  
    tense and polarity 1: 47, 56, 162–3, 181  
American sign language 2: 339  
Amharic 1: 140, 163, 180–1; 2: 213, 344, 349  
Angami 2: 267, 270, 272  
Anindilyakwa 2: 291  
Anywa 1: 140, 186; 2: 178  
Apalai 2: 274, 311  
Arabana 2: 182  
Arabic 1: 142, 180, 196; 2: 84, 86, 292, 321, 35  
    *also see* Classical Arabic  
Arapesh 2: 86  
Arawak language family 1: 15, 259, 303; 2: 96, 204–5, 241, 311  
Archi 2: 151  
Arrente 2: 9, 23, 25  
Athapaskan 2: 204  
Australian languages 1: 45, 165, 180, 223, 296, 308; 2: 31, 69, 97, 157, 182, 283, 291, 399  
    phonology 1: 210–11, 250, 270, 283, 288; 2: 8  
    pronouns 2: 194, 196, 213, 219, 257–9  
Avar 2: 151, 212  
Avestan 2: 85  
Awa Pit 2: 87, 178, 231–2, 272, 274, 310–11, 403  
Axininca Campa 1: 159, 189  
Aymara 2: 197  
Baagandji 2: 194  
Bahnar 1: 90  
Bahuna 2: 205  
Balto-Finnic languages 1: 165  
Bandjalang 2: 87  
Baniwa of Içana 2: 158, 276  
Banjarese 2: 134  
Bantu languages 1: 226, 263; 2: 5, 137, 157  
Bardi 1: 231–2, 241, 246  
Bare 2: 9  
Basque 2: 84, 97, 180, 203, 213, 239, 250–1, 277, 379  
Batsbi (Tsova-Tush) 1: 73, 91; 2: 121, 141, 157  
Bella Coola 2: 454  
Bengali 1: 17, 55, 161, 180; 2: 8, 130, 150, 202, 234, 244, 292, 358  
Berber 1: 52; 2: 63, 91, 106  
Biblical Hebrew 2: 9, 369  
Bilin 2: 84, 88  
Bilua 2: 382, 384  
Blackfoot 2: 311  
Bolivian Quechua 2: 271  
Bora 2: 201  
Bororo 2: 81  
Boumaa Fijian, *see* Fijian  
Brahui 2: 260  
Burarra (Gun-Nartpa) 2: 157  
Buriat 2: 85, 88  
Burmese 2: 135  
Buru 2: 270  
Burushaski 1: 69–70, 90; 2: 97, 141  
Caddo 2: 205  
Caddoan languages 2: 140  
Cairene Egyptian Colloquial Arabic 2: 292, 353  
Cambodian, *see* Khmer  
Cantonese 2: 180, 192  
Carib languages 2: 92–4, 197, 311

- Cashinawa 1: 123–4  
 Cavineña 2: 165, 181  
 Cayuga: 2: 60, 217–8, 250–1  
 Celtic languages 2: 12  
 Central Pomo 2: 140  
 Chadic languages 2: 197  
 Chamorro 2: 78  
 Chechen 2: 8, 148, 151  
 Chemehuevi 2: 81  
 Cherokee 1: 65, 90; 2: 77–8  
 Chimariko 2: 410  
 Chipweyan 1: 115, 179; 2: 191  
 Chinese, *see* Mandarin Chinese, Cantonese  
 Chitimacha 2: 90  
 Chrau 2: 302  
 Chukchi 1: 165, 181  
 Classical Arabic 2: 9  
     *also see* Arabic  
 Classical Armenian 2: 141  
 Classical Greek 2: 263, 265  
     *also see* Greek  
 Classical Nahautl 2: 61  
 Coast Timshian 2: 217  
 Colloquial Arabic 1: 142, 196; 2: 292, 353  
 Colloquial Burmese 2: 340, 364  
 Colloquial Czech 2: 236  
 Colloquial Welsh 2: 183, 302  
 Comanche 1: 159, 180  
 Coos 2: 37, 42, 64  
 Cree 2: 18  
 Creek 2: 278, 280, 283–7, 293  
 Cup'ik Eskimo 2: 19, 21, 23, 44  
 Cushitic languages 2: 300
- Dagbani 2: 92, 299  
 Dakota 1: 84, 141, 180; 2: 278, 281–4, 288–9  
     *also see* Lakota  
 Danish 2: 342, 364  
 Dëne Suñiné, *see* Chipewyan  
 Dhaasanac 2: 299–300, 343–4, 352  
 Dhalanji 2: 156  
 Diegueño 2: 334  
 Diyari 2: 169, 181  
 Dhamindhung 2: 217  
 Djambarrpuyngu 2: 412  
 Djapu (a Yolngu dialect) 1: 280, 288  
 Djeebbana, *see* Ndjébbana  
 Djingulu 2: 167–8  
 Dolakha Newar 2: 165, 175, 258, 342, 357,  
     390, 404, 414  
 Dravidian languages 1: 305; 2: 234, 260  
 Dumi 2: 187  
 Dutch 2: 6
- Dyrbal 1: 19–20, 61, 68–9, 82–3, 96, 133,  
     144–8, 172–7; 221–2; 2: 22, 206–7, 238–9  
     adjectives 1: 26; 2: 63, 85–8, 99–104  
     antipassive 1: 148, 165–8, 172–3, 188–9,  
     234  
     avoidance style (Jalnguy) 1: 215–16,  
     293–5, 308  
     cases and genitive 1: 44–5, 126–8, 188–91,  
     213  
     complementation strategies 2: 351, 398,  
     400–13  
     constituent order and word order  
     1: 22–4, 233–4  
     demonstratives 2: 229, 237, 242, 244, 248,  
     250  
     derivational suffixes 1: 41, 219–20;  
     2: 14–17  
     lexicon 1: 22–3, 256, 283, 290–1, 300–8,  
     344  
     locational suffixes 1: 16–17, 69; 2: 320–1  
     noun classes 1: 27–31, 55, 238; 2: 85–6  
     phonology 1: 198–200, 210–12, 265–6,  
     272–3, 278, 283; 2: 30  
     possession 1: 5–6; 2: 268–9, 272, 275  
     reduplication 1: 140, 294; 2: 56, 87  
     relative clauses 2: 319–24, 334, 341–2, 349,  
     351–2, 364, 407–8  
     structure/word class correspondence  
     1: 110; 2: 41  
     transitivity 2: 121–2, 131, 134, 150, 157
- East Tucanoan 2: 89  
 Edo 2: 109–10, 126  
 Egyptian 2: 183  
 Egyptian Colloquial Arabic 1: 142, 180;  
     2: 321  
 Emerillon 2: 114  
 Emmi 2: 87, 90, 97–8  
 Enga 1: 257, 262  
 English 1: 22–3, 76, 79, 107–8, 118–22, 137–9,  
     172–6, 225, 241, 255, 283, 301, 305;  
     2: 27–30  
     adjectives 1: 52–3, 113, 240, 303; 2: 70–1,  
     76, 80, 90, 99, 104, 114  
     clause linking: 1: 133–7, 228  
     comparatives 1: 177–9, 201–2  
     complement clauses 1: 30–34, 55, 97,  
     129–32; 2: 370–90, 394–401, 408, 411,  
     417–20  
     constituent order 1: 37–8, 72–3  
     copula constructions 1: 100–1; 2: 163,  
     170–4, 177, 180, 183, 186–7  
     definiteness 1: 160–1, 180

- English (*cont.*)  
 demonstratives 2: 223–8, 231–6, 245–52, 260  
 derivational processes 1: 149–52, 185, 217–19; 2: 16, 46–9  
 irregular forms 1: 15, 140, 205–6  
 number marking 1: 143, 158, 184–5, 302; 2: 55  
 orthography 1: 67, 90, 281  
 passive 1: 166–8, 231  
 phrasal verbs 1: 36–7, 290  
 phonology 1: 9, 264–6, 272, 277  
 possession 1: 5–6, 229; 2: 262–6, 295–9, 302, 311  
 pronouns 1: 10, 18, 47, 115; 2: 200–4, 207, 210  
 relative clauses 1: 23; 2: 313–6, 319–20, 324–5, 329, 346–51, 359–62, 367  
 semantic types 1: 42–3, 50, 104–6, 394–411  
 time specification 1: 118–22, 154  
 transitivity 1: 98–105; 2: 118–9, 124, 127–34, 144, 147, 149, 154, 147  
 word 2: 2–5, 17–26, 36  
 word classes 1: 24–7; 2: 39–40, 61
- Eskimo 1: 65, 90; 2: 19, 21, 23, 44–5, 204, 243, 275, 291, 304
- Estonian 1: 51, 164–5; 2: 8, 149–50, 170
- Evenki 2: 16, 303
- Ewe 1: 5–6, 54; 2: 253, 278, 280, 285, 288
- Fijian 1: 23–4, 59–60, 82–3, 90, 108, 141, 199–200, 255, 286–8, 301–3, 307–9; 2: 23–4, 161–4  
 adjectives, *see* distinguishing word classes  
 complement clauses 2: 386, 393, 400–2, 420  
 constituent order 1: 74–5, 91, 233  
 demonstratives 2: 229–33, 240, 242, 251  
 distinguishing word classes 2: 37–8, 42, 46, 52–6, 65, 71–2, 78–83, 100, 108–12  
 number words 1: 191–3, 213  
 phonology 1: 66–7, 197–9, 278, 281–2  
 possession 1: 303; 2: 270–8, 284–8, 298, 307–10  
 pronouns 2: 191–2, 202, 208, 215, 218  
 relative clauses 1: 186–8, 213; 2: 321, 327–9, 336, 339, 353–6  
 transitivity 2: 131–2, 135, 143–7, 157–8  
 sentence 1: 133, 179  
 word 2: 4, 8–9, 20–26, 36
- Finnish 1: 159; 2: 86, 180, 298, 392  
 adjective class 1: 63, 68, 88  
 cases 1: 12, 223; 2: 148, 158, 167  
 Finno-Ugric languages 2: 277  
 Flinders Island language 2: 218  
 Fox 2: 250–1  
 French 1: 160, 270; 2: 4–7, 114, 180, 227, 278, 298–9, 320  
 demonstratives 2: 231–2, 238, 242  
 gender 1: 12, 155–7, 163  
 pronouns 1: 18, 202, 204  
 Fula 2: 86
- Gaaguju 2: 182
- Gala 1: 21; 2: 201
- Gapapaiwa 1: 6, 54; 2: 281–5, 290
- Georgian 1: 45; 2: 239, 250–1, 364
- German 1: 68, 160, 224; 2: 4–5, 41, 227, 239, 292, 298–9, 316  
 gender 1: 12, 155, 163  
 phonology 1: 224, 267, 272  
*see also* Pennsylvania German
- Germanic languages 1: 15; 2: 27, 364
- Girramay: dialect of Dyirbal, q. v.
- Goemai 2: 419
- Gog-Nar: 2: 291
- Gokana 2: 251
- Gola 1: 161, 180
- Gooniyandi 2: 182
- Greek 1: 76, 155, 160, 225–6; 2: 5, 10, 26, 68, 177, 390  
 cases and genitive 2: 45–6, 225  
 gender 1: 155; 2: 68  
 pronouns 2: 190, 201  
*see also* Classical Greek
- Greenlandic Eskimo 2: 275, 304
- Guajiro 2: 286
- Guarani 2: 8, 140, 302, 310
- Gugada (dialect of the Western Desert language) 1: 260
- Gugadj 2: 291
- Gumbaynggirr 1: 60, 90; 2: 113, 182
- Gurindji 2: 197–8
- Guugu Yimidhirr 2: 142, 158, 182
- Haida 2: 5–6, 54, 64, 140
- Hanis 2: 37, 42, 64
- Hanunóo 2: 140
- Harar Oromo, *see* Oromo
- Hārauti 2: 210
- Hausa 1: 60; 2: 234, 246, 251, 352, 360, 419  
 adjectives 2: 74–6, 107, 114
- Haya 2: 311

- Hdi 1: 15, 179  
 Hebrew 1: 12, 163–4; 2: 353, 363  
*see also* Modern Hebrew  
 Hindi 1: 165; 2: 135, 162, 174, 177, 358  
 Hixkaryana 2: 92–3, 239, 270–1, 316  
 Hmar 2: 317–8, 333, 339  
 Hopi 2: 240  
 Hua 2: 84, 91–2, 199, 243, 272–3, 285, 293, 368  
 Huichol 2: 135  
 Hungarian 1: 12, 20, 55, 2: 5, 11, 15, 63, 88, 181, 346, 353, 364
- Icelandic 2: 148–50  
 Igbo 2: 53, 109; 2: 269–70, 277–8, 340  
 adjectives 2: 63, 74–5, 81, 84, 114  
 Ijo 2: 64  
 Ilocano 2: 197, 258, 321  
 Indo-Aryan 2: 141, 234, 357  
 Indo-European 1: 8, 15, 155–6, 224; 2: 4, 41, 97, 213, 277, 301, 360, 364  
 Indonesian 1: 178, 181; 2: 181, 231–3, 241, 268, 292, 386, 420  
 Ingush 2: 88, 148, 151  
 Iranian 2: 141  
 Iraqw 2: 200, 216, 221, 292  
 Irish 2: 373, 381, 393, 402  
 Iroquoian 2: 60  
 Israeli sign language 2: 12  
 Italian 2: 21
- Jacalteco 1: 18, 55, 118, 179; 2: 92, 304, 404  
 Jalnguy, *see* Dyirbal avoidance style  
 Japanese 2: 94–5, 98–9, 169, 277, 298, 304, 349–50  
 pronouns and demonstratives 2: 201–3, 231–3, 239, 259  
 Jarawara 1: 23, 60, 82–3, 117, 137, 140, 152, 233, 255, 290, 324; 2: 17, 23–4, 84, 238  
 adjective class 2: 74–6, 86, 114  
 complementation 2: 372, 379–89, 393, 400, 403  
 copula clauses 2: 165, 170, 173–4, 180, 184  
 gender 1: 62–3, 156–7, 238  
 lexicon 1: 301–2, 302, 306, 308  
 number system 1: 11, 159  
 phonology 1: 278; 2: 23, 29  
 pronouns 2: 205, 215, 221  
 possession 1: 230; 2: 278, 281–5, 293, 296–300, 304, 311–12  
 relative clauses 2: 319, 321, 349, 352, 356  
 sentence 1: 75, 91, 133, 179  
 transitivity 2: 124–5, 132–5, 144–7, 154, 157
- Jersey Norman French 2: 272–3  
 Jirrbal: dialect of Dyirbal, q.v.  
 Juúhoan 2: 260
- Kabardian 1: 288  
 Kabyle 2: 167–8  
 Kaike 2: 222–3  
 Kalispel 2: 53  
 Kalkatungu 2: 135  
 Kamaiurá 2: 70, 78, 80, 213, 221, 293, 342, 404  
 Kambera 1: 24, 55; 2: 352  
 Kamula 2: 75–6, 114  
 Kana 2: 136, 178, 192, 254, 299  
 Kanada 2: 200, 260  
 Karajá 2: 201, 340, 363, 373  
 Karbi 2: 169, 272  
 Karok 2: 64  
 Kayardild 1: 82–3; 2: 36  
 number system 1: 10–11, 55; 2: 191  
 pronouns 1: 115, 179; 2: 208  
 Kazakh 2: 11  
 Kewa 2: 10  
 Kham 1: 82–3; 2: 63, 74, 125, 319, 408–9  
 Khmer 1: 69; 2: 202, 231–3, 242, 277  
 Kinyarwanda 2: 442  
 Kiowa 2: 217–8, 283, 331–2  
 Kiranti 2: 333, 349  
 Kiriri 2: 277  
 Koasati 1: 82–3, 152, 180, 233; 2: 93, 200, 316  
 Kobon 2: 331, 343, 353  
 Koiari 2: 136, 194  
 Kolokuma dialect of Ijo 2: 64  
 Koran (or !Oro) 2: 259  
 Korean 1: 65, 73, 90, 265; 2: 277  
 adjectives 2: 63, 77, 82–3, 99, 103  
 relative clauses 2: 320, 333, 338, 352  
 state verbs 1: 19–20, 55  
 Korku 2: 42, 50  
 Koromfe 2: 178  
 Koyukon Athabaskan 2: 281–4, 289, 292, 301, 310  
 Krahn/Wobé 2: 76  
 Kresh 1: 163, 181  
 Kugu Muminh (or Kugu Nganhcara) 2: 213  
 Kuman 2: 200  
 Kurdish 1: 165, 181  
 Kurukh 2: 178  
 Kwakiutl 2: 244
- Ladakhi 2: 291  
 Lahu 1: 64, 81, 83; 2: 243



- Lak 1: 12, 15–16, 55; 2: 242  
 Lakota 2: 140, 199, 344, 364  
*also see* Dakota  
 Lango 1: 6, 11–12, 54, 279, 288; 2: 125, 231–2, 239  
   possession 2: 278, 280, 283–8, 298, 303  
 Lao 1: 83; 2: 113, 236, 259, 329, 338, 347  
 Lardil 1: 17, 55; 2: 182  
 Latin 1: 79–80, 97, 155, 240, 282; 2: 85, 124, 150, 190, 213, 265, 299  
   cases and prepositions 1: 9, 45–6, 96, 165, 224–5, 299; 2: 123, 150, 265  
   constituent order and word order 1: 37–8, 71–2  
   copula clauses 1: 101, 254; 2: 160, 183  
   fusional character 1: 43–4, 49, 55, 117, 144–6, 217, 220; 2: 58, 216  
   word 2: 2–5, 10, 15, 17  
   word classes 1: 25–6, 52, 102, 110, 194; 2: 38–41, 63, 68  
 Latvian 2: 298  
 Lavukaleve 2: 261  
 Lezgian 1: 152, 189; 2: 151, 158, 347  
 Lillooet 2: 51, 53, 241, 244  
 Longgu 1: 10–11, 55; 2: 252  
 Luisño 1: 50, 56  
 Luritja (dialect of the Western Desert language) 2: 211  
 Lushootseed 2: 51, 53  
  
 Maasai 1: 73; 2: 86  
 Mabuig: dialect of West(ern) Torres Strait language, q.v.  
 Macedonian 2: 361  
 Macushi 1: 73, 91; 2: 93–4, 399  
 Madi-Madi (dialect of Wemba-Wemba) 2: 220–1  
 Maká 2: 286  
 Makah 2: 51, 53, 401–2  
 Malagasy 1: 18–19, 55; 2: 320, 324  
 Malay 1: 83  
 Malayalam 2: 91–2, 180, 183, 260, 336, 357  
 Mali (Baining) 2: 120, 157  
 Mam 1: 82–3, 168, 181; 2: 64, 90  
 Mamu: dialect of Dyirbal  
 Manambu 1: 83, 164, 181, 277, 306–8; 2: 13–14, 362; 408–12  
   copula clauses 2: 169, 173, 176–7, 182, 184, 300  
   pronouns 2: 199, 201, 206–7  
   relative clauses 2: 316, 349–50, 352  
   transitivity 2: 124, 145, 154, 157  
  
 Manange 2: 95  
 Mandarin Chinese 1: 5–7, 110–11, 179, 227; 2: 42–3, 173, 183  
   adjectives 1: 52; 2: 63, 69–70, 78–9, 81, 83, 88, 95  
   demonstratives 2: 233, 246–7  
   phonology 1: 267, 277, 279, 288  
   possession 2: 268–9, 272, 278  
   relative clauses 2: 330, 339, 342  
   word 2: 4–5, 30–31, 35  
 Mangap-Mbula 2: 231–2  
 Mangarayi 2: 198, 291  
 Mangghuer 2: 179, 329  
 Mantauran (Rukai) 1: 83; 2: 382, 387, 409  
 Mao Naga 2: 42, 54  
 Mapuche 2: 229, 242  
 Margi 2: 197  
 Maricopa 1: 5–6, 54, 163, 181  
 Martuthunira 1: 250, 267, 276; 2: 157, 182  
 Matses 2: 204, 410, 419  
 Mayali 2: 311  
 Mayan languages 2: 80, 141  
 Mbyá (variety of Guaraní) 2: 8  
 Mende 2: 212  
 Menomini 1: 328  
 Middle English 1: 51  
 Mingrelian 2: 168  
 Miya 2: 201, 360  
 Modern Hebrew 2: 325, 342  
   complementation 2: 373–4, 379, 385, 393, 399, 403, 405  
   *see also* Hebrew  
 Modern Standard Arabic 2: 84  
 Mohawk 2: 284  
 Mojave 2: 81, 167–8, 333–4, 337, 339, 343, 350, 352  
 Mokilese 2: 101, 399  
 Moses-Columbia Salish (Nxaʔamxćin) 2: 22, 27  
 Motuna 1: 330; 2: 178, 195, 300  
 Muna (Sulawesi) 2: 244, 252  
 Mundari 2: 43–4, 49, 50, 178  
 Mupun 2: 79, 82–3, 169, 231–3, 330, 352, 364  
 Murinypata 2: 332  
 Muskogean 2: 96  
 Mÿky 2: 269–70, 272  
  
 Nakkara 2: 135  
 Nanai 2: 283  
 Navajo 2: 327, 332–3, 337, 343  
 Ndjébbana 2: 293  
 Nenets 1: 110; 2: 42

- Ngajan: dialect of Dyrbal, q.v.  
 Ngandi 1: 163, 181  
 Ngiyambaa 1: 161–2, 180; 2: 213  
 Nhangu 2: 196  
 Nilo-Saharan languages 2: 97  
 Nishnaabemwin: dialect of Ojibwe, q.v.  
 Njangumarta 2: 182  
 Nootka  
   possession 1: 5–6, 11–12, 54  
   structure/word class correspondence  
     1: 111, 179; 2: 37, 45, 51–2, 56–9, 61,  
     103  
 North-east Ambae 2: 45, 82, 113, 406, 408  
 North-eastern Neo-Aramaic language 2: 11  
 Northern Sotho 2: 5, 17  
 Northern Subanen 2: 75–6, 114, 165–6, 237,  
   248  
 Nuer 2: 177  
 Nunggubuyu 2: 63, 89, 97–9, 201–2, 209  
 Nuuchahnulth, *see* Nootka  
 Nyawaygi 1: 277, 284–6, 300, 303
- Oceanic languages 1: 313; 2: 8, 45, 54, 70,  
   101, 202, 205, 259,  
   possession 2: 276–7, 284, 288, 290  
 Ojibwe 2: 181, 321, 384  
 Old English 1: 62; 2: 3, 220, 227  
 Old Norse 2: 342, 364  
 Olgolo 1: 9, 54  
 Oroco 2: 167–8, 227, 299–300
- Paamese 2: 155  
 Palikur 2: 244  
 Panare 2: 246, 278–9, 285, 288, 292, 327, 329,  
   337, 379  
 Papantla Totonac 2: 77, 80, 84, 88  
 Papuan languages 1: 308, 313; 2: 198–99  
 Parecis 1: 140; 2: 205  
 Pāri 2: 142  
 Passamaquoddy 2: 188  
 Patjtjamalh 2: 215, 221  
 Pennsylvania German 1: 18, 55; 2: 393, 405  
 Persian 2: 330, 352, 360  
 Pitjantjatjara (dialect of the Western Desert  
   language) 2: 8  
 Pitta-Pitta 1: 165; 2: 182  
 Polish 2: 135, 170  
 Pomoan languages 2: 204  
 Ponapean 2: 200, 231–2, 239  
 Portuguese 1: 18, 67, 160, 237–8, 281; 2: 21–2,  
   129, 176, 260  
 Punjabi 2: 152, 158, 303, 358  
 Purki 2: 203
- Quechua 1: 9, 45; 2: 60, 85, 97, 181, 184, 197,  
   204, 271, 409  
   demonstratives 2: 239, 250–1  
   relative clauses 2: 333, 339, 349  
 Qeqchi Maya 2: 386  
 Qiang 2: 79, 114, 165  
 Quileute 1: 137, 179; 2: 60
- Rarotongan 2: 70  
 Rembarnga 2: 359, 362  
 Rotuman 1: 69, 90  
 Rukai (Mantauran) 1: 83; 2: 382, 387, 409  
 Rumanian 2: 201, 353, 360  
 Russian 1: 225, 281, 290; 2: 4, 88, 135, 292,  
   316, 372  
   copula clauses 2: 162, 167, 170, 181, 187  
   genders 1: 12, 156  
   tense and aspect 1: 162, 181
- Sahaptin 2: 97  
 Salinan 1: 227  
 Salish(an) languages 1: 190; 2: 96, 241  
   distinguishing noun from verb 2: 22, 27,  
   38, 43, 51–56, 59–61  
 Samoan 1: 141, 210; 2: 13, 44–5, 61, 126  
 Sango 2: 183  
 Sanskrit 1: 45–6; 2: 68, 137, 190  
   phonology 1: 8, 10, 12, 272, 279–80  
 Sanuma 2: 193  
 Sarcee 1: 220–1; 2: 242  
 Sare 2: 75, 114  
 Semelai 1: 83, 187; 2: 82  
 Semitic 1: 139, 143, 301, 350; 2: 9, 189, 201,  
   310  
 Serbo-Croatian 2: 336  
 Shih 2: 246  
 Shipibo-Konibo 2: 166, 322, 333  
 Shoshone 2: 244  
 Shuswap 2: 401, 420  
 Sierra Miwok 1: 84–5; 2: 97  
 Sign languages 2: 12, 339  
 Sinhala 2: 326  
 Siouan (languages) 1: 84; 2: 35, 64, 96, 197,  
   243  
 Siriono 1: 140  
 Siuslawan (Lower Umpqua) 2: 64  
 Slave 2: 199, 278–9, 283, 287, 298  
 Slavic languages 1: 86, 153–4  
 Somali 2: 66, 75–6, 114, 200, 301  
 Sorowahá 2: 259  
 Southeastern Tepehuan 2: 243  
 Southern Paiute 1: 84, 265; 2: 99, 197  
 Southern Sotho 2: 5

- Spanish 1: 67, 281; 2: 63, 70, 129, 148, 213, 216, 260, 300  
 copula clause 2: 168–9, 175–6, 178, 183  
 Sranan 2: 176  
 Sumerian 1: 84, 159, 180; 2: 162  
 Sundanese 2: 10  
 Supyire 1: 12, 55; 2: 236, 246, 250–1, 343, 364  
 Suruí 2: 209  
 Swahili 1: 12, 85, 157, 177; 2: 86, 239, 277
- Tachelhit 1: 5–6, 12, 54  
 Tagalog 2: 37, 52–3, 56, 59  
 Takelma 1: 84, 218; 2: 63, 89–90  
 Tamambo 2: 217, 299  
 Tamil 2: 108, 174, 177, 233, 242, 250–1, 260, 302, 320, 357  
 Tarascan 2: 97  
 Tariana 1: 20, 83, 164, 181, 290; 2: 54, 84, 241, 293  
 adjectives 1: 194–6, 213; 2: 83, 88–9  
 complement clauses 2: 379, 382–6, 390, 393, 406, 409–10  
 copula clauses 2: 163–4, 169, 179  
 evidentials and tense 1: 15, 18–19, 44  
 relative clauses 2: 321, 330, 348–9  
 pronouns 2: 205, 210–11, 221  
 transitivity 2: 124, 126, 145, 147  
 word 2: 8, 17, 21, 33  
 Tarma Quechua 2: 181, 184  
 Tawala 2: 135  
 Telugu 1: 69, 307–8; 2: 108, 135, 224, 242, 247, 260  
 Temiar 2: 78  
 Tennet 2: 167–8, 172  
 Teribe 2: 64, 90  
 Thai 2: 5, 63–5, 113, 119, 202, 239, 300  
 Tialo 2: 272, 298  
 Tibetan 2: 10  
 Tibeto-Burman languages 2: 222, 258, 319, 333–4, 342, 363  
 Tigak 2: 80  
 Tiriyo (or Trio) 2: 92–3, 197–8, 311  
 Tiwi 1: 39, 40, 55, 116, 205, 227; 2: 199  
 Tjajtjala (dialect of Wemba-Wemba) 2: 219  
 Tlingit 2: 64  
 Toba-Batak 2: 82  
 Tok Pisin 1: 21, 320; 2: 343, 364  
 Tongan 1: 160, 180, 284, 288; 2: 54  
 transitivity 1: 99–100, 179; 2: 116–8, 154, 161  
 Tonkawa 1: 140; 2: 60, 215
- Totonac 2: 77, 80, 84, 88, 90  
 Trio (or Tiriyo) 2: 92–3, 197–8, 311  
 Trumai 1: 100, 179  
 Tsimshian 2: 96, 278–9, 285–6, 292  
*see also* Coast Tsimshian  
 Tsova-Tush (Georgian name for Batsbi) 2: 121, 157  
 Tswana 2: 5  
 Tübatulabal 1: 88  
 Tucano 1: 13, 15, 55, 47, 56  
 Tucanoan languages 1: 15, 20, 210, 221; 2: 89, 205  
 Tukang Besi 2: 61, 82  
 Tunica 1: 84; 2: 64, 91, 201, 221–2, 287, 293  
 Tupuri 2: 81  
 Turkic languages 1: 226; 2: 11, 88  
 Turkish 1: 9, 12, 20, 55, 161, 227; 2: 64, 271  
 word 2: 10, 12, 15, 17, 23–4  
 Tuscarora 2: 239, 241, 293, 300  
 Tuvaluan 2: 322, 349, 351  
 Tuyuca 1: 13–14, 55, 162, 181  
 Tyaddyuwurru (dialect of Wemba-Wemba) 2: 219  
 Tzotzil 1: 69–70, 90, 304, 308; 2: 391
- Udmurt 1: 164, 181  
 Ungarinjin 1: 260, 262  
 Upper Necaxa Totonac 2: 84, 90  
 Uradhi 2: 363  
 Urarina 2: 178, 181  
 Urdu 2: 347, 358  
 Ute 2: 301, 338, 349
- Venda 2: 87  
 Veps 1: 164  
 Vietnamese 1: 148, 226; 2: 58, 63, 78, 234, 277
- Waga-Waga 2: 182  
 Waikurúan languages 2: 243  
 Wakashan languages 1: 37, 43, 45, 51–56, 59–60  
 Walmatjari 2: 8, 182  
 Wambaya 2: 217  
 Wappo 2: 331, 349, 372  
 Wardaman 1: 82–3, 277; 2: 182  
 Warekena 1: 191–6, 213, 245; 2: 8–9, 126, 156–7; 275, 393, 320–1  
 Warlpiri 1: 282, 306; 2: 135, 213, 218–9, 358, 362  
 Warray 2: 182  
 Warrgamay 1: 197, 203, 238, 277, 300; 2: 24, 237, 248

- Watjarri 2: 213, 241  
Welsh 2: 320, 326  
Wemba-Wemba 2: 219–21  
Weri 2: 198  
West Greenlandic (Eskimo) 2: 275,  
304–5  
Western Desert language 1: 260, 262,  
307; 2: 8, 182, 211, 258  
West(ern) Torres Strait language 1: 12–14,  
23, 53, 55  
Wetan 2: 348  
White Hmong 2: 379, 385–6, 393, 398,  
404–6  
Wirangu 2: 182  
Wiyot 2: 286–7  
Worora 1: 260  
Wunambal 2: 191  
  
Xhosa 2: 5  
  
Yagua 2: 200, 241, 272, 274, 292, 298, 346,  
364  
Yana 1: 227  
Yaqui 2: 211  
Yawuru 1: 52, 56; 2: 63, 151  
Yiddish 2: 10, 17  
  
Yidiñ 1: 50–1, 118, 133, 146–8, 170, 179–80,  
255; 2: 101, 207–8, 259  
classifiers 1: 18, 157; 2: 86  
demonstratives 2: 226, 236, 242, 245  
lexicon 2: 297–9, 307–8  
phonology 1: 206–9, 213, 273, 288  
possession 2: 278, 280, 283, 288, 296, 298,  
302  
transitivity 2: 128, 139, 144, 157  
Yimas 1: 23, 55, 82–3, 162, 181; 2: 9, 23–4,  
169, 178, 199, 353–4, 393–4  
adjectives 2: 63, 74, 76  
Yingkarta 2: 8  
Yir-Yoront 2: 69, 104, 182  
Yokuts 2: 2, 22, 97  
Yoruba 2: 74, 92, 169  
Yuma 1: 140  
Yuman languages 1: 5, 26, 55; 2: 120, 333–4  
Yurok 1: 70, 90  
Yuwaalaraay 2: 180  
  
Zayse 2: 170, 237, 243  
Zekkara 2: 348  
Zoque 2: 8  
Zulu 2: 5  
Zuni 2: 96, 103

# Subject Index

Note that entries in the glossary (which appears in both volumes) are not included in this index.

- A, S and O core arguments 1: 76–7, 98–100, 122–5, 228–9; 2: 116–23, 129–33, 138–40, 147–55, 161–2, 166–70, 292–4
- ablative 1: 132, 145–7, 224, 231–2, 282; 2: 291
- absolutive 1: 76, 122–3; 2: 116–23, 137, 145–52, 165, 167
- accent, *see* stress
- accessibility hierarchy 2: 320–4
- accompaniment 1: 126
- accusative 1: 9, 76, 122–3, 161; 2: 116–23, 147–52, 162, 167–8, 172
- active articulator 1: 269
- active voice 1: 167, 240
- active/stative, *see* split-S marking, fluid-S marking
- Activity type of complement clause 2: 382–421
- addition type of clause linking 1: 134–6
- adjective class 1: 52–3, 112–14, 194–6, 243–5, 304–5; 2: 62–114
- criteria for recognition 2: 70–73
- distinguished from noun class 2: 69, 84–8, 106–7
- distinguished from verb class 2: 77–83, 105–6
- grammatical properties 2: 63–5
- semantic content 2: 73–6
- adjoined relative clause 2: 358–9
- adjunct 1: 101–2
- adposition 1: 73, 127, 224–5, 231–3
- adverb 1: 109, 301; 2: 76, 82, 88
- AFFECT semantic type 1: 104; 2: 127–33, 147, 394
- affective case 2: 151
- affinal kin 1: 6–7
- affix 1: 221–5
- affix(ation) 1: 141–4, 217–8, 269
- AGE semantic type 1: 114; 2: 73–6, 79, 82–5, 104
- Agent semantic role 1: 99, 104–5; 2: 127–33, 153
- agglutinating language 1: 226–7
- agreement 1: 220, 230
- airstream mechanisms 1: 271
- alienable possession 1: 5–7, 11–12, 230; 2: 277–312
- allative 1: 132, 227–8
- allomorph 1: 179, 185
- allophone 1: 180, 264–6
- alphabet 1: 264–5
- also see* orthography
- alternative syntactic frame 1: 98–9, 105–6
- ambitransitive 1: 77–8, 103–4; 165, 305; 2: 100, 124–6, 143–7, 154–7, 300
- analytic language 1: 226–8
- analysis, linguistic 1: 182–99, 243–7
- anaphora 1: 332; 2: 247–61
- ANNOYING semantic type 2: 129, 397
- antipassive 1: 165–8, 172–4, 207–8; 2: 237
- apical place of articulation 1: 267, 276
- applicative 1: 165, 168–71; 2: 123, 186
- apposition, *see* verbless clause, complementation strategies
- archiphoneme 1: 272
- argument, *see* core argument, peripheral argument
- argument identity 1: 175–7
- article 1: 160–1; 2: 51, 55
- articulator 1: 268
- aspect 1: 86, 154, 162–5; 2: 52–4, 177–8, 181–5, 188, 301
- motivating split marking 2: 141
- aspiration 1: 250, 271–2; 2: 9–10
- associated motion affixes 1: 50–1, 180
- atelic 1: 153
- ATTENTION semantic type 1: 104; 2: 127–30, 146–52, 385–413
- Attribution semantic relation 1: 101; 2: 159, 171–84, 188
- augmented term in pronoun system 1: 115; 2: 196–9, 253
- auxiliary 1: 96, 125, 130, 154, 166
- aversive 1: 127
- avoidance style 1: 214–16, 293–5
- base 2: 37
- BEGINNING semantic type 2: 402–17
- benefactive semantic relation 2: 159, 171–84

- beneficiary 1: 108, 126  
bilateral opposition 1: 70, 235–6  
binarism 1: 70–1  
blood relation 1: 6–7  
body part terms 1: 5–7, 22–3, 55, 303  
bound form 1: 145, 217  
bound pronoun 1: 39–40, 55, 82, 116–7,  
125–6, 159; 2: 169, 209–23  
boundedness 1: 153
- cardinal vowels 1: 173–4  
case 1: 12–13, 43–5, 85, 125–6, 164–5, 224–5;  
2: 55, 88  
cataphora 2: 247–61  
causal 2: 291, 362–3  
causative 1: 165, 168–78; 2: 17, 165, 186  
certainty, degree of 1: 153  
changing valency 1: 165–71  
circumfix 1: 141  
classifiers 1: 18, 87, 157–8; 2: 55, 86, 248  
clause 1: 75–6, 93–102, 132–7, 228  
clause linking 1: 94–5, 133–7; 2: 352, 374–5,  
410–14  
clause structure 1: 97–102, 110–12, 254  
click 1: 271, 282; 2: 29  
clitic 1: 221–5; 2: 20–2, 215–8, 254  
co-existing phonological systems 1: 283  
Cogitator semantic role 1: 104–5;  
2: 127–30  
cohesiveness 2: 14–15  
COLOUR semantic type 1: 53, 114, 194, 196,  
304; 2: 73–6, 79, 92–5, 104, 114  
colour terms 1: 256, 291–2  
comitative 1: 145–8  
command 1: 95–6  
common argument in a relative clause  
construction 1: 105, 246; 2: 313–69  
comparative construction 1: 62, 113, 177–9,  
210; 2: 64–6, 71–2, 77, 82, 88–91, 265,  
320, 342, 347, 364  
complement clauses 1: 27, 30–1, 91, 94,  
128–32, 201–3; 2: 171, 185, 361,  
370–411  
grammatical criteria for 2: 375–81  
grammatical parameters for 2: 384–9  
types and meanings 2: 388–94  
complement-taking verbs 2: 253, 370–424  
complementation strategies 1: 83; 2: 351–2,  
405–15  
apposition strategy 2: 409–15  
clause chaining strategy 2: 410–14  
nominalization strategy 2: 408–14  
purposive strategy 2: 399–415  
relative clause strategy 2: 399–414  
serial verb construction strategy  
2: 404–14  
complementary distribution 1: 285  
complementizer 1: 333  
completion 1: 153  
composition 1: 153–4  
compounding 1: 138–9, 304–5; 2: 23, 26,  
56, 155  
concord 1: 230  
condensed relative clause 2: 359–60  
conditional 1: 135–6  
configurational language 1: 72  
congruent/non-congruent, *see*  
conjunction/disjunct contrast  
conjugation 1: 207–12, 239  
conjunction/disjunct contrast 1: 334; 2: 222–3,  
259  
conjunction 1: 134–7  
conjunctive writing system 2: 6  
consanguineal relation 1: 6–7  
consequence type of clause linking  
1: 134–6  
consonant system 1: 7–8, 250, 266–73  
constituent 1: 232  
constituent order 1: 37–8, 71–5, 126, 233–4,  
254–5; 2: 164–5  
construct state 2: 310  
construction 1: 228–9  
content interrogative/question 1: 95–6, 159;  
2: 233–4, 346, 364, 368–9  
*also see* interrogative/indefinite word  
continuous 1: 153  
contrast type of clause linking 1: 134–6  
contrastive focus 1: 174–5  
convenience sample 1: 263  
copula clause 1: 100–1; 2: 66–114, 159, 188,  
300–3  
copula complement 1: 100–1; 2: 66–114, 159,  
188  
copula subject 1: 100–1; 2: 66–114, 159–168  
core argument 1: 97–101, 122–8, 246–7;  
2: 116–59  
*also see* A, S and O  
co-relative construction 2: 356–8  
CORPOREAL semantic type 1: 54, 300; 2: 147,  
394  
correlative construction, *see* co-relative  
construction  
coverb 1: 52, 103, 305  
creole 1: 21–2; 2: 176, 343
- dativ 1: 128; 2: 148–52, 290–1  
DECIDING semantic type 2: 397  
declarative mood 1: 95–7

- definiteness 1: 49, 160–2; 2: 55  
 deictic reference, *see* demonstrative  
 demonstrative 1: 71, 108, 117, 159; 2: 183, 188, 227–47, 346, 364, 369  
   local adverbial demonstrative 2: 224–51  
   manner adverbial demonstrative 2: 224, 229, 233  
   nominal demonstrative 2: 224–51  
   verbal demonstrative 2: 224, 229–31, 242, 251  
 dependencies between grammatical systems 1: 162–5, 181, 255  
 derivation 1: 142–52, 180, 218–21; 2: 15–16, 46–50, 56, 61, 385–6  
 determiner 1: 27–9, 128, 180  
 detransitivizing derivations 1: 165–8, 175–7  
 dictionary 1: 48, 215  
 DIFFICULTY semantic type 2: 74, 76, 95  
 DIMENSION semantic type 1: 53, 114, 194–6, 304; 2: 73–6, 79, 92–5, 104, 114  
 diphthong 1: 198–9  
 direct speech 1: 307–8; 2: 171, 397–8, 419  
 disjunct/conjunct contrast 1: 334; 2: 222–3, 259  
 disjunction type of clause linkage 1: 136–7  
 disjunctive writing system 2: 6  
 dissimilation 1: 270  
 ditransitive, *see* extended transitive  
 Donor semantic role 1: 63–4, 115, 229; 2: 127–8, 134–7  
 double case 1: 45, 56  
 dual number 1: 9–10, 158; 2: 191–217  
 durative 1: 153  
  
 E syntactic function 2: 116–19, 161–2  
 ejective 1: 271, 313  
 elaboration, *see* pronoun elaboration  
 enclitic, *see* clitic  
 environment affecting language 1: 15–17  
 equipollent opposition 1: 236, 272  
 ergativity 1: 76, 82, 86, 89, 91, 123–8, 188–9, 246, 261; 2: 116–23, 147–56, 162, 165, 167, 169, 291  
 essive 2: 170  
 ethics of fieldwork 1: 311  
*Ethnologue* 1: xiii, 73, 91; 2: xiv  
 evaluation 1: 4  
 evidentiality 1: 13, 18, 56, 87, 153, 162–4, 260–1; 2: 260  
 exclusive 1: 335; 2: 194–6, 258  
 Existence semantic relation 2: 160, 174  
 Experiencer semantic role 1: 53–4, 104–5; 2: 127–30, 150–1  
  
 explanation 1: 205–13  
 extended intransitive 1: 99–100, 229; 2: 116–24, 144, 150  
 extended transitive (or ditransitive) 1: 99–100; 2: 116–18, 134  
 extra-language typology 1: 247–8  
 extent (non-spatial setting) 1: 153  
  
 Fact type of complement clause 2: 380–421  
 feminine  
   as unmarked gender 1: 240  
   *also see* noun classes, gender  
 fieldwork 1: 297–9, 209–30  
 finite 1: 80, 91  
 fluid-S marking 1: 77–8, 124–5; 2: 121, 126, 141  
 focal clause 1: 133–6  
 focus 1: 174–5  
 focus system 2: 52  
 foot 1: 148, 206  
 formal markedness 1: 237–40  
 formal theories 1: 3–4, 183–4  
 fourth person pronoun 1: 260–1; 2: 203–5  
 free form 1: 145, 217  
 fricative 1: 269  
 functional markedness 1: 237–40  
 fused relative clause 2: 356–60  
 fusion of morphemes 2: 215–6  
 fusional language 1: 226–7  
 future time marking 1: 154  
  
 gender 1: 12–13, 43, 87, 155–8, 180, 290, 335; 2: 54–5, 86–7, 200–222, 246, 259, 297  
   *also see* noun classes  
 generic noun 1: 300–2  
 genitive 1: 44–5, 73; 2: 123, 148–50, 167, 268–312  
 Gift semantic role 1: 53, 104–6, 229; 2: 127, 134–7  
 GIVING semantic type 1: 104–6; 2: 127–8, 134–7, 145–6, 157, 394  
 glossing, conventions for 1: 61, 216  
 glottalic airstream mechanism 1: 271  
 glottalization 1: 280  
 government 1: 231  
 gradual opposition 1: 236, 272  
 grammatical word 1: 93, 108, 116, 138, 221–4; 2: 1–36  
   criteria for 2: 12–19  
  
 half-conjunction 1: 134–6  
 harmonic pronoun 1: 17  
 'have' 2: 290, 298–302

- head (of noun phrase) 1: 229–32; 2: 296–8  
heterorganic 1: 276  
historical explanation 1: 62–3, 205, 209–13  
homonymy, multiple 1: 290  
homorganic 1: 198, 276  
honorific pronoun 1: 17–18  
HUMAN PROPENSITY semantic type 1: 53,  
114, 304; 2: 73–6, 79, 81, 92–5, 104, 114,  
146
- iconicity 2: 289–90  
ideophone 1: 302; 2: 30  
Identity semantic relation 1: 101; 2: 159,  
170–83, 187  
immersion fieldwork 1: 317–25  
imperative 1: 13, 77, 95–7; 2: 78, 154, 181,  
185–6  
imperfect 1: 153  
imperfective 1: 154  
impersonal form 2: 204–5  
Impression semantic role 1: 104–5;  
2: 127–30, 150–1  
inalienable possession 1: 5–7, 11–12, 230;  
2: 277–312  
inclusive 2: 194–6, 258  
inclusory construction, *see* pronoun  
elaboration  
indefinite form 2: 205  
index of comparison 1: 177–8  
indicative mood, *see* declarative mood  
indirect speech 2: 397–8, 419  
inductive basis for generalizations 1: 2,  
184h  
infinitive 1: 79–80  
infix 1: 141, 146–7, 180, 217; 2: 52, 215, 259,  
305  
inflection 1: 96–7, 142–7, 203, 213–21;  
2: 15–16  
ingressive airstream mechanism 1: 271  
inner locative 1: 119  
instrument(al) 1: 98–9, 126–8, 188–9;  
2: 121–3, 170  
interjection 1: 283; 2: 27–30, 36  
internal change 1: 140  
interrogative mood 1: 95–7; 2: 186  
interrogative/indefinite word 1: 17–18, 216;  
2: 233–4  
*also see* content interrogative/question  
intonation 1: 24, 67–8, 75, 95–6, 133, 137,  
187–8, 283  
intra-language typology 1: 247–8  
intransitive, *see* transitivity  
irrealis 1: 153
- isolating language 1: 226–7  
item-and-arrangement model 1: 146  
item-and-process model 1: 146
- Judgment to complement clause 1: 31–33,  
42; 2: 395
- karaka 1: 45  
kin term 1: 5–7, 26, 256, 262, 300; 2: 40, 47,  
262–6, 278, 298
- labile, *see* ambitransitive  
laminal place of articulation 1: 267, 276  
language contact 1: 15, 83, 209, 283; 2: 205,  
220  
language planning 1: 20  
lateral 1: 269, 275–6  
lexicon (and lexeme) 1: 47–54, 214–7, 253–5,  
289–308; 2: 4–5  
lexicostatistics 1: 215, 240  
LIKING semantic type 1: 32–4, 104–5;  
2: 127–30, 148–50, 157, 397–413  
liquid 1: 269  
Location semantic relation 1: 101; 2: 159–61,  
169, 172, 176, 179, 180–8  
locative 1: 128; 2: 151  
locutor/non-locutor, *see* conjunct/disjunct  
contrast  
logophoric pronoun 2: 252–4, 419  
long vowel 1: 196–9, 209–13, 281–2
- Manip semantic role 1: 99, 104–5; 2: 127–33  
manner of articulation 1: 269–71, 276  
mark of comparison 1: 177–8  
marked nominative 2: 167–8, 172  
markedness 1: 235–41, 272–3; 2: 120, 137–40,  
166–8, 181, 246–7  
masculine  
as unmarked gender 1: 237  
*also see* noun classes, genders  
medial clause 2: 410  
Medium semantic role 1: 53, 104–6; 2:  
127–31  
Message semantic role 1: 53, 104–6; 2:  
127–31, 136  
middle 1: 235  
minimal pair 1: 273  
minimal term in pronoun system 1: 115;  
2: 196–9, 258  
modal verb 1: 109; 2: 402  
modality 1: 96, 153; 2: 52–4, 301  
modifier 1: 106–8, 230  
mood 1: 95–6



- mora 1: 197–8, 281–2  
 morpheme 1: 145, 180  
 morphological processes 1: 83–4, 138–44, 217–8; 2: 53  
 morphology 1: 85–6, 89, 138–52, 180, 217–21  
 ‘mother-in-law language’, *see* avoidance style  
 MOTION semantic type 1: 119; 2: 145–6, 394  
 multilateral opposition 1: 70–1, 235–6  
 mutation 2: 12  
  
 naming 2: 173–4  
 narrative case 2: 168  
 nasal 1: 269, 276  
 nasalization 2: 9–10  
 negation 1: 83, 137–8; 2: 57, 186, 385, 401–2  
 negative copula 2: 178–9  
 neutralization 1: 164, 272; 2: 199–200, 217  
 nominal hierarchy 1: 123, 179; 2: 137–41  
 nominalization 1: 129–31, 150–1; 2: 263–7, 292, 326, 363, 403–14  
 nominative 1: 76, 122–3; 2: 116–23, 147–52, 167–72  
 non-canonical marking of core arguments 1: 127; 2: 147–52  
 non-canonical relative clauses 2: 356–61  
 non-configurational languages 1: 72  
 non-nuclear verb 1: 293–5  
 non-restrictive relative clause 2: 314–5, 352–4  
 non-spatial setting 1: 152–5  
 noun 1: 51–3, 102–3, 302–4; 2: 37–61  
   criteria for 2: 38–41  
   grammatical categories associated with 2: 54–5  
 noun classes (and genders) 1: 12–13, 27–31, 87, 155–8, 163–5; 2: 54–5, 86–9, 239, 245–6  
 noun incorporation 1: 68–9, 89; 2: 155, 310–11  
 noun phrase (NP) 1: 106–8, 229–30; 2: 50–51, 68–9, 85  
 nuclear verb 1: 293–5  
 number system 1: 9–11, 43, 87, 143, 153–9, 163–5, 184–5, 237; 2: 55, 87, 154, 191–2, 246  
 numeral (number) 1: 175, 207, 301; 2: 72, 93  
  
 O, S and A core arguments 1: 76–7, 98–100; 122–5, 228–9; 2: 116–23, 129–33, 139–40, 147–55, 161–2, 166–70, 292–4  
 object incorporation, *see* noun incorporation  
  
 obstruent 1: 269  
 omission of copula verb 2: 180–2  
 onomatopoeia 1: 68–9; 2: 30  
 oppositions 1: 70–1, 236, 272–3  
 orthography 1: 66–7, 90, 286  
   *also see* alphabet  
 outer locative 1: 119  
 overlapping between word classes, semantic 2: 99–103  
  
 parameter of comparison 1: 177–9  
 particle 2: 37  
 partitive 2: 148–50, 158, 167  
 passive articulator 1: 268  
 passive voice 1: 165–8, 172–4, 240–1, 245–6; 2: 136–7, 183, 373–4, 377–8  
 paucal number 1: 9–10, 158; 2: 101–2, 199, 202  
 pause 2: 9–10, 18–26, 35  
 Perceiver semantic role 1: 53–4, 104–5; 2: 117, 127–30, 150–3  
 perfect 1: 153  
 perfective 1: 153–4  
 peripheral argument 1: 97–101, 126–8, 189–91; 2: 116  
 peripheral place of articulation 1: 270, 276  
 person system 1: 115–16, 163–5; 2: 191–2, 245–6  
   *also see* pronoun  
 pertensive 2: 268–312  
 phase of activity 1: 153, 155  
 phoneme 1: 88, 145, 264–6, 287  
 phonoaesthesia, *see* sound symbolism  
 phonological rules 2: 11–12  
 phonological word 1: 93, 197–200, 206–7, 221–5, 249–51, 267; 2: 1–36  
   criteria for 2: 7–12  
 phonology 1: 200–1, 216–7, 249–51, 264–88  
 phonotactics 1: 273–9, 290  
 phrasal verbs in English 1: 35–7, 67, 290; 2: 5, 395  
 phrase 1: 106–10, 254–5  
 PHYSICAL PROPERTY semantic type 1: 53, 114, 202, 205, 314; 2: 73–6, 79, 92–5, 104, 114  
 pidgin 1: 21  
 pivot 1: 168, 172–5, 235; 2: 237, 321, 411–12  
 place of articulation 1: 268–9, 276  
 plural number 1: 9–10, 158–9; 2: 191–261  
 polar question/interrogative 1: 95–6; 2: 166  
 polarity 1: 138, 162–5  
 politeness strategies 1: 17–18; 2: 201–3  
 polysynthetic 1: 226–8, 241

- portmanteau form 2: 215–6
- Possessed, parameters relating to 2: 277–90
- possession 1: 5–7, 11–12; 2: 51, 55, 219–22, 262–312
- Possession semantic relation 2: 171, 174–81
- possessive phrase 1: 107, 229–30
- possessive relation, nature of 2: 274–7
- Possessor, parameters relating to 2: 271–4
- possible consequence type of clause  
linking 1: 152
- postposition, *see* adposition
- posture verb 1: 257, 262, 307
- Potential type of complement clause  
2: 392–421
- pragmatics 1: 68, 95–7, 246, 325, 327; 2: 61, 109, 119, 161, 172–3, 179, 332, 334, 352, 356, 398, 403
- predicate 1: 78–9, 98–101; 2: 40–1, 51–2, 77–8, 162–4
- prediction 1: 4
- prefix, *see* affix(ation)
- preposition, *see* adposition
- preverb 2: 5
- primary sources, need to refer to  
1: 64–6
- primary verbs 1: 54, 131; 2: 394
- primitive languages, lack of 1: 21
- privative marking 2: 11
- privative opposition 1: 236, 272
- proclitic, *see* clitic
- progressive 1: 153
- pronoun (personal) 1: 115–17; 2: 189–223, 247–61  
harmonic 1: 17  
honorific 1: 17–18  
*also see* bound pronoun
- pronoun elaboration 2: 207–10, 259
- proper noun/name 1: 102, 108
- prosody 1: 251, 279–83; 2: 10–11
- prototypical pattern 1: 4–9
- psychological reality of phonological  
units 1: 26; 2: 26, 31
- pulmonic airstream mechanism 1: 271
- punctual 1: 153
- purposive construction 1: 96–7, 124, 136;  
2: 142, 291, 399, 415
- QUALIFICATION semantic type 2: 74
- QUANTIFICATION semantic type 2: 74, 76, 93, 95
- quantifier 1: 107, 301; 2: 50–1, 59, 81, 101, 189, 230
- question 1: 95–6
- realis 1: 153
- reality 1: 153
- Recipient semantic role 1: 53, 104–6, 126–8, 229; 2: 127, 134–7
- reciprocal 1: 176–7
- recursion 2: 16–17
- reduplication 1: 139–40, 180, 253–4, 262, 294; 2: 13–16, 24–7, 55–6, 207  
for distinguishing between word  
classes 2: 65–6, 69, 78–9, 87, 114
- reflexive 1: 77, 176–7; 2: 154, 185, 347
- relative clause 1: 23–4, 89, 93–4, 107, 120, 186–8; 2: 81, 142, 185, 313–69, 374–5, 399–414  
function of 2: 350–2  
marking of 2: 338–48  
meaning of 2: 352–4  
non-canonical 2: 356–61  
restrictive and non-restrictive 2: 314–6, 352–4  
structure of 2: 348–50
- relative pronoun 2: 344–8, 363–4
- REST semantic type 1: 119; 2: 145–6, 394
- restrictive relative clause 2: 314–6, 352–4
- retroflex 1: 268, 279–80; 2: 10
- rhotic 1: 248, 265–6, 269–71, 284–6, 288
- root 1: 138–48, 199–200, 217–8
- rounding 1: 273–4
- S, A and O core arguments 1: 76–7, 98–100; 122–5, 228–9; 2: 116–23, 129–33, 139–40, 147–55, 161–2, 166–70, 292–4
- S = A type ambitransitive, *see*  
ambitransitive
- S = O type ambitransitive, *see*  
ambitransitive
- sampling 1: 257–63
- sandhi rules 2: 12
- science, linguistics as a branch of  
1: 1–4
- scrambling rule 1: 72
- secondary concept 1: 50, 131; 2: 394–5, 399–400
- secondary predicate 1: 41–3
- secondary verb 1: 54, 131; 2: 394–414
- segmental features as criteria for  
phonological word 2: 7–10
- self repair 2: 19, 30
- semantic overlap between word classes  
2: 99–103
- semantic roles 1: 53–4, 104–6; 2: 127
- semantic types 1: 31–3, 53–4, 102–6, 300;  
2: 127, 394–421

- semi-vowel 1: 269–71, 275, 284  
 sentence 1: 75–6, 91, 132–7  
 serial verb construction 1: 83, 109, 132, 139, 178, 290, 305; 2: 52, 58–60, 69, 82, 136, 404–21  
 shifter 1: 114; 2: 189  
     *also see* pronoun, demonstrative, temporal item  
 sign languages 1: 90; 2: 12  
 SIMILARITY semantic type 2: 74, 76  
 singular number 1: 9–10, 158–9  
 social niceties 2: 201–2  
 sonorant 1: 269  
 sound symbolism 1: 68–70, 90; 2: 242  
 spatial item 1: 107, 118–22, 152  
 spatial shifter 1: 114  
 Speaker semantic role 1: 53–4, 104–6; 2: 127–31  
 SPEAKING semantic type 1: 104–6; 2: 127–31, 146, 397–413  
 SPEED semantic type 2: 73–6, 79, 93–5, 104  
 split ergativity 1: 122–6; 2: 137–42, 206, 208, 212, 234  
 split-S marking 1: 77–8, 82, 124–5; 2: 120, 126, 140–1, 155–6, 221  
 stance verb 2: 161, 182–8  
 standard error in sampling 1: 258–9  
 standard of comparison 1: 177–9  
 statement 1: 95–6  
 stative/active, *see* split-S marking, fluid-S marking  
 stem 1: 138–42  
 Stimulus semantic role 1: 53, 104–5; 2: 127–30, 150–1  
 stop 1: 269, 276  
 stress (or accent) 1: 251, 280–3; 2: 10–12  
 stress shift 1: 14, 160–1, 180  
 subject 1: 76–7, 98, 229  
 subjunctive 1: 97; 2: 183, 391  
 substitution anaphora/cataphora 2: 248–50  
 subtraction 1: 140–1  
 suffix, *see* affix(ation)  
 supporting clause 1: 133–6  
 suprasegmental 1: 279, 297; 2: 10  
 switch-reference marking 1: 82, 174; 2: 185  
 syllable 1: 249, 277–9  
 syllable structure 1: 9, 249–50, 275–9  
 symbolic type of morphological process 1: 226  
 synharmonism 2: 11  
 synthetic language 1: 226–8  
 systems, grammatical 1: 247, 252  
 tabooing 2: 31  
 Target semantic role 1: 99, 104–5; 2: 127–33  
 telic 1: 153  
 temporal item 1: 107, 114, 118–22, 248; 2: 107, 222, 242  
 temporal type of clause linking 1: 134–5  
 tense 1: 12–14, 154–5, 162–5, 239; 2: 52–4, 181–4, 301  
     motivating split marking 2: 141  
 textual anaphora/cataphora 2: 248–50  
 thesaurus 1: 296–9, 308, 319  
 THINKING semantic type 1: 32–4, 104; 2: 127–30, 146–7, 396–413  
 third person 2: 189–90, 203  
 Thought semantic role 1: 104; 2: 127–30  
 time word, *see* temporal item  
 tones 1: 140, 251, 279; 2: 10–12, 178  
 topic 1: 171–5, 234–5; 2: 172  
 topicalization 1: 235  
 traditional grammar 1: 7, 114; 2: 123, 225, 369  
 transitivity 1: 89, 103–5, 165, 168–71; 2: 115–58  
 tree structure 1: 48, 292  
 trial number 1: 9–10, 158  
 tripartite marking of S, A and O 1: 123–4; 2: 118–9, 139, 154–7  
 typology, linguistic 1: 242–63  
 unaccusative 2: 155–6  
 unergative 2: 155–6  
 unit augmented term in pronoun system 2: 180–5, 294  
 unmarked, *see* formal markedness, functional markedness  
 valency-changing derivation 1: 165–71, 227, 235–6  
 VALUE semantic type 1: 114; 2: 73–6, 79, 92–5, 104, 114  
 verb 1: 52–4, 103, 305–6; 2: 37–61  
     criteria for 2: 38–41  
     grammatical categories associated with 2: 52–4  
 verb phrase 1: 108–10  
 verbalization 1: 150–1  
 verbless clause 1: 341; 2: 160–188  
 verbless clause complement 1: 341; 2: 160–188  
 verbless clause subject 1: 341; 2: 160–188  
 vicious circles in definitions 1: 292

- visibility 2: 243–4  
vocabulary, *see* lexicon  
voice system 1: 167, 273–5  
voicing 1: 271–2  
vowel harmony 1: 142, 251, 274, 279;  
2: 7, 10, 12, 20, 24, 32–3  
vowel system 1: 7–8, 249–50
- WANTING semantic type 2: 403–13  
word 1: 92–3, 223–4; 2: 1–36  
definitions of 2: 1–5, 34  
word classes, recognition of 1: 25–7, 102,  
191–3; 2: 38–41
- word classes and clause structure 1: 110–12;  
2: 41–50  
word order 1: 37–8, 71–5, 233–4; 2: 15  
*also see* constituent order  
word-and-paradigm model 1: 144  
writing a grammar 1: 57–63  
writing systems 2: 5–6
- yes/no question, *see* polar  
question/interrogative
- zero 1: 143–4  
zero derivation 2: 46–50

**Books by R. M. W. Dixon**

## ON LINGUISTICS

Linguistic science and logic

What *is* language? A new approach to linguistic description

The Dyirbal language of North Queensland

A grammar of Yidij

The languages of Australia

Where have all the adjectives gone? and other essays in semantics and  
syntax

Searching for Aboriginal languages, memoirs of a field worker

A grammar of Boumaa Fijian

Words of our country: stories, place names and vocabulary in Yidiny,  
the Aboriginal language of the Cairns-Yarrabah region

Ergativity

The rise and fall of languages

Australian languages: their nature and development

The Jarawara language of southern Amazonia

A semantic approach to English grammar

*with Bruce Moore, W. S. Ramson, and Mandy Thomas*

Australian Aboriginal words in English, their origin and meaning

*with Grace Koch*Dyirbal song poetry, the oral literature of an Australian rainforest  
people

## ON MUSIC

*with John Godrich*

Recording the blues

*with John Godrich and Howard Rye*

Blues and gospel records, 1890–1943

NOVELS (*under the name Hosanna Brown*)

I spy, you die

Death upon a spear

## EDITOR OF BOOKS ON LINGUISTICS

Grammatical categories in Australian languages

Studies in ergativity

*with Barry J. Blake*

Handbook of Australian languages, Vols 1–5

*with Martin Duwell*

The honey ant men's long song, and other Aboriginal song poems

Little Eva at Moonlight Creek: further Aboriginal song poems

*with Alexandra Y. Aikhenvald*

The Amazonian languages

Changing valency: case studies in transitivity

Areal diffusion and genetic inheritance: problems in comparative linguistics

Word, a cross-linguistic typology

Studies in evidentiality

Adjective classes: a cross-linguistic typology

Serial verb constructions: a cross-linguistic typology

Complementation: a cross-linguistic typology

Grammars in contact: a cross-linguistic typology

The semantics of clause linking: a cross-linguistic typology

*with Alexandra Y. Aikhenvald and Masayuki Onishi*

Non-canonical marking of subjects and objects