



Maggie Tallerman

third edition

Understanding
Syntax

UNDERSTANDING LANGUAGE SERIES

Understanding

Syntax

Third edition

Understanding Language series

Series editors:

Bernard Comrie, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
Greville Corbett, Surrey Morphology Group, University of Surrey, UK

The Understanding Language series provides approachable, yet authoritative, introductions to major topics in linguistics. Ideal for students with little or no prior knowledge of linguistics, each book carefully explains the basics, emphasising understanding of the essential notions rather than arguing for a particular theoretical position.

Understanding Child Language Acquisition (forthcoming)

Caroline Rowland

ISBN 978 1 444 15265 4

Understanding Language Testing

Dan Douglas

ISBN 978 0 340 98343 0

Understanding Morphology Second Edition

Martin Haspelmath

Andrea D. Sims

ISBN 978 0 340 95001 2

Understanding Phonetics

Patricia Ashby

ISBN 978 0 340 92827 1

Understanding Phonology Third Edition

Carlos Gussenhoven

Haike Jacobs

ISBN 978 1 444 1204 7

Understanding Pragmatics

Jef Verschueren

ISBN 978 0 340 64623 6

Understanding Second Language Learning

Lourdes Ortega

ISBN 978 0 340 90559 3

Understanding Semantics

Sebastian Löbner

ISBN 978 0 340 73198 7

Understanding Sociolinguistics (forthcoming)

Enam Al-Wer

ISBN 978 0 340 80774 3

For more information on any of these titles, or to order, go to www.hoddereducation.com

Understanding

Syntax

Third edition

Maggie Tallerman



HODDER
EDUCATION

AN HACHETTE UK COMPANY

First edition published in 1998 by Hodder Arnold
Second edition published in 2005 by Hodder Arnold
This edition published by Hodder Education

© 2011 Maggie Tallerman

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronically or mechanically, including photocopying, recording or any information storage or retrieval system, without either prior permission in writing from the publisher or a licence permitting restricted copying. In the United Kingdom such licences are issued by the Copyright Licensing Agency: Saffron House, 6–10 Kirby Street, London EC1N 8TS.

Hachette UK's policy is to use papers that are natural, renewable and recyclable products and made from wood grown in sustainable forests. The logging and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

The advice and information in this book are believed to be true and accurate at the date of going to press, but neither the author[s] nor the publisher can accept any legal responsibility or liability for any errors or omissions.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN: 978 1 444 11205 4

Impression number 10 9 8 7 6 5 4 3 2 1

Year 2021 2020 2019 2018 2017 2016 2015 2014 2013 2012 2011

Cover image of two cornflowers (close up) © Tomashko/Fotolia.com

Typeset in 11 on 12pt Minion by Phoenix Photosetting

Printed and bound in the UK by CPI Antony Rowe

What do you think about this book? Or any other Hodder Education title?
Please send your comments to educationenquiries@hodder.co.uk

*For S.J. and Maggie the younger,
the linguists in the family*

This page intentionally left blank

Contents

Acknowledgements	xi
Note to the instructor	xii
Note to the student	xiii
List of abbreviations used in examples	xv
List of tables	xvi
1 What is syntax?	1
1.1 Some concepts and misconceptions	1
1.1.1 What is the study of syntax about?	1
1.1.2 Language change	7
1.2 Use of linguistic examples	11
1.2.1 Why not just use examples from English?	11
1.2.2 How to read linguistic examples	13
1.3 Why do languages have syntax?	19
1.3.1 Word order	19
1.3.2 Promotion and demotion processes	22
1.3.3 All languages have structure	24
Further reading	26
Exercises	26
2 Words belong to different classes	32
2.1 Identifying word classes	32
2.1.1 How can we tell that words belong to different classes?	32
2.1.2 Starting to identify nouns, adjectives and verbs	33
2.1.3 An illustration: How do speakers of a language identify word classes?	36
2.2 Verbs	39
2.2.1 An introduction to verb classes	39
2.2.2 Verbs and their grammatical categories	41
2.3 Nouns	46
2.3.1 Semantic roles for noun phrases	46
2.3.2 Syntactic roles for noun phrases	47
2.3.3 Nouns and their grammatical categories	50
2.3.4 Nouns, definiteness and determiners	53
2.4 Adjectives	56
2.4.1 Positions and functions of adjectives	56
2.4.2 Adjectives and intensifiers	57
2.4.3 Adjectives and their grammatical categories	58
2.4.4 Are adjectives essential?	58

2.5	Adverbs	61
2.5.1	Adverbs and adjectives	61
2.5.2	The adjunct function	63
2.6	Prepositions	63
2.6.1	Identifying prepositions in English	63
2.6.2	Postpositions	65
2.6.3	Grammatical categories for adpositions	65
2.7	Conclusion	66
	Further reading	67
	Exercises	67
3	Looking inside sentences	73
3.1	Finiteness and auxiliaries	73
3.1.1	Independent clauses	73
3.1.2	Finiteness	74
3.1.3	Main verbs and verbal auxiliaries	76
3.1.4	Ways to express the grammatical categories for verbs	78
3.1.5	Non-finite verbs	80
3.1.6	Co-ordination of clauses	83
3.1.7	Summary	84
3.2	Introduction to subordination	84
3.2.1	Complement clauses	84
3.2.2	Adjunct or adverbial clauses	87
3.2.3	Identifying subordinate clauses	88
3.2.4	Root and subordinate clauses: Some distinctions	89
3.2.5	Some cross-linguistic variation in subordination	92
3.2.6	Summary: Properties of subordinate clauses and root clauses	93
3.3	Major cross-linguistic variations	94
3.3.1	The co-ordination strategy	94
3.3.2	Nominalization	95
3.3.3	Serial verbs	96
3.3.4	Summary	99
	Further reading	99
	Exercises	100
4	Heads and their dependents	108
4.1	Heads and their dependents	108
4.1.1	What is a head?	108
4.1.2	The influence of heads on their dependents	109
4.1.3	Summary: The properties of heads	111
4.1.4	More about dependents: Adjuncts and complements	111
4.1.5	More about verb classes: Verbs and their complements	114
4.1.6	Other heads and their complements	116
4.1.7	Summary: The main properties of complements vs. adjuncts	117
4.1.8	Is the noun phrase really a determiner phrase?	118
4.1.9	Phrases within phrases	119

4.2	Where does the head occur in a phrase? Head-initial and head-final languages	120
4.2.1	Head-initial languages	120
4.2.2	Head-final languages	121
4.2.3	An exercise on head-initial and head-final constructions	122
4.3	Head-marking and dependent-marking languages	123
4.3.1	Definitions and illustrations: Syntactic relationships between heads and dependents	123
4.3.2	Head preposition/postposition and its NP object	124
4.3.3	The clause: A head verb and the arguments of the verb	126
4.3.4	Head noun and dependent possessor NP	128
4.3.5	Head noun and dependent AP	129
4.3.6	An exercise on head-marking and dependent-marking	130
4.3.7	Some typological distinctions between languages	132
4.3.8	Summary	134
	Further reading	134
	Exercises	135
5	How do we identify constituents?	141
5.1	Discovering the structure of sentences	141
5.1.1	Evidence of structure in sentences	141
5.1.2	Some syntactic tests for constituent structure	143
5.1.3	Introduction to constituent structure trees	148
5.1.4	Summary	153
5.2	Relationships within the tree	153
5.3	Developing detailed tree diagrams and tests for constituent structure	156
5.3.1	Verb classes and constituent structure tests	157
5.3.2	The co-ordination test for constituency	163
5.3.3	Do all languages have the same constituents?	165
5.4	Summary	166
	Further reading	167
	Exercises	167
6	Relationships within the clause	172
6.1	Indicating grammatical relations in the clause	172
6.2	Order of phrases within the clause	173
6.2.1	Basic and marked orders	173
6.2.2	Statistical patterns	175
6.3	Case systems	177
6.3.1	Ways of dividing core arguments	177
6.3.2	Nominative/accusative systems	179
6.3.3	Ergative/absolutive systems	180
6.3.4	Split systems I	182
6.3.5	Marked and unmarked forms	184
6.4	Agreement and cross-referencing	184
6.4.1	What does verb agreement involve?	184
6.4.2	Nominative/accusative agreement systems	185
6.4.3	Ergative/absolutive agreement systems	188

6.4.4	Split systems II	189
6.5	Grammatical relations	190
6.5.1	Investigating core grammatical relations	190
6.5.2	Subjects: Typical cross-linguistic properties	190
6.5.3	An examination of subjects in specific languages	192
6.5.4	Objects	199
6.6	Free word order: A case study	201
6.7	Summary	204
	Further reading	204
	Exercises	205
7	Processes that change grammatical relations	211
7.1	Passives and impersonals	211
7.1.1	The passive construction and transitive verbs	211
7.1.2	The impersonal construction	216
7.2	The antipassive	217
7.2.1	Basic facts	217
7.2.2	Primary grammatical relations and grammatical pivots	220
7.3	The applicative construction	225
7.4	The causative construction	229
7.5	Summary	233
	Further reading	233
	Exercises	234
8	<i>Wh</i>-constructions: Questions and relative clauses	243
8.1	<i>Wh</i> -questions	243
8.1.1	Languages with <i>wh</i> -movement	243
8.1.2	Languages with <i>wh</i> -in-situ <i>wh</i> -questions	247
8.1.3	Multiple <i>wh</i> -questions	249
8.2	Relative clauses	251
8.2.1	Relative clauses in English	251
8.2.2	Cross-linguistic variation in relative clauses	254
8.3	Focus movements and scrambling	259
8.4	Some conclusions	261
	Further reading	262
	Exercises	262
9	Asking questions about syntax	271
9.1	Syntactic description: What questions to investigate	271
9.2	A case study: Grammatical sketch of colloquial Welsh	274
9.3	Some questions concerning syntax	281
9.4	Last words: More syntax ahead	284
	Sources of data used in examples	287
	Glossary	291
	References	296
	Language index	304
	Subject index	307

Acknowledgements

Over the 13 years or so since the first edition of this book was published, I have been overwhelmed by the interest shown in the material it presents, and by the kindness of very many people from around the world. I have received dozens of emails, often from complete strangers, volunteering corrections to data, offering new data, suggesting ways in which the book could be improved, discussing fine linguistic points at great length, offering to read drafts of new material, and generally providing constructive criticism. Doubtless, I have overlooked some of you in the list that follows; for this, I heartily apologise, and I stress my genuine gratitude to all who helped make this third edition a better text. Many thanks, then, to the following colleagues, friends and students whose real and virtual presence has helped so much in the writing of the first, second and third editions of this textbook: Muteb Alqarni, Clayton Ashton, Seiki Ayano, Ute Bohnacker, Bob Borsley, Siobhan Casson, Zedric Dimalanta, Joe Emonds, Tom Ernst, Stuart Forbes, Don Frantz, Anders Holmberg, Chris Johns, Andreas Kathol, Jagdish Kaur, Daniela Kolbe, Lan Yin Kong, Nedzad Leko, Joan Maling, Anna Margetts, Jenny Marjoribanks, Roger Maylor, Sadat Peyambar, Tenzin Rigzin, Caroline Gray Robinson, Stuart Payton Robinson, Anna Siewierska, Carlota S. Smith, Rex Sprouse, Maite Taboada, Höski Thráinsson, Graham Thurgood, Antoine Trux, Ian Turner, Robert D. Van Valin, Nigel Vincent, Emiel Visser, Stephen M. Wechsler, Ian Woo and Monaliza Sarbini Zin. None of the above should be held responsible for any remaining errors.

I also owe a great debt of thanks to the series editors, Bernard Comrie and Grev Corbett, who improved this work in immeasurable ways. I hope that this new edition will be of credit to both these linguists, because their own work has inspired me throughout. Of course, full credit for any shortcomings remains with the author.

Finally, especial thanks to my husband, S. J. Hannahs, for massive support, both practical and moral, for reading and commenting on drafts, and for generally putting up with me during the preparation of this edition. Since the first edition was published, our family has gained another linguist. This third edition is thus dedicated to the linguists in the family, S. J. and Maggie the younger.

Note to the instructor

Changes to the third edition

If you have used this textbook before, you will find that this new edition contains essentially the same material as the second edition, but that every chapter has been substantially revised, and there is also some significant new material. I hope to have improved the clarity of discussion and level of explanation for all the most complex concepts that are introduced, and I have attempted to anticipate more precisely the needs of the beginning student with no background whatever in language studies. Please let me know if you find this helpful.

The organization of material has changed quite a lot, in Chapters 2 and 3 especially. Also, I have introduced more subsection headings, so that material is easier to find, and so that each section does not become too long. Chapters 6 and 7 remain highly demanding, but I am hopeful that the attentive reader can follow the discussion, and I have given more pointers and reminders of earlier discussions than in the previous edition.

Many new exercises have been added, and some of the old ones removed, in cases where I felt that they didn't work too well. Importantly, the exercises have now been tailored to fit more closely the contents of the chapter that they belong to. This means that exercises have been moved around a great deal since the 2005 edition, so please do look in the language index if you can't find an exercise which you previously found useful.

Chapter 9 has undergone substantial revision and expansion. It now contains a section suggesting a set of questions that a student might investigate in order to produce a basic sketch of the syntax and morphosyntax of a language. Following this is a case study of colloquial Welsh, which illustrates what such a sketch might look like. There is also a discussion in this chapter of some broader questions concerning the human language faculty, especially in light of some recent debates within linguistics.

As always, I'd be glad to hear from any instructors about the success or otherwise of any of the changes I've made, and I'm also happy to receive data corrections and suggestions for further improvements.

*Maggie Tallerman
Newcastle University
March 2011*

exercises, please email me and I will try to help by suggesting a strategy, but I won't tell you the answers unless I can be sure that I'm not giving away the answers to a set assignment! For that reason, students should ask their instructors to email. My email address is: maggie.tallerman@ncl.ac.uk.

I will also be happy to receive corrections to data or to claims I make about any language, or further illustrations, or suggestions for new exercises.

*Maggie Tallerman
Newcastle University
March 2011*

List of abbreviations used in examples

1	first person	INTRANS	intransitive
2	second person	M, MASC	masculine
3	third person	NEG	negative
ABS	absolutive	NOM	nominative
ACC	accusative	NONPAST	nonpast tense
APPLIC	applicative	OBJ	object (or object marker)
AUX	auxiliary	OBL	oblique (case)
CAUS	causative	PAST	past tense
COMP	complementizer	PERF	perfect (tense/aspect)
CONJ	conjunction	PERFCTV	perfective (tense/aspect)
DEF	definite	PL	plural
DEF ART	definite article	PN	proper noun marker
DEM	demonstrative	POSS	possessive marker
DET	determiner	PRED	predicate marker
ERG	ergative	PRES	present tense
EXC	exclusive	PROG	progressive
F, FEM	feminine	PRT	particle
FUT	future tense	PUNC	punctual (aspect)
GEN	genitive	RM	relative clause marker
IMPER	imperative	SEQ	sequential
IMPF	imperfect (tense/aspect)	SG	singular
INC	inclusive	SJTV	subjunctive (mood)
INDEF	indefinite	SM	subject marker
INDIC	indicative mood	SU	subject (or subject marker)
INFIN	infinitive	TRANS	transitive marker

List of tables

Table 1.1	Present tense of French verb <i>parler</i> , ‘to speak’	16
Table 1.2	Glosses for person and number	18
Table 4.1	Syntactic relationships between a head and its dependent	123
Table 6.1	The core arguments	177
Table 6.2	The major case systems	179
Table 7.1	Accusative and ergative case systems	215
Table 7.2	Primary grammatical relations	220
Table 9.1	Inflectional paradigm for the Welsh preposition <i>wrth</i> , ‘at’	276
Table 9.2	Inflectional paradigm for the past tense of the Welsh verb <i>gweld</i> , ‘see’	278

What is syntax?

1.1 SOME CONCEPTS AND MISCONCEPTIONS

1.1.1 What is the study of syntax about?

This book is about the property of human language known as syntax. ‘Syntax’ means ‘sentence construction’: how words group together to make phrases and sentences. Some people also use the term GRAMMAR to mean the same as syntax, although most linguists follow the more recent practice whereby the grammar of a language includes all of its organizing principles: information about the sound system, about the form of words, how we adjust language according to context, and so on; syntax is only one part of this grammar.

The term ‘syntax’ is also used to mean the *study* of the syntactic properties of languages. In this sense, it’s used in the same way as we use ‘stylistics’ to mean the study of literary style. We’re going to be studying how languages organize their syntax, so the scope of our study includes the classification of words, the order of words in phrases and sentences, the structure of phrases and sentences, and the different sentence constructions that languages use. We’ll be looking at examples of sentence structure from many different languages in this book, some related to English and others not. All languages have syntax, although that syntax may look radically different from that of English. My aim is to help you understand the way syntax works in languages, and to introduce the most important syntactic concepts and technical terms which you’ll need in order to see how syntax works. We’ll encounter many grammatical terms, including ‘noun’, ‘verb’, ‘preposition’, ‘relative clause’, ‘subject’, ‘nominative’, ‘agreement’ and ‘passive’. I don’t expect you to know the meanings of any of these in advance. Often, terms are not formally defined when they are used for the first time, but they are illustrated so you can understand the concept, in preparation for a fuller discussion later on. More complex terms and concepts (such as ‘case’ and ‘agreement’) are discussed more than once, and a picture of their meaning is built up over several chapters.

To help you understand what the study of syntax is about, we first need to discuss some things it isn’t about. When you read that ‘syntax’ is part of ‘grammar’, you may have certain impressions that differ from the aims of this book. So first, although we will be talking about grammar, this is not a DESCRIPTIVE GRAMMAR of English or any other language. Such books are certainly available, but they usually aim to catalogue the regularities and peculiarities of one language rather than looking at

the organizing principles of language in general. Second, I won't be trying to improve your 'grammar' of English. A PRESCRIPTIVE GRAMMAR (one that prescribes how the author thinks you should speak) might aim to teach you where to use *who* and *whom*; or when to say *me and Kim* and when to say *Kim and I*; it might tell you not to say *different than* or *different to*, or tell you to avoid split infinitives such as *to boldly go*. These things aren't on our agenda, because they're essentially a matter of taste – they are social, not linguistic, matters.

In fact, as a linguist, my view is that if you're a native speaker of English, no matter what your dialect, then you already know English grammar perfectly. And if you're a native speaker of a different language, then you know the grammar of that language perfectly. By this, I don't mean that you know (consciously) a few prescriptive rules, such as those mentioned in the last paragraph, but that you know (unconsciously) the much more impressive mental grammar of your own language – as do all its native speakers. Although we've all learnt this grammar, we can think of it as knowledge that we've never been taught, and it's also knowledge that we can't take out and examine. By the age of around seven, children have a fairly complete knowledge of the grammar of their native languages, and much of what happens after that age is learning more vocabulary. We can think of this as parallel to 'learning' how to walk. Children can't be taught to walk; we all do it naturally when we're ready, and we can't say how we do it. Even if we come to understand exactly what muscle movements are required, and what brain circuitry is involved, we still don't 'know' how we walk. Learning our native language is just the same: it happens without outside intervention and the resulting knowledge is inaccessible to us.

Here, you may object that you *were* taught the grammar of your native language. Perhaps you think that your parents set about teaching you it, or that you learnt it at school. But this is a misconception. All normally developing children in every culture learn their native language or languages to perfection without any formal teaching. Nothing more is required than the simple exposure to ordinary, live, human language within a society. To test whether this is true, we just need to ask if all cultures teach their children 'grammar'. Since the answer is a resounding 'no', we can be sure that all children must be capable of constructing a mental grammar of their native languages without any formal instruction. Most linguists now believe that, in order to do this, human infants are born pre-programmed to learn language, in much the same way as we are pre-programmed to walk upright. All that's needed for language to emerge is appropriate input data – hearing language (or seeing it; sign languages are full languages too) and taking part in interactions within the home and the wider society.

So if you weren't taught the grammar of your native language, what was it you were being taught when your parents tried to get you not to say things like *I ain't done nowt wrong*, or *He's more happier than what I am*, or when your school teachers tried to stop you from using a preposition to end a sentence with? (Like the sentence I just wrote.) Again, consider learning to walk. Although children learn to do this perfectly without any parental instruction, their parents might not like the way the child slouches along, or scuffs the toes of their shoes on the ground. They may tell the child to stand up straight, or to stop wearing out their shoes. It's not that the child's

form *I aren't* must be wrong too. It's true that speakers who accept (3) don't ever say *I are not*. But the argument is flawed: standard English is just as illogical. Look how the statement in (4a) is turned into a question in (4b):

- (4) a. **I'm not** going with you.
 b. **Aren't I** going with you?

Example (4) does not conform to the usual rules of English grammar, which form questions by inverting the word order in *I can't* to give *can't I*, and *I should* to give *should I*, and so on. Given these rules, the 'logically' expected form in (4b) would be *amn't I* (and in fact this form is found in some dialects). If the standard English in (4) fails to follow the usual rules, then we can hardly criticize (3) for lack of logic. And since *aren't I* is OK, there's no logical reason for dismissing *I aren't*. The dialects that allow either *I aren't* or *amn't I* could actually be considered more logical than standard English, since they follow the general rule, while the standard dialect, in (4), has an irregularity.

It is clear, then, that socially stigmatized forms of language are potentially just as 'logical' as standard English. Speakers of non-standard dialects are, of course, following a set of mental rules, in just the same way that speakers of the most prestigious dialects are. The various dialects of a language in fact share the majority of their rules, and diverge in very few areas, but the extent of the differences tends to be exaggerated because they arouse such strong feelings. In sum, speakers of prestige dialects may feel that only their variety of English is 'grammatically correct', but these views cannot be defended on either logical or linguistic grounds.

If, by way of contrast, some speaker of English produced examples like (5), then we could justifiably claim that they were speaking ungrammatically:

- (5) *I do didn't wrong anything.
 *Do wrong didn't anything I.

Such examples completely contravene the mental rules of all dialects of English. We all agree on this, yet speakers of English haven't been taught that the sentences in (5) are bad. Our judgements must therefore be part of the shared mental grammar of English.

Most of the rules of this mental grammar are never dealt with by prescriptive or teaching grammars. So no grammar of English would ever explain that, although we can say both (6a) and (6b), we can't have questions like (7) (the gap ___ indicates an understood but 'missing' element, represented by the question word *what*):

- (6) a. They're eating eggs and chips.
 b. What are they eating ___?
 (7) *What are they eating eggs and ___?

The rules that make (7) impossible are so immutable and fundamental that they hardly seem to count as a subject for discussion: native speakers never stop to

- b. Why hydestow (i.e. *hidest thou*) the keyes ... ?
 ‘Why do you hide the keys?’

So these are two ways in which MAIN verbs in Middle English (verbs that aren't auxiliaries) behave differently than in modern English. You may also have noticed that ‘do-support’ is used in modern English for emphasis too. We had an example just now: *Middle English did allow this construction.*

Although you may not be surprised that changes like this occurred over a period of several hundred years (with do-support becoming standard by around 1700), it may be less obvious that English changed in the twentieth century, and indeed, is still changing constantly. But there are plenty of syntactic changes in progress right now. At the moment, these are restricted to certain dialects or to non-standard British English, but all the examples of change discussed later are spreading, and some may eventually become standard English. First, consider TAG QUESTIONS such as those in bold in (17):

- (17) a. It is a hot day, **isn't it?**
 b. I can come, **can't I?**
 c. We still lost in the end, **didn't we?**

These questions ‘tagged onto’ the end of a statement are formed by specific rules in standard English which match the tag to the statement. A positive statement like *It is ...* gets a negative tag, *Isn't it*. Most importantly, an auxiliary used in the statement must be used in the tag (*I can* and *can't I*) and the pronoun (such as *it*, *I*, *we*) in the statement is also in the tag. In (17c), there's no auxiliary: main verbs like *lose* can't occur in tags (**lost we*) so do-support occurs, as in other questions. But in some dialects of British English, a single tag question *init* is used in each of the contexts shown in (17). Example (18) illustrates. The tag *init* is a reduced form of *isn't it*, a form which, in standard English, is possible only if the statement contains *is*. In *init* dialects, however, this has become an invariant tag, so that as well as the grammatically standard *It's a hot day, init?*, we find:

- (18) a. I can come, **init?**
 b. We still lost in the end, **init?**

Some other varieties of English, such as Indian English, already have an invariant *isn't it* tag. And in some languages, an invariant tag is completely standard, as in French: *n'est-ce pas* (literally, ‘isn't it?’) occurs whatever the form of the statement:

- (19) a. Il va arriver demain, n'est-ce pas? (French)
 he goes arrive tomorrow TAG
 ‘He will arrive tomorrow, won't he?’
 b. Nous n' avons pas de pain, n'est-ce pas?
 we NEG have NEG of bread TAG
 ‘We haven't got any bread, have we?’

use of *they* in (22): it's used not as a plural pronoun, but rather as a gender-neutral singular pronoun. This is clear in (22a), where *any candidate* was addressed to a group of males and females; but *they* can also be used as in (22b) and (22c), where the speakers must know the actual sex of the person referred to. (I can confirm that this is the case for the response in (22c), since I was that speaker, and heard myself say this!)

Interestingly, this development seems to have occurred independently of any desire to use non-sexist language; British English has not, for example, adopted such forms as *waitperson*, often used in American English.

To summarize, I argued in Section 1.1 that all native speakers of a language share an internal grammar, although they have never been taught its rules. Evidence for this is that we largely agree about what is and what is not a possible sentence of our language, though speakers are likely to differ over their acceptance of certain non-standard or dialectal variants. What is more, languages which are unrelated share many common properties and constructions, suggesting that human beings have an innate language faculty. Finally, we saw that language changes through time, and I gave some examples of ongoing changes. I now demonstrate how to make use of examples from other languages.

1.2 USE OF LINGUISTIC EXAMPLES

1.2.1 Why not just use examples from English?

This book contains examples from a wide variety of languages, including English. At first you may find it difficult to study examples from unfamiliar languages, and perhaps you wonder why we don't just use examples from English. There are two main reasons for using foreign-language examples: to learn about the differences between languages, and to learn about the similarities between them.

First, then, languages don't all look the same, and examining just our own language and its immediate relatives doesn't show how much languages can differ. Imagine that you've met only two languages, English and German, two closely related Germanic languages from northern Europe. Example (23), from German, is a word-for-word translation of the English.

(23) der schöne Wasserfall (German)
the pretty waterfall

You might imagine that the translation of this phrase would look the same in any language: first a word for 'the', then a word for 'pretty' or 'beautiful', then a word for 'waterfall'. But this is not so. In Spanish, for instance, we'd get (24):

(24) la cascada hermosa (Spanish)
the waterfall beautiful
'the beautiful waterfall'

Here, the word order is different in one respect: the word for ‘beautiful’ follows ‘waterfall’. Otherwise, the Spanish is not too different from the English: it has just the same three words, and a word for ‘the’ in the same position. This isn’t too surprising, as Spanish is also related to English, although more distantly than German. But in certain other languages, the equivalent to ‘the’ comes at the end of the phrase, as in Indonesian *surat itu* ‘letter the’ illustrated in (10), or else there may be no word for ‘the’ at all, as in Japanese and Chinese, or in some languages there isn’t even a direct translation of the adjective *beautiful*.

The world’s languages have many interesting and important syntactic features that I’d like you to know about. English has some but not all of these features, so if we only looked at English you’d miss out on the rest. In (25), we see one example, from Spanish:

- (25) Es nuevo. (Spanish)
 is new
 ‘It’s new.’

Example (25) has no word for ‘it’; it literally means ‘is new’ – an impossible sentence in English. Spanish typically drops the subject pronoun meaning ‘it’ in such examples; for this reason, it’s known as a PRO-DROP language. Many languages have examples parallel to this, but confining the discussion to English would never reveal that. In yet other languages, such as Arabic and Indonesian, the three-word English sentence *It is new* translates as ‘It new’ (this is illustrated in Chapter 2). These simple examples show that we can’t expect sentences in other languages to be word-for-word translations of English sentences. So we study other languages to discover the range of constructions and features they contain – in order to find out about LINGUISTIC DIVERSITY.

The second reason for looking at examples from other languages is that linguists want to discover the common properties that languages share – their *homogeneity* or sameness. One of the most important discoveries of modern linguistics is that languages don’t vary from each other at random, but are remarkably alike. Certain features occur in all languages. For instance, every language distinguishes a word class of NOUNS (words like *tree*, *liquid*, *expression* and *student*) from a word class of VERBS (words like *liquefy*, *learn*, *enjoy* and *grow*), although some languages have no other major word classes. (Chapter 2 examines word classes.) To discover this kind of information, linguists need to examine a representative sample of languages from different language families and different geographical areas.

Most linguists want to uncover the central patterns common to all languages. Although specific constructions are not universal (= common to all languages), all languages use a sub-set of the same basic tools of grammar. Each language has a wordlist or LEXICON which all its speakers share, and that wordlist always contains words from several different classes. All languages combine these words into phrases and sentences, and can manipulate the order of the phrases for various purposes – perhaps to ask questions, or to emphasize different parts of a sentence, or to show who’s doing what to whom. This is syntax, and it forms the subject matter of the chapters ahead.

1.2.2 How to read linguistic examples

1.2.2.1 *The layout of examples*

Your first task as a syntactician is to learn to make use of examples from other languages. This book contains examples from over 100 different languages. Of course, I don't speak most of these – the examples come from other linguists, or from native speakers of the language (and sometimes native-speaker linguists). But I can utilize these examples because linguists set them out in a specific way for students and researchers who don't speak the language.

Examples of this special layout occur in the two Spanish illustrations in (24) and (25). Each consists of three lines. The first line is from the source language under consideration. The third line is a translation from the source language into English. You need this line to know what the original example means, but it's not the most important part of the example, because it only tells you about English – it tells you nothing about the source language. The really important line is the second one, called the GLOSS. The gloss is a literal translation of the original language. Each meaningful part of the original is translated, whether it corresponds exactly to a word in English or not. Look back at (2): French *ne* is GLOSED (translated) simply as NEGATIVE because there's no English word that directly corresponds to it.

To see why the gloss is so important, consider (26) and (27), from Japanese and from Welsh. I have left out the gloss line. Both examples mean the same thing in the sense that they can receive the same English translation:

(26) Sensei-ga gakusei ni tegami-o kaita. (Japanese)
'The teacher wrote a letter to the student.'

(27) Ysgrifennodd yr athro lythyr at y myfyriwr. (Welsh)
'The teacher wrote a letter to the student.'

Let's suppose the point I'm trying to make is that sentences in Japanese, Welsh and English all have different word orders. Unless you happen to know both Japanese and Welsh, you won't be able to work this out from (26) and (27). In (28) and (29), I give the full examples, with glosses:²

(28) Sensei-ga gakusei ni tegami-o kaita. (Japanese)
teacher student to letter wrote
'The teacher wrote a letter to the student.'

(29) Ysgrifennodd yr athro lythyr at y myfyriwr. (Welsh)
wrote the teacher letter to the student
'The teacher wrote a letter to the student.'

2. To simplify matters, I leave two small words in the Japanese un glossed: *ga* indicates that *sensei* '(the) teacher' is the subject (here, the one writing) and *o* indicates that *tegami* '(the) letter' is the object (here, the thing being written). These terms come up again later, and in Chapters 2 and 6, so don't worry if they are unfamiliar to you now.

glosses contain only lexical information. The Rapa Nui example in (30), though, has two items glossed as *NONPAST* and *PROGRESSIVE* (which indicates an ongoing action). This information concerns grammatical categories such as *TENSE* and *ASPECT* (more on these in Chapter 2). The point is that there are no separate words in English – members of the English *LEXICON* or vocabulary – that can translate this grammatical information, so it is glossed using the technical terms that describe its function in the source language.

All languages contain grammatical information. In (31), we show this by suggesting a precise gloss of an example from English, treating it as if it were a foreign language, and representing the grammatical information, as usual, in small capitals.

- (31) The student-s ask-ed for these book-s.
DEF.ART student-PL ask-PAST for DEM.PL book-PL
 ‘The students asked for these books.’

Taking these glosses as illustration, we can now explain the usual linguistic conventions. There are some familiar lexical items, ‘student’, ‘ask’, ‘for’ and ‘book’, but I’ve glossed *the* by referring to the grammatical information it represents: it’s a *DEFINITE ARTICLE* – a word meaning ‘the’, as opposed to an *INDEFINITE ARTICLE* – a word meaning ‘a’. I also glossed *these* as *DEM.PL.*: *these* is a *DEMONSTRATIVE* word, a ‘pointing’ word from the set *this, that, these, those*. It’s also *PLURAL*, therefore used before a plural word like *books*. Throughout the book, though, I will normally try where possible to use glosses you can recognize as words.

Apart from the lexical and grammatical information, the gloss also contains pieces of information separated by a hyphen (-). A hyphen preceding or following a piece of grammatical information in the gloss means that the grammatical element is attached to the word, or to another grammatical element, and can’t be a separate word. Crucially, though, such grammatical elements have their own meaning. So the glosses *book-PL* and *student-PL* indicate that *books* and *students* are plural nouns; *-s* is a plural ending. And *-ed* is a past tense ending. I’ve also used the hyphen in the first line in (31) to indicate the boundaries in the source language between the grammatical information and the lexical items, although not all examples in this book follow this convention.

Grammatical elements attached to the beginning or end of a word, or to other pieces of grammatical information, are called *AFFIXES* (meaning something attached). Generally, then, a hyphen in the gloss indicates an affix, such as the plural *-s*. Grammatical affixes come in two main varieties: suffixes and prefixes. English plural *-s*, progressive *-ing* and past tense *-ed* are *SUFFIXES*; they’re attached to the end of words. *PREFIXES* are attached at the beginning of words; examples from English are *un-* as in *untidy* and *re-* as in *re-seal*.

Elements of meaning such as ‘ask’ and ‘past tense’, ‘un-’ and ‘plural’ are known as *MORPHEMES*. As you can see, some of these represent independent words, but not all. The study of word forms is known as morphology, and though this is generally outside the scope of this book, we will often meet examples that show the interface between morphology and syntax – morphosyntax. Glosses in an example essentially

inform the reader about the morphosyntax of the words used, as well as just giving their literal meaning.

Sometimes, we recognize that a word contains more than one piece of information, that is, more than one morpheme, but these meaning elements have no discernible boundaries. For instance, if (31) had been *The students took these books home*, we would recognize that the verb *took* is past tense, just as *asked* is, but *took* is irregular, and doesn't have a past tense *-ed* suffix. We can't tell what part of *took* means 'past'. Linguists generally indicate this in the gloss using a colon (:) or a dot: thus, *took* would be glossed 'take:PAST' or 'take.PAST'. This convention means that a single source word contains more than one morpheme (such as 'take' and 'past tense') but there are no clear boundaries between these morphemes.

We also use this convention if we just don't wish to show the boundaries in a particular example, usually for the sake of keeping things clear or simple for the reader. Illustrating again with *asked*, I could show it as in (31) as *ask-ed*, with the hyphen indicating a morpheme boundary in the source word, and gloss it as *ask-PAST*, again showing the morpheme boundary. Alternatively, I could show *asked* in the source line, and *ask.PAST* in the gloss. Typically, we use this convention when we don't need to emphasize the detailed morphosyntax of the example.

1.2.2.3 The categories of person and number

In this section, I discuss the conventions used to represent the grammatical categories of PERSON and NUMBER, using examples from French and Kwamera (spoken in Vanuatu in the Pacific).

If you have learnt a foreign language, you will probably be used to meeting tables of verb forms like Table 1.1, from French.

Table 1.1
Present tense of French verb *parler*, 'to speak'

	Singular	Plural
1st	je parle	nous parlons
2nd	tu parles	vous parlez
3rd	il/elle parle	ils/elles parlent

Such tables, known as PARADIGMS, display the set of related forms that a particular lexical word has in a given grammatical context. The paradigm in Table 1.1 shows the set of forms that makes up the present tense of the verb *parler*, 'to speak'. Reading down the column headed Singular, the forms mean 'I speak, you (singular) speak, he/she speaks'. In the column headed Plural, the forms mean 'we speak, you (plural) speak, they (masculine/feminine) speak'.

The labels 1st (FIRST), 2nd (SECOND) and 3rd (THIRD) in the first column designate the grammatical category called PERSON. First person indicates the speaker, or a group of people that includes the speaker: so both the 'I' and 'we' forms are first person. Second person indicates the addressee(s): the 'you' forms. Third person indicates

some third party, an individual or group other than the speaker and addressee: the ‘he/she/it’ and ‘they’ forms.

The category of NUMBER refers to the distinction between SINGULAR (one person) and NON-SINGULAR (more than one person). In French, as in most European languages, number is either ‘singular’ or ‘plural’. Note, though, that French distinguishes between *tu parles*, ‘you (singular) speak’, and *vous parlez*, ‘you (plural) speak’. English once had this distinction too: *thou* meant ‘you (singular)’, equivalent to *tu*; and some varieties of modern English also have second person plural forms such as *you all* or *yous* (for instance, *yous* occurs in parts of both northeast and northwest England). Some languages divide non-singular into several categories, such as a category referring to two people (a DUAL), a category for three people (a TRIAL), and additionally a plural, used for referring to more than three people. For example, the Polynesian language Kwamera has just such a system.

Kwamera also has more PERSON distinctions than are familiar in European languages. First person in this language divides into INCLUSIVE and EXCLUSIVE forms. ‘Inclusive’ means ‘we (as in me and you, speaker and addressee)’, and ‘exclusive’ means ‘we (speaker and other party, excluding you, the addressee)’. Imagine that a friend says ‘We could go and see a film tonight’. You reply ‘We? Do you mean you and me (*we inclusive*) or you and your boyfriend (*we exclusive*)?’ English doesn’t have different forms of ‘we’ to specify this information, but Kwamera does:

- (32) a. **sa-ha-akw** (Kwamera)
 IINC-PLURAL-break.up
 ‘We all break up.’ (inclusive ‘we’)
- b. **ia-ha-vehe**
 IEXC-PLURAL-come
 ‘We came.’ (exclusive ‘we’)

Before going any further, it’s vitally important that you understand how to read the information in examples like this. The English translations in (32) contain several words – four separate words in (32a), for instance. But the Kwamera source examples each contain just *one* word, though this incorporates several distinct pieces of lexical and grammatical information. I’ll explain using (32a), where there is a verb STEM, *akw*, and two prefixes attached to it – prefixes are grammatical elements which precede a stem. The *ha-* form closest to the verb stem means ‘plural’, and the *sa-* form on the outside means ‘first person inclusive’. Together, these buy the meaning ‘inclusive we’. English and Kwamera differ in a crucial way here. English has a separate pronoun *we* – it’s a distinct, independent word on its own, not part of the verb. This is known as a FREE PRONOUN. But in the Kwamera, there is no separate word for ‘we’ at all: instead, that meaning is expressed by using grammatical elements attached to the verb itself. The forms *sa-* and *ia-* can’t be separated from the verb, and don’t occur on their own, and so are known as BOUND PRONOMINALS (there is more discussion of this in Section 4.3).

Note that in Kwamera, there are separate affixes representing the categories of PERSON and NUMBER, whereas in English the pronoun *we* represents both person

(1st) and number (plural) simultaneously. So the pronominal prefix *ia-* in (32b) represents not ‘I’ or ‘we’, but just first person: it becomes ‘we’ only when the plural prefix *ha-* follows. This means that the same ‘first person exclusive’ form *ia-* also translates as ‘I’:

- (33) *ia-pkagkiari-mha*
 IEXC-talk-NEG
 ‘I didn’t talk.’

In future examples I gloss person and number as in Table 1.2, unless the language has some special inclusive and exclusive forms as in Kwamera. The first and second columns give the glosses and their meaning – this is grammatical information – and the third column lists the pronouns which in English are associated with this grammatical information, to help you remember.

Table 1.2
 Glosses for person and number

Gloss	Meaning	English pronouns
1SG	first person singular	‘I/me’
2SG	second person singular	‘you [singular]’
3SG	third person singular	‘he/him; she/her; it’
1PL	first person plural	‘we/us’
2PL	second person plural	‘you [plural]’
3PL	third person plural	‘they/them’

If the gloss specifies just the person, ‘1’, ‘2’ or ‘3’, but doesn’t mention singular or plural, this means that the particular language being glossed does not have number distinctions in this instance.

1.2.2.4 Writing systems and glosses

Not all languages use the Roman alphabet (the one you’re reading now). For example, Russian uses the Cyrillic alphabet, and Chinese and Japanese both use writing systems based on characters rather than an alphabet. But there usually exist conventions for writing such languages in the Roman alphabet, and this enables linguists to make use of the examples. I mostly follow the published source that my data come from, although some labels for glosses are changed to bring them into line with my practice. Additionally, I standardize glosses that are more detailed or less detailed than we need. Occasionally I simplify by not glossing some item, especially if we haven’t yet met the appropriate grammatical term. I will often omit the tones in examples from languages such as Chinese: these specify pronunciation and are not vital to our discussion of syntax. Finally, note that some languages don’t have a writing system at all, since they’ve never been written down. In this case, linguists typically give a phonetic representation of the original language. For that reason,

some of the examples don't start with capital letters; the phonetic alphabet doesn't follow the conventions of a writing system.

1.3 WHY DO LANGUAGES HAVE SYNTAX?

Speakers manipulate sentences in all sorts of ways because they're trying to convey different meanings. Syntax allows speakers to express all the meanings that they need to put across. In the simplest cases, this might mean altering the basic word order of a sentence, to emphasize or downplay a particular phrase, or to ask a question, or else grouping words together in different ways to modify the meaning. This section gives a preliminary idea of some of the typical syntactic constructions found in languages, and demonstrates that languages really do have syntactic structure.

1.3.1 Word order

In English, the WORD ORDER is pretty fixed. There are three main elements in the sentence in (34): *Kim*, the one drinking the tea; *drank*, the verb, which expresses what Kim did; and *the tea*, expressing what is being drunk. We use the term 'word order' (more accurately, as we will see later on, 'constituent order') to discuss the order in which these three main parts of a sentence occur in a language. In English, the three elements occur in the order shown in (34a). This is the normal word order, and all variants of it are impossible (therefore starred) except for (34f), which has a restricted special usage.

- (34) a. Kim drank the tea.
b. *Kim the tea drank.
c. *Drank Kim the tea.
d. *Drank the tea Kim.
e. *The tea drank Kim.
f. The tea Kim drank.

Most of the logically possible variations are impossible in English. However, each of the word orders in (34) is attested (= found) among the world's languages, though some are much more common than others (see Chapter 6). The three most common basic word orders in languages other than English are those of (34a), (b) and (c). We saw in Section 1.2.2.1 that Japanese has the basic word order of (34b), and Welsh the basic order of (34c). Malagasy, spoken in Madagascar, has the basic order in (34d). The two word orders in (34e) and (34f) are the rarest basic word orders in the languages of the world, although they are found in the Carib language family of the Amazon basin. For example, Hixkaryana has the word order in (34e).

It is generally possible to determine the basic, neutral word order in a language, but the flexibility or rigidity of the basic word order differs widely among the world's languages. English has a fixed basic word order, while Russian has a very flexible word order, and Japanese allows many different orders provided the verb comes at the end of the sentence, as in (28). In English, some of the starred (ungrammatical)

verb (*counted out, considered*) and they express what is being counted out, what is considered. In (41) and (42), the phrases *their catch* and *pointed gnashers* appear in a new, promoted position in the sentence. They have changed their grammatical function, and become the subjects of the passive sentences. How do we know that these phrases are now subjects? One major indication is that *their catch* and *pointed gnashers* appear immediately before the verb, in the normal sentence-initial position of English subjects. (We will see more tests for subjecthood in English in Chapter 2.) This advancement to subject position in (41) and (42) makes the promoted phrases more salient: it focuses attention on *their catch* and *pointed gnashers*. By contrast, the phrases that were the subjects of the active sentences in (43) and (44), namely *the market boss* and *the girls here*, are no longer subjects. In the passive sentences in (41) and (42), they have been demoted to a lower position. Demotion in this case means that they are consigned to a *by*-phrase, outside the core of the sentence. Notice that this *by*-phrase is entirely optional: we could omit it, and just have, for instance, *Their catch was counted out*. Compare that optionality with what we find in (43): both the subject *the market boss* and the object *their catch* are core elements of the sentence, and neither can be omitted. (Try this.)

You should now be starting to have some feeling for the purpose and usual positions of different parts of the sentence. Before leaving the topic of the passive construction, note that in English (and in many other languages) it is signalled by changes in the form of the verb: compare (45a) and (45b), where the verbs are in bold type.

- (45) a. Kim **broke** the vase. (active)
 b. The vase **was broken** by Kim. (passive)

The examples in (46) show the corresponding properties in Japanese. Example (46a) is the active sentence, (46b) its passive counterpart:

- (46) a. Sensei ga John o sikat-ta. (active) (Japanese)
 teacher SUBJECT John OBJECT scold-PAST
 ‘The teacher scolded John.’
 b. John ga sensei ni sikar-are-ta. (passive)
 John SUBJECT teacher by scold-PASSIVE-PAST
 ‘John was scolded by the teacher.’

In (46a),³ the ‘teacher’ phrase *sensei ga* is the subject, and *John* is marked as the object of the ‘scold’ verb by the *o* marker. In (46b), *John* is promoted to subject position and the ‘teacher’ phrase is demoted. It appears in the equivalent of a *by*-phrase, *sensei ni* ‘teacher by’ – note that Japanese has a different word order from English here. The

3. The verb stem (the form before the affix is added) is *sikar*, but this changes to *sikat* before the past tense suffix *-ta*.

verb also has a special passive suffix, *-are*. Please make sure you understand the way these examples are structured before moving on.

Passives and other promotion and demotion constructions are discussed in detail in Chapter 7.

1.3.3 All languages have structure

All languages, whether living or dead, have syntactic structure, including, of course, sign languages (such as British Sign Language). This means that a language doesn't just consist of strings of words, but that the words group together to form phrases, and the phrases group together to form larger phrases and sentences. Linguists describe this phrases-within-phrases pattern as HIERARCHICAL STRUCTURE.

One kind of hierarchical structure is seen in EMBEDDED sentences, a construction in which sentences occur within other sentences, such as *Chris told Lee [Kim couldn't swim]*. This property is known as RECURSION. Here, the sentence in brackets – *Kim couldn't swim* – is the embedded sentence. It serves to tell you what it was that Chris told Lee. More examples from English are given in (47): the embedded sentences are again in square brackets.

- (47) I wonder [if Lee will arrive late].
 The claim [that she doesn't like Kim] is very surprising.
 [That we've no coffee left] isn't my fault.
 We asked [how to get to the station].

For the first three phrases in brackets in (47), you can check that they really are sentences by removing the words *if* and *that* which introduce them: you then get perfectly good independent sentences such as *Lee will arrive late*. But this doesn't work for all embedded sentences, as is clear from *how to get to the station*; we will see much more on this in Chapter 3. Try to decide what properties this final example has that set it apart from the other embedded sentences in (47).

There are no limits to the number of embedded sentences that can be strung together. So given a sentence like *Kim couldn't swim* we can turn it into *Lee thought that Kim couldn't swim*, then *I said that Lee thought that Kim couldn't swim*, and so on. This means it's never possible to construct a 'longest sentence'.

I end this chapter with two short practical demonstrations that syntactic structure really exists, in other words that speakers of a language share the same mental representations of this structure. First, look at the examples in (48):

- (48) a. I charged up the battery.
 b. I charged up the street.

At first glance these sentences appear to be structurally identical. Of course, you might be aware when you read them that *charge* means something different in (48a) and (48b), but otherwise, the only difference seems to be that *street* replaces *battery*. And yet the syntactic behaviour of the two sentences is entirely different. As always, the asterisks indicate ungrammatical sentences:

- (49) a. I charged the battery up.
b. *I charged the street up.
- (50) a. *It was up the battery I charged (not the engine).
b. It was up the street I charged (not the corridor).
- (51) a. *I charged up Lee's battery and (then) up Kim's too.
b. I charged up Lee's street and (then) up Kim's too.

Since native speakers of English agree about where the asterisks showing an ungrammatical sentence should be placed, we must all share an unconscious knowledge of the sentence structure of English. Even though pairs of sentences like those in (48) look the same, they in fact have different structures. Different sets of words group together to form phrases in each case, and linguists represent this using brackets:

- (52) a. I [charged up] the battery.
b. I charged [up the street].

In (52a), the brackets show that there's a phrase *charge up*. This makes sense if you think that the only thing you can do with a battery is charge it up; you can't charge it down, over, across, or anything else. So *up* belongs with *charge* in (52a). In (52b), it doesn't: instead, there's a syntactic unit *up the street*, which can be moved around the sentence for focus, as for example in (50b), *It was **up the street** I charged*. And in (52b), *up* can be replaced with a number of other words: *I charged **down the street/over the street/across the street** and so on*. The very fact that *up* forms a unit with *charge* in (52a) but with *the street* in (52b) is responsible for the patterns of (un)grammaticality in (49) to (51). We'll return in detail to questions of structure and the grouping of words to form phrases in Chapter 5.

As a second demonstration of syntactic structure, let's examine possessive -'s in English, as in *Lee's friend*. You might assume at first that this possessive ending simply attaches to a noun, a word such as *Lee* or *government*, as in *the government's dilemma*. But consider (53):

- (53) a. I saw the woman next door's children.
b. What was that guy who retired last month's name?
c. The student I lent the book to's room-mate said she'd left.

Each example in (53) shows that -'s actually attaches to the end of a whole phrase, not to the single noun at the end of the phrase: we know that the door doesn't have children, and that the answer to (53b) couldn't be *November*. And *to* in (53c) isn't even a noun (you may find this example a little odd, because it belongs in spoken rather than written English. Try saying it aloud a few times.) Native speakers also know that you can't attach the -'s to the noun it logically seems to belong to: **What was that **guy's** who retired last month name?* The fact that we agree where to attach the

-ʒ shows once again that sentences do have structure, and that we have an intuitive knowledge of it. The phrases that -ʒ attaches to are shown in brackets:

- (54) a. [the woman next door]ʒ
 b. [that guy who retired last month]ʒ
 c. [the student I lent the book to]ʒ

Demonstrations of this nature could be given from any language, because the rules of the syntax of all languages are STRUCTURE DEPENDENT. This is why no language has rules that, for example, form questions from statements by reversing the order of the words in the sentence – such a rule wouldn't depend on the structure of the sentence at all, and so can't work. When, as linguists, we try to figure out the syntactic structures of a language, we rely on the judgements of native speakers to tell us whether our example sentences are possible or impossible (the latter being starred). These GRAMMATICALITY JUDGEMENTS, along with examples that are collected from a spoken or written corpus of the language, form the data of the science of linguistics. It doesn't matter that native speakers usually can't tell us why they feel that a particular sentence is good or bad; the very fact that they have these intuitions shows up the structural differences and similarities between sentences.

FURTHER READING

An excellent general introduction to linguistics and to the views of linguists on language acquisition is Fromkin, Rodman and Hyams (2007). I can also strongly recommend Jackendoff (1993) and Pinker (1994). Baker (2001) is a recent introduction to the view of language learning most associated with the work of the linguist Noam Chomsky. Chomsky's idea that speakers possess a subconscious 'knowledge' of their native languages is explored accessibly in the early chapters of his (1986) book, *Knowledge of language*. On language change, see McMahon (1994) and Millar (2007). On person, see Siewierska (2004); on gender, see Corbett (1991); on number, see Corbett (2000); and on agreement, see Corbett (2006). Also on person, number and related issues see Whaley (1997: Chapter 10) and Comrie (1989: Chapter 9). The topics raised in Section 1.3 all appear again in later chapters: word order is in Chapter 6, promotion and demotion processes in Chapter 7, and syntactic structure in Chapters 4 and 5.

EXERCISES

1. In Chapter 1, I argued that dialectal forms of English cannot be criticized for lack of 'logic'. The table that follows lists both the standard English forms of the REFLEXIVE PRONOUNS (the 'self' forms) and the forms found in a northern dialect of British English. Which dialect has the more regular pattern for the formation of its reflexive pronouns? Why? Be as specific as you can about how the reflexives are formed in each case, using the correct technical terms if you can.

Standard dialect	Northern dialect
myself	myself
yourself	yourself
himself	hisself
herself	herself
ourselves	theirself
yourselves	ourselves
themselves	yourselves
	theirselves

I've omitted *itself*, which is the same across all dialects, and doesn't shed any light on the question. Also, the form *myself* is generally pronounced more like *meself* in this northern dialect, but I take this to be a phonetic reduction which is not relevant to the question. The northern dialect has one more reflexive pronoun than standard English. What is it, and what do you think it's used for? (If you're not a native speaker of this dialect, you may find it helpful to look back at Section 1.1.) Some English speakers have a singular form *themself*: comment on how this fits into the set of forms in standard English, and say how it is used (if you're not a speaker who has this form, see if you can imagine how it's used).

Finally, do you think that 'more logical' equals 'better', as far as languages are concerned?

2. In Chapter 1, we discussed syntactic changes in progress in English, such as the use of *init* in tag questions. Another change which seems to be gaining ground in British English concerns the COMPARISON of adjectives (words like *red*, *happy*, *difficult* in English).

Task: (i) Consider first what words can fit into the gap in (1); compile for yourself a list which you think is representative:

- (1) This is ___ than that one.

Standard English uses either a single word from a list like (a) or a phrase from a list like (b):

- | | |
|-------------|--------------------|
| (a) funnier | (b) more difficult |
| older | more dramatic |
| dearer | more generous |
| cleverer | more unpleasant |
| pleasanter | more expensive |

So the COMPARATIVE form of adjectives in standard English can consist of either ADJECTIVE + *-er* suffix, as in list (a), or *more* plus ADJECTIVE as in list (b).

(ii) Work out the standard English rule that governs the choice of these alternatives. Your answer should account for the fact that standard English does not have forms like those in (2):

(2) *more difficulter, *more unpleasanter, *more generouser

The forms in (2) are ungrammatical in all dialects, as far as I know, but many speakers of non-standard English can fill the gap in (1) with the forms in (3):

(3) more funnier, more dearer, more cleverer

(iii) Describe as accurately as you can the syntactic change in progress here, using the correct grammatical terms if you can.

3. The data in this exercise are from Icelandic, a Germanic language which is related quite closely to English, and are taken from Sigurðsson (2006).

Task: All these examples illustrate a single, specific, grammatical difference between English and Icelandic. What is this? You should use the correct grammatical term, which is given earlier on in this chapter. If you find other syntactic differences between the two languages in any example, list these too. Finally, in what specific ways can you see that Icelandic is syntactically similar to English? Use the correct terminology wherever possible. **NB!** The Icelandic alphabet includes some characters that are not used in English, but this has no bearing on the answer.

(1) Kona sat á bekk.
woman sat on bench
'A woman sat on a bench.'

(2) Ég keypti skemmtilega bók í morgun.
I bought interesting book in morning
'I bought an interesting book this morning.'

(3) Ólafur er prófessor.
Ólafur is professor
'Ólafur is a professor.'

(4) Það er maður í garðinum.
there is man in garden.the
'There is a man in the garden.'

(5) Sá sem er að tala er Íslendingur.
the.one who is to talk is Icelander
'The one who is talking is an Icelander.'

4. In (1) through (8), which follow, are some more examples from Kwamera, and two closely related languages, Lenakel and Southwest Tanna, all languages of the

Republic of Vanuatu in the southwest Pacific; data are from Lindstrom and Lynch (1994) and Lynch (1998). I give the original and the gloss, and your task is to suggest a translation. You will find it useful to look back at Section 1.2.2.3, where the Kwamera examples are discussed in the text.

Hint

There is rarely just one correct way of translating each example; the important part here is to make sure you understand the role of the grammatical information in the glosses (in small capitals).

- (1) t-r-uv-irapw (Kwamera)
FUTURE-3SG-move-outwards
- (2) t-r-am-apri (Kwamera)
FUTURE-3SG-CONTINUOUS-sleep

Translate the ‘continuous’ prefix *am-* using an *-ing* form of the verb in English, as in *I was sleeping, I’ve been sleeping, I will be sleeping*. Both Kwamera *am-* and English *-ing* denote an ongoing action here.

- (3) iak-imiki kuri u (Kwamera)
1EXC-dislike dog this

The prefix *iak-* is the form of the first person exclusive which occurs before vowels.

- (4) k-rou-ánumwi (Kwamera)
1INC-DUAL-drink

Note that the first person inclusive prefix, *k-*, does not have the same form as the first person inclusive *sa-* that we saw in (32a) in the text. The reason for this is that *sa-* is used in conjunction with a plural prefix, while *k-* co-occurs with a dual prefix.

- (5) K-ím-hal-vín-uas. (Lenakel)
3PL-PAST-TRIAL-go.there.together
- (6) R-ím-avhi-in mun (Lenakel)
3SG-PAST-read-TRANS again

The suffix *-in* marks the verb as TRANSITIVE (more in Chapter 2): this means that it is understood to have an object (see Section 1.3.1), and this should affect your translation.

- (7) Kímlu i-ímn-la-gín (Southwest Tanna)
we.two.EXC 1PL-PAST-DUAL-afraid

- Tinrin has a grammatical category ‘dual’, like Kwamera, discussed in this chapter. But this is not relevant to your answer.
- It will help to compare the structure of (1) through (6), which have the *nrâ* marker in question, with that of (7) through (9), which do not. What syntactic factor distinguishes these two groups of data?

(1) rru fi pwere ânramwâ **nrâ** truu truu-truare
 3.DUAL go to there ??? the.DUAL DUAL-brother
 ‘The two brothers went there.’

(2) nrâ nyôrrô **nrâ** wa mwîê mwâ
 3SG cook ??? the woman that
 ‘That woman cooked (something).’

(3) u truumwêrrê mirri **nrâ** nro
 1SG always hungry ??? I
 ‘I am always hungry.’

(4) nrâ tewùrrù nranri **nrâ** toni
 3SG tie.up goat ??? Tony
 ‘Tony tied up the goat.’

(5) nrâ truu tôbwerrî-nrî **nrâ** magasâ
 3SG stay close-PASSIVE ??? shop
 ‘The shop is closed.’

(6) nrâ tôbwerrî-nrî **nrâ** magasâ rugi midi
 3SG close-PASSIVE ??? shop at noon
 ‘The shop is/was closed at noon.’

(7) wa mwâ ha hêê-rò
 the house this belonging-me
 ‘This house is mine.’

(8) rri truu hubo ei nrü
 3PL stay near to 2SG
 ‘They live near you.’

(9) toni nrâ tuo nrî padrêrrê-tave
 Tony 3SG put 3SG side-bed
 ‘Tony put it beside the bed.’

Words belong to different classes

Section 2.1 should be useful to readers who have little previous experience of word classes, or ‘parts of speech’. This section concentrates on English. Then in Sections 2.2 to 2.5, we look at the major LEXICAL word classes occurring cross-linguistically, namely verbs (2.2), nouns (2.3), adjectives (2.4) and adverbs (2.5). Although all languages distinguish a class of verbs from a class of nouns, it is less clear whether or not all languages have a separate adjective word class, as we will see. Adverbs are widespread, but not universal. Section 2.6 discusses adpositions, also a widespread word class cross-linguistically. Each section discusses the distribution, function and MORPHOSYNTACTIC properties of the word class it describes. All the major word classes are associated with a typical set of grammatical categories. We concentrate here on the most common categories found cross-linguistically.

2.1 IDENTIFYING WORD CLASSES

2.1.1 How can we tell that words belong to different classes?

It is easy to demonstrate that words in a language fall into different classes. For example, only certain single words can fill the gap in (1) to complete the sentence:

- (1) Kim wanted to ____ .

The gap can be filled as in (2), but not as in (3):

- (2) Kim wanted to leave/browse/relax/sleep.
- (3) a. *Kim wanted to departure/browser/relaxation.
b. *Kim wanted to underneath/overhead.
c. *Kim wanted to energetic/thoughtful/green/sad.

The words that can fill the gap are all VERBS. Verbs appear in a variety of other positions too, but if we have to find one word to complete (1), it must be a verb. So the words that are impossible in (3) are not verbs: they must belong to other word classes. Note that to try this test you don’t need a definition of ‘verb’, because you’re simply applying your knowledge of English: you know without being told that only certain words fit in (1). From now on, you can use this test as follows: any single word which can fill the gap in (1) must be a verb.

Second, *engine* can never take the typical adjective endings *-er*, *-est*, as in *untidier*, *happiest* (nor can we say **more engine*, **most engine*). So *engine* never has the same set of word forms as an adjective either. But it does take the plural *-s* suffix of nouns, as in *Kim drives engines*.

Another way to use distributional evidence is to show that nouns and adjectives are MODIFIED by different word classes: they keep different company. So, like other nouns, *engine* can itself be modified by an adjective, such as *electric*. But it can't be modified by an ADVERB such as *electrically* (the meaning intended in (8) is that the engine is electric, not Kim):

- (8) Kim is an **electric engine** driver.
*Kim is an **electrically engine** driver.

This is typical behaviour for a noun. But adjectives behave in a different way: they are not modified by other adjectives – such as *unbelievable* in (9) – but by adverbs, such as *unbelievably*. So the asterisks are the opposite way round in (8) and (9).

- (9) *Kim is an **unbelievable skilful** driver.
Kim is an **unbelievably skilful** driver.

This distributional test distinguishes adjectives like *skilful* from nouns like *engine*. To account for all the examples seen here, we simply need to say that nouns such as *driver* can be modified either by adjectives (*skilful driver*), or by other nouns (*engine driver*).

Now consider verbs such as *vegetate* and *survive*; these don't seem to be events, actions, processes or states (or 'doing' words!), but the formal distribution test in (1) shows that they are indeed verbs (e.g. *Kim wanted to vegetate*). As before, you may have to adjust the test slightly in order to fit the meaning of the verb. Again, these verbs take the past tense *-ed* suffix (*vegetated*, *survived*). They also take two other endings that are found on verbs: *-s* and *-ing*. Only verbs in English take *all three* of these suffixes, *-ed*, *-s*, *-ing*. But unfortunately, it's hard to use these suffixes independently to identify verbs, since they each have other grammatical roles. For instance, *boring* has the *-ing* suffix and can be a verb: *Kim's boring me to death*. But it's an adjective in *Kim's very boring*, as we can tell by *very*, which only modifies adjectives: we don't get: **Kim's very boring me to death*.

The formal methods that linguists use to identify word classes concentrate both on MORPHOLOGICAL criteria and on SYNTACTIC criteria. Morphology is the study of word form. Recurring patterns in the form of words, particularly in the affixes that they take, indicate that a group of words belong to the same class. We've seen several examples already: for instance, the observation above that only verbs take all three endings *-ed*, *-s* and *-ing*. This kind of evidence is based on the MORPHOSYNTAX of verbs: the morphology that they take in specific syntactic contexts. We will see plenty more morphosyntactic categories as we go along.

Syntactic criteria show that each word class has a unique pattern of distribution. First, there are certain slots in a sentence that can only be filled by members of one

that of a verb. Of course, nouns aren't simply randomly interchangeable with verbs. We can tell that the words *dog* and *eyes* in (11) really are verbs by substituting more typical verbs:

- (12) a. Revived ferry sale fears **disturb/jeopardize/irritate** islanders.
 b. Treasury **considers/postpones/denies** wider prescription charges.

How effective would the headlines be if we changed them as follows?

- (13) a. Revived ferry sale fears **have dogged** islanders.
 b. Treasury **to eye** wider prescription charges.

These don't achieve the same effect at all because it's now (too!) obvious that *dogged* and *eye* are verbs. You don't have to know the meaning of 'verb' to pick up the various clues to word class that (13) contains – as a speaker of English, you use these clues subconsciously all the time.

In (11b) the form *eyes* was particularly clever, because out of context, it might be either a noun or a verb – both word classes happen to have an *-s* suffix in English, though it performs very different work in each case. So (11b) at first leads us astray by playing on the fact that the word *eyes* can be a noun or a verb. In (13b), though, the use of *to eye* makes it clear at once that *eye* is a verb. Nouns can't fit into that slot:

- (14) *Treasury **to ear/denial/postponement** wider prescription charges.

Although evidence from morphology (word form) can often be used to distinguish word classes, it's not always available. Furthermore, some languages – such as Chinese or Vietnamese – have very few grammatical affixes. For example, nouns in Chinese are not marked for a singular/plural distinction, so for instance the word *xin* translates as both 'letter' and 'letters'. In such languages there isn't much morphological variation, so word form won't usually help to identify word class.

Syntactic evidence to distinguish word classes typically is available, however. In (13a), the verb *dogged* is followed by *islanders*; and in (13b) the verb *eye* is followed by the 'noun phrase' *wider prescription charges*. In fact, these phrases (or ones like them) have to be present, or else the sentences will be ungrammatical (check this for yourself). Here, then, is another distribution test for verbs: certain verbs must be accompanied by a NOUN PHRASE like *islanders* or *wider prescription charges*. In technical terms, as we saw in Chapter 1, this phrase is the OBJECT of the verb. Verbs that need an object (often termed a DIRECT OBJECT) are known as TRANSITIVE verbs.

For completeness, notice that there's also an adjective *dogged* – it has a different (two-syllable) pronunciation from the verb, and means something like 'determined'. The adjective occurs, for instance, in *these dogged islanders*, where it modifies the noun *islanders*. As you'll expect by now, it has a different distribution to that of a verb or a noun. For instance, using the tests in (7), we can get *Kim's as dogged as Chris* and *Kim is so dogged*, but a noun or verb won't work here: **Kim's as dog as Chris*.

The newspaper headlines in (11) make use of words from just three different word classes: *ferry*, *sale*, *fears*, *islanders*, *treasury*, *prescription* and *charges* are all nouns;

dog and *eyes* are verbs, as is *revived* in this usage, and *wider* is an adjective. The majority of words in the headlines are nouns and verbs – these word classes are indispensable and, cross-linguistically, are always the most important word classes. All languages seem to have distinct classes of nouns and verbs, so these are true LANGUAGE UNIVERSALS (= a property found in all languages). Also, nouns and verbs in most languages are OPEN CLASS words: this means that we can add new words to these classes. For example, the nouns *byte*, *blog*, *software* and *laser* are all recent innovations in English, as are the verbs *breathalyse* and *decoke* (to remove carbon deposits from an engine).

In English and other European languages, adjectives (and maybe adverbs) are also open class words, but not all languages have an open class of adjectives, that is, a class to which new adjectives can be added. For example, Igbo, a language of the Benue-Congo family spoken in Nigeria, has a CLOSED CLASS of adjectives with just eight words in it. In fact, perhaps not all languages have a class of adjectives or adverbs at all (see Sections 2.4 and 2.5).

Adding a couple of other typical headlines, we also find the word class PREPOSITION – shown in bold in (15) – but no other word classes. *Bird* is slang for a prison sentence – the headline is about a woman illegally feeding pigeons:

- (15) MPs' report urges action **within** four years **on** design changes.
Pigeon woman is cured **by** spell **of** bird.

(From *The Guardian* 29.7.95)

Prepositions aren't open class words, and some languages have very few or even no prepositions. English, however, has a large class of prepositions conveying many different meanings. From the newspaper headlines, you can see that in English the four classes N (nouns), V (verbs), A (adjectives) and P (prepositions) contain the words we need most when we're trying to write in 'telegraphese'. Cross-linguistically, we can expect the classes N, V and A to be the major LEXICAL word classes, containing most members, and expressing most of the important meanings.

Some prepositions don't really carry much meaning, and are used for purely grammatical purposes: *by* and *of* are like this in (15). Headlines can often dispense with words that mainly bear grammatical information. This is why headlines don't typically contain the grammatical 'little words' like ARTICLES (*the*, *a* in English) which don't have much semantic content; in other words, meaning. All languages have words that express grammatical information, such as definiteness (*the*) or indefiniteness (*a*), or the DEMONSTRATIVES (*this*, *that*, *these*, *those*), or negation (*not*); a language may well not have counterparts to these specific grammatical elements, but there will certainly be grammatical words of some kind. These purely grammatical words are known as FUNCTIONAL categories, and they contrast with LEXICAL categories, which are rich in meaning. Other functional categories include CONJUNCTIONS (such as *and*, *or*, *but*) and PRONOUNS (such as *she*, *her*, *they*, *them*). We will see more as we go along.

We've now seen something of the way speakers of English 'decide' (subconsciously) the classes of the words they encounter. We've also begun to see how linguists

discover the different word classes by running a set of diagnostic tests based on morphological and syntactic evidence.

To summarize, we've argued in this section that words fall into different classes. Evidence comes partly from morphosyntax: each word class has its own unique set of affixes. But morphological evidence of this kind is not always available, so syntactic evidence is vital too. Each word class fits into certain slots which are unique to it, and each class co-occurs with (keeps the company of) specific words from other classes. Furthermore, each word class has specific functions, performing certain tasks in a sentence.

We next turn to a wider examination of the major lexical word classes, looking at their typical behaviour cross-linguistically. Section 2.2 looks at verbs; Section 2.3, nouns; Section 2.4, adjectives; Section 2.5, adverbs; and Section 2.6, prepositions.

2.2 VERBS

2.2.1 An introduction to verb classes

The major function of verbs is to express what is known as 'predication'. A PREDICATE expresses an 'event' in the sentence, which may be quite literally an event (such as *collapse* or *explode*) but also includes actions, processes, situations, states and so on. Though the role of predicate is typically fulfilled by a verb, we will see later that this isn't always the case.

In all languages, verbs fall into various syntactic sub-classes. Three of the most important are discussed in this section, starting in (16) with the sub-class of intransitive verbs. The verbs are in bold:

- (16) a. Lee **sneezed**.
The volcano **erupted**.
- b. ótáù **síkáàna** (Mbalanhu)
night falls
'The night falls / is falling.'
- c. **Bhéic** sé. (Irish)
yelled he
'He yelled.'

Each of these verbs requires a single participant, the entity involved in the event or action which the verbs express. The participants in these examples are *Lee*, *the volcano*, *ótáù*, *sé*. In linguistic terminology, we say that the participant is the ARGUMENT of the verb. ('Argument' is a technical term, and doesn't mean that the verb and the participant are quarrelling!) Verbs with only one participant or argument are called INTRANSITIVE verbs. Note that it may well be the case that this single argument is an entire phrase, maybe even referring to many people: *Lee and Kim sneezed*; *All the students sneezed*. But nonetheless, the verb *sneeze* has just the one argument.

You should have the following results. In (21a), *capitulated* is intransitive – it has only one argument, *Lee and Kim*: note that you can get this answer without actually knowing the meaning of the verb. *Gave* is ditransitive, although note that in (21b) the participants appear in a different order than that in (19): the recipient (*the children*) in this example comes before what is given. *Assassinated* is a transitive verb, which is why (21c) is impossible: its direct object is missing. Both verbs in (21d) are transitive: both *love* and *detest* have two arguments. You may have thought that these are intransitive verbs, because there is no argument immediately following the verb. But this is wrong, as we can tell because **Kim loves/detest*s is ungrammatical. The direct object arguments, *sprouts* and *cabbage*, would normally be positioned immediately after the verb, but in (21d), each of them has been moved from its usual position, for emphasis. Even displaced in this way, *sprouts* and *cabbage* still fulfil the requirements of both verbs for an ‘item loved/detested’ participant. So even if an argument is displaced from its usual position, it still ‘counts’ as an argument of the verb that it’s associated with.

In English, there are very many verbs that are ‘ambitransitive’: these can be either transitive or intransitive, such as *sing*, *cook*, *read*, *eat*. Fewer verbs can only be transitive (*devour*, *reject*) or only intransitive (*erupt*, *disappear*). This situation is not necessarily the same for all languages. For instance, in Jarawara, an Amazonian language (Dixon 2004b), about half the verbs are strictly intransitive, somewhat fewer are strictly intransitive, and maybe a third of the total are ambitransitive.

More verb classes are illustrated as we go along, in Chapters 3, 4 and 5. What we have seen in this section is that across all languages, verbs occur with specific ‘core’ arguments: these are the arguments required by the verb. The verb also selects the particular grammatical properties of its arguments, as we’ve seen. This relationship between a verb and its arguments is one kind of DEPENDENCY: a relationship contracted between elements in a sentence. We will see other kinds of dependencies throughout this book.

2.2.2 Verbs and their grammatical categories

Verbs have more cross-linguistic differences in the grammatical categories they express than any other word class. The major categories are illustrated here.

2.2.2.1 Tense and aspect

These are the most common morphosyntactic categories associated with verbs, and this discussion provides only a brief sketch of these extensive categories. Starting with English, you may be surprised to learn that morphologically speaking (in terms of form) English verbs have only two tenses, namely present and past:

- (22) a. Kim **helps** Lee every day.
 b. Kim **helped** Lee every day.

The present tense of the verb in (22a) is marked by the *-s* inflection (ending), although this only occurs on the third person singular form: so in *I help(*s) Lee*, the verb has no actual suffix. This tense is sometimes referred to as ‘non-past’, a more accurate label, because most ‘present’ tense verbs don’t refer to something that is happening right now. So (22a), for example, refers to a habitual event. The PAST TENSE in (22b) is marked with the *-ed* suffix, and this doesn’t change for person and number. These *-s* and *-ed* endings are the only pieces of regular verbal morphology that represent tense in English, although *-s* actually has a dual role, as we’ll see later.

What about the future tense? English certainly has ways of referring to future time: one is to use the present tense of an AUXILIARY element *will*: *She will help Lee tomorrow*. But the main verb, *help*, doesn’t inflect here. There is no ‘future’ verbal morphology equivalent to the *-s* present tense or *-ed* past tense endings. The present tense of a verb can also refer to future time – as in *She leaves the country tomorrow* – or we can say *She is leaving the country tomorrow*, using another auxiliary, *is*. Note that the *-ing* suffix here isn’t a tense marking: it can occur with any time reference, as in *She was leaving*, *She will be leaving*.

Tense is defined by Comrie (1985a: 9) as the ‘grammaticalized expression of location in time’. The point is that different languages will ‘choose’ to grammaticalize (= represent grammatically) different contrasts in time – these are its tenses. This does not mean that a language can only refer to the points in time for which it has a morphological marker for tense, as we’ve already shown for ‘future’ in English. Other languages may have many more tense distinctions than English, or even fewer tenses, even none at all. Some Austronesian languages (e.g. Leti, Saliba) have no grammatical tense: there is no verbal morphology which represents tense in these languages, nor are there separate tense markers or auxiliaries. There are words that refer to time, however, such as Saliba *lahi* ‘yesterday’ and *malaitom* ‘tomorrow’.

Most languages have a basic two-way tense opposition: either between past and non-past tenses – like English – or else between future and non-future tenses. Within these major divides, some languages have much finer tense distinctions, particularly the African Bantu family, and native Australian and American languages. The Wishram-Wasco dialect of Chinook, a native American language spoken in the states of Oregon and Washington, has four past tenses represented by different inflections, or markings on the verb, shown in bold:

- | | | | | |
|------|----|---|----------------------------|-----------|
| (23) | a. | ga- čiu _x | ‘He did it some time ago.’ | (Chinook) |
| | b. | ni- čí _x | ‘He did it long ago.’ | |
| | c. | na- číú _x ^w -a | ‘He did it recently.’ | |
| | d. | i- čí _x | ‘He just did it.’ | |

Note that the tense inflections are prefixes in this language.

A category closely related to tense is that of ASPECT. Aspect marks such properties as whether an action is ongoing or completed. For example, in *Kim was eating his dinner*, the verb *was* is past tense but we understand that the eating event wasn’t over. This sentence has the PROGRESSIVE aspect, marked in English partly by the *-ing*

- (28) a. Kalulu a-na-b-a mkazi wa njovu. (Chichewa)
 hare SU-PAST-steal-ASPECT wife of elephant
 ‘The hare stole the elephant’s wife.’
- b. Mkazi wa njovu a-na-b-**edw**-a (ndi kalulu).
 wife of elephant SU-PAST-steal-PASSIVE-ASPECT by hare
 ‘The elephant’s wife was stolen (by the hare).’

In both Chichewa and English, the passive affects the arguments of the verb, and also the form of the verb itself. The noun phrase *mkazi wa njovu*, ‘the elephant’s wife’, is the direct object in (28a), and becomes the subject in the passive (28b): in the terminology introduced in Chapter 1, it gets promoted to subject position. The subject of the active sentence, *kalulu*, ‘the hare’, is demoted in the passive: it becomes the object of a preposition *ndi/by*, or it can be omitted entirely. The valency of the ‘steal’ verb is altered in the passive: in (28a), it takes two core arguments, a subject and a direct object, while in (28b), it has only one core argument: *mkazi wa njovu*, ‘the elephant’s wife’. The phrase *ndi kalulu* ‘by the hare’ is not a core argument: it can be omitted entirely.

The passive in English is characterized by an auxiliary *be* or *get* (as in *It got stolen*) plus the PAST PARTICIPLE form of the main verb (*stolen, seen, killed*) but there’s no specific passive affix. Chichewa, however, marks the passive directly on the verb, using the *-edw* suffix in (28b).

2.2.2.4 Agreement

Verbs in many languages ‘agree with’ one or more of their arguments (see Chapter 6). This means that various properties of the noun phrase arguments are also marked on the verb, the most common properties being person and number, and then gender or noun class. The situation most familiar to speakers of European languages is that of subject/verb agreement. English has very little verbal agreement – only the third person singular in the present tense is overtly marked (for example, *I play* versus *He plays*). This is the dual role of the *-s* suffix mentioned earlier: it represents both 3SG and present tense.

The Australian language Gunin also has subject/verb agreement, but in Gunin it is the gender of the subject that is cross-referenced (morphologically marked) on the verb, as shown in (29). Gunin has five genders, one denoting all humans (male or female) and four covering all non-human nouns (see Section 2.3.3.2 for more on gender).

- (29) a. benyjin **bi**-yangga (Gunin)
 man GENDER-goes
 ‘The man is walking.’
- b. leewa gadi **a**-yangga
 dog run GENDER-goes
 ‘The dog is running.’

Cross-linguistically, verbs often agree with their objects as well as their subjects. This example is from a Malayo-Polynesian language, Kambara:

One hypothesis might have been that subjects all bear the same semantic role. But we already know from Section 2.3.1 that this is not the case: different verbs require their subjects to bear different roles. So in (36a), the subject is an AGENT, in (36b) and (d), an EXPERIENCER, and in (36c), a PATIENT (apples are the ‘thing grown’).

Looking at the distribution of the phrases in bold, we might conclude that subjects precede the verb in English. This is certainly true, and as noted in Chapter 1 it is indeed one of the ways we can tell subjects in English. It is definitely not true of all languages, however, as we saw for Irish in (16c), where the verb precedes the subject. Having observed that ‘English subjects precede the verb’, you may wonder if *every* NP that precedes the verb in English is a subject. We particularly need to know the answer to this in (36d), where two NPs precede the verb. Only *she* is marked in bold, though. How do we know that *she* is the subject and not *apples*? There are two ways of testing this, and these tests give us two further properties of subjects in English.

First, subjects in English control SUBJECT/VERB AGREEMENT: verbs and auxiliaries change in form to match or ‘agree’ with particular features of the subject, such as person and number. So in (36a), the verb *buys*, third person singular, agrees with the singular subject, *this woman*, while in (36b) and (36c) we get plural auxiliaries *are/were* to match plural subjects – *all those people* and *apples*. If you aren’t quite satisfied that *apples* really is the subject in an example like (36c), perhaps because of its semantic role, note that the subject/verb agreement test proves that *apples* really is the subject: we get *apples were grown* rather than **Apples was grown* (at least in standard English). This confirms that subjects are defined by their syntactic properties, not by their semantic roles. It also shows that we must distinguish between the semantic role and the grammatical relation of an NP: remember that subjects are often agents, but not always. Turning to (36d), the verb *enjoys* is a third person singular form: it agrees with *she* (3SG) and not with *apples*, which is plural. So *she* is the subject of the verb *enjoys*.

The second test for subjecthood in English involves CASE-MARKING. Pronouns have a special form in English which is restricted to the subject position. This test is appropriate for the subject of a verb (or auxiliary) that is FINITE, such as *loves* (present tense) or *tasted* (past tense). We’ll explore the verbal property of finiteness further in Chapter 3, but for now you can consider it to be equivalent to ‘bearing tense’. The correct subject pronouns are in bold (examples are again from standard English):

- (37) a. **She**/*her loves apples.
 b. **We**/*us don’t grow that kind of apple.
 c. **They**/*them saw her/*she
 us/*we.
 d. Those apples tasted great to her/*she
 us/*we.

CASE means that the form of a noun phrase or a pronoun changes according to its grammatical relation (more details in Chapter 6). In the pronoun pairs *I/me*, *we/*

apples; Lee enjoys **all the varieties of apples that we grow in the orchard**.

A third grammatical relation is that of PREPOSITIONAL OBJECT, taken by the NPs in bold in *on the bus*, *by train*, *with three friends* and also by *her* in *to her*. The words *on*, *by*, *with* and *to* are prepositions (see Section 2.6). English subject pronouns have a special form, as we saw above, but the objects of both verbs and prepositions share the same form: for instance, words such as *her* and *us* can be the object of either a verb, like *saw*, or a preposition, like *to*.

Although the most typical function of NPs is as arguments of a verb, noun phrases can in fact also be predicates, expressing an event or a situation:

- (40) a. Zainal **guru** **saya**. (Malay)
 Zainal teacher my
 ‘Zainal is my teacher.’
 b. Marija **rebēnok**. (Russian)
 Mary child
 ‘Mary is a child.’

The NP predicates, in bold, are *guru saya* ‘my teacher’ and *rebēnok* ‘a child’. The English translations also have NP predicates: *my teacher*, *a child*. However, in English the predicate NP is linked to the subject by *is*, a form of the verb *be*. Such linking verbs are known as COPULA verbs. In the examples in (40), though, there is no copula, which is actually a common situation cross-linguistically. In fact, even in English we can omit the copula to express disbelief: *Zainal a teacher? Who would ever have believed it?* This knowledge may also help you with the subject slot in ‘__ extinct’ in (39), which also omits the copula.

We have seen so far that noun phrases most often function as participants or ARGUMENTS of verbs. They can be classified in terms of their semantic functions (agent, theme and so on) or in terms of their syntactic functions, known as grammatical relations – for instance, subject, direct object, and prepositional object. We return to grammatical relations in Chapter 6.

2.3.3 Nouns and their grammatical categories

2.3.3.1 Number

Many languages mark nouns and noun phrases according to whether they are singular or plural. Typical examples are shown from an Austronesian language, Saliba, which like English has plural suffixes on nouns:

- (41) a. natu-gu b. natu-gu-**wao** (Saliba)
 child-my child-my-PLURAL
 ‘my child’ ‘my children’
 c. natu-m d. natu-m-**wao**
 child-your child-your-PLURAL
 ‘your child’ ‘your children’

- (43) a. **mo**-tswadi b. **ba**-tswadi (Northern Sotho)
 CLASS 1-parent CLASS 2-parent
 ‘parent’ ‘parents’
- (44) a. **le**-oto b. **ma**-oto
 CLASS 5-foot CLASS 6-foot
 ‘foot’ ‘feet’

Here, class 1 indicates human beings, and class 2 is the plural of class 1. Class 5 (and its plural, class 6) indicates body parts, but is also used for nouns representing many other concepts, including natural phenomena, fruit and vegetables, various birds and animals, and nationalities, among other things. The meaning of a noun therefore does not correlate strictly with noun class.

2.3.3.3 Possession

Possessive constructions are often quite complex. For a start, a language may regard some types of noun as not referring to possessable things at all, including features of the natural world such as rocks or rivers. In terms of possessable nouns, it’s very common to find a division between what is known as ALIENABLE and INALIENABLE possession. Typically, nouns for body parts or for a person’s relatives are in the inalienably possessed class; these include terms for things that you can’t put aside or dispose of. Alienable possession covers other types of noun, such as possessions, animals or food. These examples are from Jarawara, a language of southern Amazonia:

- (45) a. Okomobi kaa taokana b. ami kaa jomee (Jarawara)
 Okomobi POSS gun mother POSS dog
 ‘Okomobi’s gun’ ‘mother’s dog’
- (46) a. Okomobi teme b. ami tame
 Okomobi foot.M mother foot.F
 ‘Okomobi’s foot’ ‘mother’s foot’

Alienable possession is illustrated in (45), and inalienable possession in (46). You can see that alienable possession requires the use of an extra possessive morpheme, *kaa*, whereas inalienable possession merely involves placing two nouns side-by-side. Cross-linguistically, this is expected: alienable possession typically involves additional morphology, while inalienable possession just involves the juxtaposition of the nouns.

2.3.3.4 Case

Case is a grammatical property that occurs in many languages, but by no means all, and indicates the grammatical function of an NP in a phrase or sentence. Case marks,

for example, whether a noun phrase is a subject or an object of a verb: it denotes the relationship the NP has to that verb. Not all languages have case-marking: this means they don't mark the grammatical function of an NP on that NP in any way. English has very little case morphology: we saw earlier that only pronouns have a special form when they fulfil the grammatical relation 'subject' of a finite verb. Even then, the forms *you* and *it* have no distinctive case-marking. Some languages have even less case-marking than English:

- (47) a. Saya benci dia. (Malay)
 I hate he/she
 'I hate him/her.'
- b. Dia benci saya.
 he/she hate I
 'She/he hates me.'

Note that the subject and object forms of each pronoun do not differ from each other in form, so that *saya* in Malay translates as both 'I' and 'me', and *dia* translates as *he/she* as well as *him/her* (the pronoun having no gender distinction either).

Conversely, some languages have rich case systems, such as Turkish, Finnish, Latin and the Slavonic languages (e.g. Russian and Polish). Examples from Latin are shown in (48). The 'nominative' case (NOM) indicates the grammatical relation of subject, and the 'accusative' case (ACC) indicates the grammatical relation of direct object:

- (48) a. Nauta puellam amat. (Latin)
 sailor.NOM girl.ACC loves
 'The sailor loves the girl.'
- b. Puellam nauta amat.
 girl.ACC sailor.NOM loves
 'The sailor loves the girl.'

Note how flexible the word order is in Latin: since the grammatical relation of the noun phrases is always marked on the NPs themselves, they don't have to occur in a fixed order, unlike in English. So (48a) and (48b) have the exact same meaning, no matter whether it's the subject *nauta*, 'the sailor', that's initial in the sentence, or the object *puellam*, 'the girl'.

2.3.4 Nouns, definiteness and determiners

Some languages, such as the Scandinavian languages Norwegian, Swedish and Danish, can mark definiteness morphologically – via a change in the form of the noun – as well as using a definite article, a word for 'the':

- (49) a. mus-en (Swedish)
 mouse-DEF
 'the mouse'

- b. den hungriga mus-**en**
 the hungry mouse-DEF
 ‘the hungry mouse’

The suffix *-en* marks definiteness, and can co-occur with *den* ‘the’, as in (49b).

The noun itself doesn’t have any ‘definiteness’ morphology in English. Many languages, including English, can distinguish definite from indefinite nouns by using a separate functional element – an article, such as the definite article *the* or the indefinite article *a/an*. Articles are members of a larger class of functional words known as DETERMINERS. Some of the main subclasses of English determiners are shown in (50), with the determiners themselves in bold:

- (50) a. Articles: **the** paper(s); **a** problem; **an** egg
 b. Demonstratives: **this** paper; **these** papers; **that** egg; **those** eggs
 c. *Wh*-determiners: **what** colour(s); **which** paper(s)
 d. Quantifiers: **some** milk/eggs; **each** paper; **every** boy; **all** cases;
no time; **most** eggs; **few** eggs; **much** time; **any** eggs
 e. Possessive determiners: **my** child; **her/his** child; **our** child; **Lee’s** child
 f. Pronouns: **we/us** linguists; **you** boys

The reasoning behind classifying all these items as members of an overarching category ‘determiner’ is that we can only put one of them in the single slot before a noun in English: ___N. For instance, we get *this child*, but not **this my child*, or **these which eggs*. However, the situation is actually not quite as simple as this, particularly with regards to the quantifiers (words like those in (50d), which specify quantity), because we do get phrases like *my every wish*, *some of the eggs* and so on, which have a more complex syntax.

Note also that in (50e), examples of possessive determiners include whole NPs such as *Lee’s* in *Lee’s child*, or *my cousin’s child*. These seem to fill the same position as single-word possessive determiners, and we can certainly choose only one of them in the pre-noun slot: we don’t get **This is Lee’s her child*. But if the ‘determiner’ position can be a whole phrase, it again suggests that the situation is quite complex syntactically.

Finally, it might seem strange to suggest that pronouns such as *we*, *us*, *you* should be placed in the class of determiners, along with words like *some* and *the*. But the fact that pronouns don’t co-occur with determiners (**the she*) suggests that pronouns aren’t nouns. (Proper nouns – names – can’t generally take determiners in English either, but they can in certain contexts: *The Kim Jones I know has black hair*; *I can hardly recognize the London I once loved*.) Interestingly, pronouns can often replace determiners, which suggests that they may indeed be in the same word class:

- (51) **We/us linguists** aren’t stupid. (*Compare*: These linguists ...)
 I’ll give **you boys** three hours to finish the job! (*Compare*: those boys ...)

One of the properties of such determiners (*we, us, you*) is that they can occur without a following noun:

- (52) **We** ___ aren't stupid.
I'll give **you** ___ three hours to finish the job!

You might doubt that this is a general property of determiners, since *the* and *a* can't occur alone: **The/a* ___ *could be problematic*. However, plenty of other determiners can occur without a following noun and, as (53) shows, they have just the same distribution (= are found in the same places) as a full noun phrase:

- (53) **These/those** ___ are good!
I'll give **some** ___ to Lee.
I'll give **that/this** ___ away.

For reasons like these, some linguists propose that noun phrases are really 'determiner phrases'; we return to this question in Section 4.1.8.

Determiners are paired only with nouns, and don't co-occur with other word classes. For example, we get **Their expects are unrealistic*, where *expects* is a verb – the noun *expectations* would be fine. Knowing that determiners pair up with nouns, we can use them to test for word class. So if we're unsure whether or not, say, *singing* can be a noun, we can try it with a determiner: *This singing is nice; Her singing is awful*. Since these are grammatical, we can conclude that *singing* is a noun here.

Cross-linguistically, determiners are common. They typically occur either initial in the noun phrase, as in English and Japanese, or final, rather than in the middle of the phrase. This becomes clear if the noun phrase also has an adjective: (54) is from Akan, a Kwa language spoken in Ghana:

- (54) mmea nketewa no (Akan)
women PLURAL.small the
'the small women'

There are, however, many languages without the range of determiners that we find in English. For instance, many lack definite and/or indefinite articles (e.g. Russian, Finnish and Chinese). Some languages have one and not the other; for instance, Welsh has definite articles but not indefinite. But there are other ways of distinguishing definite and indefinite nouns, as illustrated by Chinese in (55) (the small functional element glossed as ASPECT serves here to indicate a completed event):

- (55) a. Ta mai pingguo le. (Chinese)
he buy apple ASPECT
'He bought an apple.'

such as *heavy*, *cantankerous*, *supportive* – someone can be supportive, for instance, to a greater or lesser extent. So *very* is unlikely to occur with non-gradable adjectives such as *definitive*, *residual*, *syntactic*.

2.4.3 Adjectives and their grammatical categories

It is common, though certainly not universal, for languages to have the morphosyntactic category known as COMPARISON. In English, we represent the comparison of adjectives in two different ways. The first is morphological, via changes in the form of the adjective itself; for instance, in *straight*, *straighter*, *straightest*, the base form of the adjective *straight* takes a comparative suffix *-er* or a superlative suffix *-est*. The second method is via the addition of a functional element, *more/most*: *more honest*, *most honest*; see exercise 2 in Chapter 1 for an indication of when each strategy is used. Some languages have an extra degree of comparison that doesn't occur in English. For instance, the Celtic family has an EQUATIVE, used in the 'as <Adjective> as' construction. Where English simply uses the base form of the adjective, Welsh has an *-ed* equative suffix:

- (63) Mae 'r cwpan cyn llawn-ed â 'r botel. (Welsh)
 is the cup as full-EQUATIVE with the bottle
 'The cup is as full as the bottle.'

The other morphosyntactic category for adjectives which is widespread is AGREEMENT. Adjectives are often marked to agree with the nouns they modify. Inherent features of the noun such as gender or number are often cross-referenced on a modifying adjective, and in some languages, the case of a noun is also cross-referenced on the adjective. In French and many other European languages, adjectives agree in gender with the noun they modify, changing their form accordingly:

- (64) a. le vin blanc b. la porte blanche (French)
 the.M wine(M) white.M the.F door(F) white.F
 'the white wine' 'the white door'

The noun *vin*, 'wine', is masculine, so the attributive adjective, *blanc*, appears in its masculine form too (as does the determiner *le*). The noun *porte*, 'door', is feminine, so we find the feminine form of the adjective, *blanche* (and also the feminine determiner *la*). It should be noted, though, that many French adjectives do not have distinct masculine and feminine forms. For instance, *noir* 'black(M)' is pronounced identically to *noire* 'black(F)' – the spelling rule, which adds *-e* in the feminine, therefore does not reflect a genuinely different form.

2.4.4 Are adjectives essential?

We saw earlier that not all languages have an open class of adjectives. For instance, Dixon (2004b) reports that the Jarawara language of southern Amazonia has a closed

‘Happiness does/feels on me’. This example uses the ‘adjectival’ noun *wapun*, ‘happy/happiness’:

- (73) *wapun* *kantk-n* *amayak*
 happy.V.SG with-V.SG COPULA.1SG
 ‘I’m happy.’

This literally means something along the lines of ‘Happiness is with me’, and we can see that, like other nouns in Yimas, *wapun* belongs to a specific noun class (class V in this case), and triggers agreement (on *kantkn* ‘with’), as do other nouns in the language.

In sum, it appears that many languages typically either use verbs in place of adjectives, or nouns, for example by saying something like ‘Kim has kindness’ rather than ‘Kim is kind’. Are there, then, languages without a recognizable class of adjectives at all? This is a controversial issue, but in fact, two linguists from very different grammatical traditions have argued recently that all languages do have a formal class of adjectives: Baker (2003), from a Chomskyan generative grammar perspective, and Dixon (2004a), from the broadly functionalist/typological perspective. In some languages, we would have to conclude, this may be a very small class of adjectives, as in Yimas. Whether or not adjectives are an essential word class, they are certainly widespread cross-linguistically.

2.5 ADVERBS

2.5.1 Adverbs and adjectives

In English, central members of the traditional word class of adverbs are words like *suddenly*, *slowly* and *gradually*. These central adverbs are formed from the related adjectives by an affix *-ly*, which turns adjectives like *sudden* into *suddenly*, and so on. Similarly, in French, *-ment* turns *sage* ‘wise’ into *sagement* ‘wisely’, and so on. We can’t, however, identify adverbs in English by their morphology. Numerous adjectives in English don’t take the *-ly* affix at all: *big*, *small*, *ill*, *young* and many more. Some irregular English adverbs have the same form as the adjective: *She works fast/hard* but not **She works fastly/hardly*. Just to confuse matters, there’s an entirely different adverb which does have the form *hardly*, as in *She hardly works*, but which has just the opposite meaning! Conversely, some *-ly* words are definitely adjectives, not adverbs: examples are *ungodly*, *kindly*, *ungainly*, *lonely*. We can tell that these are adjectives because they modify nouns but not verbs: *this ungodly hour*, but not **He speaks ungodly*. One of the chief functions of adverbs is to modify verbs, as in *Kim stopped suddenly*.

Traditionally, English adjectives are distinguished from adverbs because they don’t generally occur in the same syntactic environment. Adjectives modify nouns, such as *song*, as in (74); and adverbs modify adjectives, such as *sad*, (75a), other adverbs, such as *lucidly*, (75b), and verbs, such as *spoke*, (75c). Here, we follow the linguistic practice of putting the phrase in square brackets and indicating its category (NP, AP etc.) at the left edge:

- (74) [_{NP} a **strange** song]
- (75) a. a [_{AP} **strangely** sad] song
 b. She spoke [_{AdvP} **strangely** lucidly].
 c. She [_{VP} spoke **strangely**].

This set pattern of distribution is the only one possible in standard English: compare **a strangely song*, **She spoke strange lucidly*. In fact, in standard English adjectives and adverbs cannot occur in identical positions, but instead occur in what is called COMPLEMENTARY DISTRIBUTION: where one occurs, the other doesn't, but together they cover all the available positions. So, adjectives modify nouns, but adverbs modify the other lexical word classes; together, they modify all the available word classes, and their environments don't overlap. We can predict which will occur in any given syntactic environment. Because adjectives and adverbs complement each other in this way, some linguists consider them to be sub-classes of the same word class. We can regard this to be the adjective class, since this is more basic in form.

To qualify as sub-classes of a single word class, there must also be grammatical properties common to both groups. Adverbs and adjectives fulfil this requirement too. First, they share modifiers: they take the same intensifiers, as in *very/quite/most unusual(ly)*. Second, they can both occur in the *as _____ as* comparative construction: *as miserable as Kim*, *as miserably as Kim*.

Third, the comparative suffixes *-er*, *-est* occur on a few adverbs, such as *soon* (*sooner*, *soonest*) as well as on adjectives such as *red* (*redder*, *reddest*). There are some distinctions: (76) shows that, for example, the adjective *uncertain* can take a following *whether ...* sentence, whereas the related adverb can't.

- (76) He seems **uncertain** whether she's left or not.
 *He spoke **uncertainly** whether she'd left or not.

But on balance, the evidence for treating the central class of *-ly* adverbs in English as a sub-class of adjectives seems convincing.

In many languages, there is no formal distinction between adjectives and adverbs. German illustrates: in (77), *schlecht* 'bad' has the function of a predicative adjective, while in (78), it has the function typical of adverbs, modifying the verb.

- (77) Er ist **schlecht**. (German)
 he is bad
 'He is bad.'
- (78) Er singt **schlecht**.
 he sings bad
 'He sings badly.'

Finally, let's consider words like *still* (as in *I'm still waiting*), *yet*, *always*, *already* and *sometimes*. These aren't related to any adjective, and can't take any of the typical adjective/adverb modifiers: **very already*, **more sometimes*. However, since they

modify verbs (*Kim always ate fruit, She still reads that newspaper*) we can indeed consider them to be a sub-class of adverbs.

2.5.2 The adjunct function

As a word class, ‘adverb’ has traditionally been rather problematic, since it’s been used as a ragbag for any words that don’t neatly fit into the categories of nouns, verbs or adjectives. For instance, in traditional grammar, words like *today, tomorrow, yesterday* and *tonight*, as well as phrases such as *this week, next week*, would be termed ‘adverbs’. Here, we’ll see that they are actually nouns or noun phrases (NP). They can occur in all the typical NP positions, with typical NP functions: as subjects, (79a); direct objects, (79b); and as the objects of prepositions, (79c):

- (79) a. **Tomorrow/today/tonight/this week** seems fine.
 b. I planned **tomorrow/yesterday** very carefully.
 c. I’ll finish it by **tonight/tomorrow/next week**.

And they can also take the -s possessive ending, like other NPs: *today’s bike ride, tomorrow’s lectures, next week’s wedding*. But unlike adverbs, they can’t be modified by the intensifiers *very, quite* and so on: **very tonight, *quite tomorrow*. So we can conclude that *today, tomorrow* etc. are not adverbs at all, and in this respect, the traditional view is incorrect.

Why, then, have these NPs traditionally been termed ‘adverbs’? The reason is that – like adverbs – they often occur not as subjects, objects and so on, but rather as optional modifying phrases. Preposition phrases (PP) can also occur in this same context. For instance:

- (80) We’re leaving **next week/today/tomorrow** (NP).
 We’re leaving **in a week** (PP).
 We’re leaving **rather hurriedly** (AdvP).

What these elements (in bold) modifying *leaving* all have in common is not their word class, but rather, their syntactic function. All of them fulfil what is known as the ADJUNCT function in (80) – they are optional modifying phrases. Confusingly, this function is also referred to as the ADVERBIAL function, no doubt because it is adverbs that often fulfil this function. But as (80) shows, not all adjuncts are adverbs. As we will see in Chapter 3, adjuncts can also be entire modifying sentences.

2.6 PREPOSITIONS

2.6.1 Identifying prepositions in English

In English, but not in all languages, we find phrases like *under the floor, towards that conclusion, outside my house*, where a PREPOSITION (the word shown in bold)

has combined with a noun phrase to form a preposition phrase (PP). Perhaps the most typical role of prepositions is to mark locative and temporal information in a language – that is, information concerning location and time. In English, prepositions such as *under, over, into, on (top of), beside, towards, in (front of)* mark location, while prepositions such as *before, during, after, while, until* and *since* mark temporal information: *before the meeting, during the war, until four o'clock*. Many prepositions express either kind of meaning: *after the game, after the traffic lights; over the bridge, over the summer*. Prepositions also express the manner in which an event is carried out: *with a knife, by means of poison, in a loud voice*, and so on. There are also metaphorical uses of prepositions: compare *against the kerb* (spatial) and *against my better judgement*.

In terms of function, many PPs are optional modifiers of verbs, as in *We left [before the meeting], She sang [in a loud voice]* – the PPs are in brackets. In this grammatical function, a PP is an adjunct, as we saw in Section 2.5: an optional, modifying phrase.

Now let's start to identify the preposition class in English. Just like nouns, adjectives and adverbs in English, prepositions pair up with their own special set of modifiers: these are *straight, right, well* and *just*, and we can also add the more restricted modifier *bang*. All of these (underlined) immediately precede the prepositions (in bold) in (81):

- (81) The weight is well/just **inside** the limit.
 We were bang on target/**on** time.
 She pushed the box well/right/straight/just **under** the bed.
 Go straight/right **to** the top of the stairs!
 The library is just/right **by/beside** the town hall.

Although the ability to take these modifiers is a good test for preposition status in English, it does need to be used with caution, because some of the modifiers can occur with word classes other than prepositions (e.g. *just fine*, where *fine* is an adjective). Also, not all prepositions work with all of these modifiers, most often because their meanings are not compatible. A final note of caution is that the purely grammatical preposition *of*, as in *the top of the stairs*, cannot take a modifier either. The modifiers do, however, enable us to identify various other words as prepositions when we might otherwise not have been sure of their word class.

First, let's consider words like *afterwards* and *nearby*. As we will see, these can be classified as intransitive prepositions. So far, the prepositions we've seen were used transitively: they take an object NP. Examples are *inside the limit, on time, under the bed*, where the prepositional objects are underlined. Though most prepositions are transitive, a number can be used either transitively or intransitively, i.e. without an object; examples are *inside, over, before*, as in *That student was here before (the others)*, and *underneath*, as in *Put your case underneath (the bed)*. The prepositions *afterwards* and *nearby* differ only in that they are always intransitive:

- (82) I'll see you right/straight/just **afterwards**.
She lives right/just **nearby**.

The co-occurrence with the modifiers *right*, *straight* and *just* identify *afterwards* and *nearby* as true prepositions.

Second, consider words like *upstairs*, *overhead* and *online*. Traditionally, these would be termed 'adverbs', but using the modifiers *just* and *right* as a test for preposition status, they are shown to be prepositions:

- (83) She lives right/just **upstairs/downstairs**
The plane flew just/right **overhead**.

Third, we can re-evaluate what are traditionally termed 'verbal particles'. The term refers to the small words that go together with verbs in 'phrasal verb' expressions like *run down*, *put back*, *take over* etc. Not only do they look identical to prepositions, these 'particles' are also classified as prepositions by the *right* test. The prepositions are again in bold:

- (84) Lee ran his apartment right **down**.
Put those chocolates right **back**!

Prepositions are used widely in English, and although not all members of the word class behave in a standard way, they do share properties in common.

2.6.2 Postpositions

So far, we've considered words like *in*, *over*, *beside* in English, which are called 'prepositions'. When these prepositions are transitive, their object NP follows the P, as in *over the summer*. However, in some languages, the object NP always precedes the P, as in Japanese:

- (85) a. tookyoo **kara** (Japanese)
Tokyo from
'from Tokyo'
b. sono hito **to**
that person with
'with that person'

In Japanese, these words in bold, *kara* 'from' and *to* 'with', are not prepositions, but **POSTPOSITIONS**: they follow the NP which is their object. The cover term for the whole word class is **ADPOSITION**, meaning both prepositions and postpositions. In Chapter 4, we return to questions of word order of this kind.

2.6.3 Grammatical categories for adpositions

In most languages, there are no adpositional inflections: only the major lexical word classes noun, verb and adjective are typically associated with any morphosyntactic

categories. In other words, we don't often expect to find 'endings' (or other kinds of affix) on prepositions and postpositions. However, a minority of languages do have inflected prepositions. Well-known examples are the Celtic and the Semitic families. In the Celtic language Irish, for example, prepositions inflect to show person, number and gender:

- | | | | | | | | |
|------|----|---------------------|----|---|----|---|---------|
| (86) | a. | <i>le</i>
'with' | b. | <i>leis</i>
with.3SG.M
'with him' | c. | <i>léithi</i>
with.3SG.F
'with her' | (Irish) |
|------|----|---------------------|----|---|----|---|---------|

The preposition 'with' is *le* in its citation form (the one in the dictionary, for instance) but there is a distinct form of the preposition for each person and number, and distinct genders in the third person singular forms. In Irish, these inflected forms replace the free pronominal objects of prepositions that we find in most other languages (*with him*, *with her*, etc.).

2.7 CONCLUSION

This chapter provides an overview of the distribution, function and morphosyntax of the major lexical word classes, verb, noun, adjective and adverb, as well as the adposition class. Word classes are distinguished by their morphosyntactic categories, their functions and by their patterns of distribution. 'Distribution' covers both the slots which words can appear in, and the modifying words that co-occur with them. We saw that specific functional elements (small, closed class words) often pair up with a particular lexical word class such as a noun or an adjective. To count as a distinct word class, a set of words must have some properties which distinguish them from other word classes in the language. If we don't find any such properties, then it would be unscientific to make artificial divisions in the data. It is important, then, not to expect all languages to look the same. For instance, we shouldn't think that just because, say, English and Italian have an open class of adjectives, then all languages must have one.

We have seen that grammatical information can be represented either morphologically (that is, via changes in the form of words from major classes) or, alternatively, by the use of separate 'functional' elements. Although both methods of representing grammatical information can occur within a single language, languages tend to lean towards one method or the other. Languages that have a lot of morphology represent grammatical information without needing many of the small, purely grammatical, 'functional' words. Good examples are the African Bantu languages, native American languages and, within Europe, Greek and the Slavonic family (Russian, Polish etc.), as well as non-Indo-European languages such as Finnish and Turkish. Conversely, languages with little morphology, such as Chinese, Vietnamese, Cambodian, and Malay/Indonesian, tend to need more of the small 'functional' elements to represent grammatical information.

FURTHER READING

Elementary reading on word classes in English can be found in Aarts (2008). A vastly more detailed treatment of word classes and most other aspects of English grammar can be found in Huddleston and Pullum (2002). A smaller, student's version is Huddleston and Pullum (2005). More advanced and technical readings on identifying word classes (also known as parts-of-speech) are Lyons (1966), Schachter (1985) and Emonds (1986), papers which tackle the problems from very different angles. A more advanced and technical discussion of the lexical categories verb, noun and adjective can be found in Baker (2003), a book which argues that these three word classes can be recognized universally, despite some claims to the contrary. Hurford (1994) is an indispensable guide for the beginning syntax student, providing definitions and examples of many of the concepts that I will be using throughout. See also Aarts (2008) for the concepts of 'subject' and 'object'. On grammatical categories, see particularly Anderson (1985) and Chung and Timberlake (1985). More detailed information can be found on aspect in Comrie (1976) and on tense in Comrie (1985a); see also Whaley (1997: Chapter 12). General help with describing syntax and morphosyntax for the beginning student can be found in T. Payne (1997, 2006).

EXERCISES

1. This exercise is intended to help you consolidate the notion of 'subject' in English.

Task: (i) Identify all the subjects in each of the following examples, using the tests established in Section 2.3.2. (ii) List the categories of phrases that can be subjects, apart from NP. Add any other examples you can think of. You can turn the noun phrases into pronouns where possible, in order to test for the nominative case forms, and you can change the tense of the verb or auxiliary, in order to test for subject/verb agreement. You can also try changing the person and number of the putative subjects, to see how this affects case and agreement. Some of these examples contain more than one clause (a concept discussed in Chapter 3); this means that there may be more than one subject in some examples. Pick out all you can find. (iii) Flag up any interesting issues raised by these data, or other examples you've thought of.

- (1) Next week might turn out to be fine and sunny.
- (2) Political science and engineering are the most popular subjects at the University of Northumberland.
- (3) Some days when she's working in town, Mel cycles to work.
- (4) Yesterday's news really upset me.
- (5) Given the problems posed by the mass protests, there's little hope that this government will survive.
- (6) These old black-and-white films on TV, Kim just can't bear to watch.
- (7) From Durham to Newcastle takes 15 minutes on the train.

- (8) After the storm, a flood of disgusting brown water surged through the streets.
- (9) The regeneration of run-down inner city areas sometimes leads to a great demand for housing.
- (10) The original Georgian glass in that window was shattered by the boys' football.
- (11) Where Kim's going on holiday is a big secret.
- (12) Just jealous of my vast wealth is what you are.
2. This exercise is intended to get you to think carefully about English word classes. In each of the following examples, decide on the word class of the items in bold. Consider the evidence given throughout this chapter and note any problems posed for it by the data here. Remember to include the evidence provided by the words which modify the items in bold, and try adding different modifiers to help with your diagnoses. Give as much evidence as possible for your answers, looking at the distribution, morphology (inflections) and function of the words. It will probably help to compile your own list of the relevant properties for nouns, verbs, prepositions and adjectives. There are 16 example sentences; you should use at least 10 of these in your answer.

Hints

- Examples marked with % are restricted to certain dialects of English. Of course, you may not find them grammatical if you don't speak such a dialect, but the point is that they provide evidence for how words are used in certain varieties of the language.
- Some of the words pose quite a challenge; if you can see evidence pointing in more than one direction (for instance, some word might be an adjective or might be a preposition), note this too.
- Remember that words can fall into more than one class, in different contexts. Just because something is, say, an adjective in one context doesn't necessarily mean that it is an adjective in all other contexts.

- (1) She lives just/right/really/%real **near** the shops.
- (2) You can't get any **nearer** than the **nearest** supermarket.
- (3) We're just **delighted** to hear your good news.
- (4) We walked **lengthways** across the quad/**clockwise** round the gardens.
- (5) She studies an **unwritten** language. *Compare:* *She unwrote the language.
- (6) This film is **devoid** of meaning.
- (7) It fell straight **apart** the moment I opened it.
- (8) I'm still quite **undecided**.
- (9) %We had a right **tasty** meal.
- (10) **Afterwards**, we travelled **southwards towards** the mountains.

- (11) I'm not **that bothered** about the exams.
 (12) I'm not too **conversant** with **that software**.
 (13) This proposal is well **worth considering**.
 (14) The kids ran **aboard** (the ship) as **soon** as they could.
 (15) The boat floated **downstream** and drifted **ashore**.
 (16) **Since** the war, the journalist has lived **overseas**.

3. In (1) through (6) that follow are some examples from Malay.

Task: (i) Go through them, noting as many grammatical differences between Malay and English as you can; there are around half a dozen things to spot. Pay particular attention to the pronouns – words such as 'I/me/my', 'he/him/his' in English.

- (1) Saya sayang dia.
 I love he/she
 'I love him/her.'
- (2) Dia sayang saya.
 he/she love I
 'He/she loves me.'
- (3) Kawan saya doktor.
 friend I doctor
 'My friend is a doctor.'
- (4) Buku ini mahal.
 book this expensive
 'This book is expensive.'
- (5) Buku-buku itu murah.
 book-book those/the cheap
 'Those/the books are cheap.'
- (6) Maria membeli sepasang kasut untuk saya.
 Maria buy pair shoe for I
 'Maria bought a pair of shoes for me.'

(ii) Can you say how Malay distinguishes the subject of the sentence from the object? Is this the same as English or different?

4. The following data are from the language Zina Kotoko, a Chadic language of Cameroon, and are courtesy of Anders Holmberg. Examine each sentence and note as many grammatical differences as you can between English and Zina Kotoko; you should find up to six. Describe these differences using the correct terminology.

- (1) Mafu de majakwi
tree.PL the tall.PL
'The trees are tall.'
- (2) Adam majakwa b'da.
Adam tall.SG NEG
'Adam is not tall.'
- (3) Adam kwice asu de da ghika b'da.
Adam cut meat the with knife NEG
'Adam didn't cut the meat with a knife.'

5. Examine the data that follow, from Japanese, and try to figure out the function of the particle *-no*, which I have left un glossed. In (1) through (4), the particle is seen in its central usage. First, decide what this is:

- (1) Hanako-**no** musuko
Hanako-NO son
'Hanako's son'
- (2) boku-**no** haha
I-NO mother
'my mother'
- (3) kono e-**no** namae
this painting-NO name
'the name of this painting'
- (4) Yamada-**no** kaban
Yamada-NO bag
'Yamada's bag'

In (5) through (8), the use of the particle is extended somewhat. In what way do these examples differ from the first four? How is the particle used?

- (5) kaihatu-**no** keikaku
development-NO plan
'a plan for development'
- (6) kaji-**no** gen'in
fire-NO cause
'the cause of the fire'
- (7) suugaku-**no** sensei
mathematics-NO teacher
'a mathematics teacher'
- (8) nihon-**no** hata
Japan-NO flag
'Japanese flag'

The data in this exercise are taken from Tsujimura (1996) and Iwasaki (2002).

6. The data in this exercise are from two Melanesian languages, Nakanamanga (or Nguna), and Fijian, and are taken from Lynch (1998). Both data sets illustrate the fact that verbs in these languages have different morphosyntactic properties from verbs in familiar European languages.

Task: A distinctive grammatical category is represented by a marker on some of the verbs in these two data sets. What is it, under what circumstances does it occur in each data set, and under what circumstances does it not occur? Remember that a list of abbreviations is given at the start of this book. The data you have are representative, so you have enough data to answer the question. Discuss briefly any other interesting properties of the verbs in these examples.

A. Nakanamanga (Nguna)

- (1) A ga munu.
I INTENTIONAL drink
'I'll drink.'
- (2) A ga munu-gi noai naga.
I INTENTIONAL drink-TRANS water that
'I'll drink that water.'
- (3) A ga munu-gi-a.
I INTENTIONAL drink-TRANS-it
'I'll drink it.'

B. Fijian

- (4) E bulu
he bury
'He/she/it is buried.'
- (5) E bulu-t-a na benu.
he bury-TRANS-it the rubbish
'He/she buried the rubbish.'
- (6) E moce na gone.
he sleep the child
'The child slept/is sleeping.'
- (7) E gunu yaqona o Seru.
he drink kava the Seru
'Seru is drinking kava.'
- (8) E gunu-va na yaqona o Seru.
he drink-TRANS the kava the Seru
'Seru is drinking the kava.'
- (9) E na lako mai o Jone.
he FUT go here the John
'John will come.'

- (10) E rai-ci ira.
he see-TRANS them
'He saw them.'
- (11) E davo-ra-a.
he lie-TRANS-it
'He lay on it.'
- (12) Eratou sã lako vata sara yani.
they.few ASPECT go together intensive there
'They (few) went off there together.'
- (13) E loma-ni koya.
he love-TRANS her
'He loves her.'

Looking inside sentences

This chapter begins an examination of the internal structure of sentences which takes up the remainder of the book. Section 3.1 examines finite and non-finite verbs and auxiliaries, and distinguishes between simple sentences and complex sentences – sentences which contain other sentences. Subordination is the term used for a construction in which a sentence is embedded (or contained) within another sentence. Section 3.2 is an introduction to subordination in English and other languages. Although subordination is common cross-linguistically, not all languages seem to make much use of it. Section 3.3 examines some cross-linguistic variation in clause types, particularly in complex constructions.

3.1 FINITENESS AND AUXILIARIES

3.1.1 Independent clauses

Linguists often divide the sentence into two main parts: the SUBJECT and the PREDICATE. As we saw in Chapter 2, the central role (or ‘head’) in the predicate is normally filled by a verb, but we also find other types of predicate, such as adjectival predicates and nominal predicates. A verbal predicate consists of the head verb plus any phrases modifying the verb, or selected by the verb.

Let’s examine the data in (1). In (1a), the subject is *Kim* and the predicate *waited*; in (1b) the subject is *these guys* and the predicate *like chips*; and in (1c), the subject is the whole phrase *The first-year students in our department* and the predicate is *bought a lot of books at this stage in the year*.

- (1)
- a. Kim waited.
 - b. These guys like chips.
 - c. The first-year students in our department bought a lot of books at this stage in the year.

These examples each illustrate SIMPLE SENTENCES. ‘Simple’ here is a technical term, meaning ‘consisting of just one clause’. To avoid conflict with the (non-linguistic) idea of a sentence as something that starts with a capital letter and ends with a full stop, here I introduce the more precise term CLAUSE. The term ‘clause’ has a specific meaning: it’s a sentence that contains one predicate. As we will see in this chapter, some sentences contain only one clause, and others contain more than one clause.

grammatical categories associated with verbs, such as agreement for person and/or number. Strictly speaking, finiteness is a property of an entire clause, rather than just a verb, and for some languages, finiteness may well not be indicated via the verbal morphology at all. For instance, if a language has nominative case (see Chapter 6), this typically occurs on the subjects of finite clauses, so this is another diagnostic. Finally, some languages only have finite clauses, for instance Mohawk, Nahuatl, Nunggubuyu and Ainu.

Examples (2) through (4) illustrate independent clauses – and therefore simple sentences – in three other languages, each of which expresses finiteness in different ways. The verbs and associated morphology are in bold:

- (2) Dytyna **spyt'**. (Ukrainian)
 child sleep.PRES.3SG
 'The child is asleep.' (literally 'The child sleeps.')
- (3) **Na-bànjäl-ya** na ana-na lai nyungga. (Kambara)
 3SG.SU-put-3SG.OBJ the child-3SG at I
 'He left his child with me.' (literally 'He put his child at me.')
- (4) Ape yu ati **o de**. (Ndyuka)
 there your heart FUT be
 'Your heart will be there.'

The Ukrainian verb is marked for tense and also the person/number of the subject; all this information is fused together, so that there are no separate morphological markers for 'present tense' or 'third person'. This is very common in the verbal morphology of European languages.

In the Kambara example, the finite verb has bound pronominals: person/number markers representing both the subject and the object. But there is no tense marker at all. The 3SG.SU prefix *na-* on the verb in (3) means a third person singular subject. This is translated as a pronoun *he* in English, but the Kambara has no independent pronoun. The 3SG.OBJ suffix *-ya* marks a third person singular object, referring to the child. (You can refresh your memory for such glosses by re-reading Section 1.2.2.3.)

In Ndyuka, (4), the verb *de* 'be' itself has no morphology indicating tense (or any other morphosyntactic category), but there is an independent future tense marker, *o*. Therefore, the clause is finite.

Cross-linguistically, most independent clauses contain finite verbs, as in (2) through (4). Some languages, though, allow independent clauses consisting of a subject and a predicate that is non-verbal, as we first saw in Chapter 2. So in (5), the predicate (in bold) is just an adjective phrase *nadif katir* 'very clean', and this sentence contains no copula (i.e. no word for 'is'):

- (5) al-bēt dā **nadif katir** (Chadian Arabic)
 DEF-house this.M.SG clean very
 'This house is very clean.'

- **Modal auxiliaries**

MODAL AUXILIARIES are a group of independent words in English which express such concepts as permission, necessity or ability. In some languages similar kinds of meaning are expressed by verbal inflections. English MODALS are distinct from other auxiliaries, and also distinct from main verbs: first, the modals only occur in a finite form, and second, they don't take the third person singular *-s* inflection in the present tense. We don't get such forms as **She mays leave* or **Kim wills arrive soon*. They do, however, mostly have contrasting finite forms which are technically considered to be present and past tense, though their relationship to actual time reference is pretty complex in English. So in these pairs of modals, the first is present tense and the second, past tense: *can/could*; *shall/should*; *may/might*; *will/would*. *Must* is also a modal auxiliary, but it has no distinct past tense form. All these modals precede the bare uninflected form of the verb which is known as the INFINITIVE, such as *leave*, *arrive*. A few elements are generally regarded as modals (e.g. *ought*, *need*), and their meaning is consistent with other modal auxiliaries, but they have exceptional syntactic behaviour in various ways. For instance, they precede *to* + infinitive, as in *Lee ought to leave*, *I need to go*.

I've already noted that in English only one element per clause can be finite, and that this is the first in the sequence of auxiliary/verbal elements. You can be sure, then, that in sequences such as *may leave*, *will arrive*, *must sleep*, *can dream*, only the modal auxiliary (in bold) is finite, and therefore the main verbs (*leave*, *arrive*, *sleep*, *dream*) are all NON-FINITE here. This means that they carry no information about tense, person or number.

- **Have and be: main verbs and aspectual auxiliaries**

The elements *have* and *be* in English have two distinct uses: they can be either main verbs or auxiliaries. Ellipsis aside, when they appear as the only verb in the clause, then by definition they must be the main verb. (7) illustrates MAIN VERB *have* and *be* (in bold):

- (7) Kim **isn't** sure about that.
I **had** a cold last week.
Are you a friend of Kim's?

(8) illustrates *have* and *be* in their other function, as ASPECTUAL AUXILIARIES (in bold). Note that each example contains additional verbal elements, including the main verbs *leaving*, *written/played/sung*, *enjoying*:

- (8) a. We're just leaving.
b. Jo **has** often written/played/sung to me.
c. They **have been** enjoying better weather lately.

ASPECT is a grammatical category of verbs which expresses such information as whether the action of the verb is completed or unfinished (Section 2.2.2.1). Two kinds of aspect are illustrated in (8). Auxiliary *be*, along with the *-ing* form of the

main verb, as in (8a), gives PROGRESSIVE aspect (an unfinished or ongoing action); *been enjoying* in (8c) is also progressive. In (8b), *has written/played/sung* illustrates PERFECT aspect, which in its basic meaning refers to a completed event, but one which still has relevance to the time of the utterance. In (8c), *have been* is another example of the perfect. Note from (8c) that progressive and perfect aspect can co-occur. Perfect aspect in English requires auxiliary *have* plus a special form of the main verb known as the PAST PARTICIPLE, which in regular verbs ends in *-ed* (*played*) and *-en* (*written*) in numerous irregular verbs.

Main verb *have* and *be* can also co-occur with auxiliary *have* and *be*: *She has **had** a cold recently; They have been **having** better weather*. The auxiliary forms are underlined, and the main verb forms are in bold type. These examples also show that in English, the main verb always comes after any sequence of auxiliaries. There can be three auxiliaries or more in one clause, as in (6b): *The people in the library may have been **working** late*.

To summarize our findings for English:

Finiteness and auxiliaries in English

- A normal simple sentence in English has *one* (and only one) finite element, which may be an auxiliary or a main verb.
- The finite element always occurs *first* in the sequence of auxiliaries/verbs.
- All other auxiliary and verbal elements in the clause are therefore NON-FINITE.
- The main verb always *follows* any sequence of auxiliaries.
- English *have* and *be* occur both as main verbs and as auxiliaries.
- Auxiliary *have* + past participle of verb gives the perfect aspect, e.g. *has written, had played*.
- Auxiliary *be* + *-ing* form of verb gives the progressive aspect, e.g. *is writing, was playing*.

3.1.4 Ways to express the grammatical categories for verbs

Many Indo-European languages (the family that English belongs to) also use the equivalents of 'have' and 'be' as auxiliaries, as does the entirely unrelated European language, Basque. But cross-linguistically, there is a great deal of variation in whether auxiliaries are used at all, and if they are, what they are used for. In all languages, 'Auxiliaries are words that express the tense, aspect, mood, voice, or polarity [= negative or affirmative characteristics] of the verb with which they are associated: i.e. the same categorizations of the verb as may be expressed by means of affixes' (Schachter 1985: 41). This means that any of the morphosyntactic categories that are associated with verbs (see Section 2.2.2) can also be expressed by an auxiliary in some language or languages. We saw earlier that, in English, morphosyntactic information about finiteness can be expressed on a main verb or on an auxiliary, but not both within a single clause. In some languages, a verb and an auxiliary in

the same clause both carry the grammatical information, for example by both being marked for tense, as in the Australian language Warlpiri.

We can now see that there are three different ways of expressing all the grammatical categories for verbs: (a) via the verbal morphology itself, (b) via an auxiliary, or (c) by using an independent word. Let's look at these three strategies now. We have seen, for instance, that English expresses progressive and perfect aspect using auxiliaries plus main verbs. The Brazilian language Bare doesn't use auxiliaries; instead, it expresses both progressive and perfect aspect just by inflections on the main verb (these affixes are shown in bold). This, then, is the verbal morphology strategy:

- (9) yaharika nu-tikuwá-**ni** (Bare)
 now 1SG-lie-PROGRESSIVE
 'I am lying down now.'
- (10) i-tíkua-**na**
 3SG-lie.down-PERFECT
 'He has lain down already.'

(Note that once again these examples do not contain actual independent pronouns for 'I' and 'he', just verbal inflections which perform the same work: first person singular in (9) and third person singular in (10). These are known as pronominal affixes.)

Conversely, some languages have auxiliaries not found in English. Evenki, a Tungusic language of Siberia, has a negative auxiliary. In (11), the main verb *duku* 'write' is finite: it has tense and person/number inflections. But in (12), the finite negative auxiliary bears these inflections instead, and the main verb *duku* 'write' is non-finite (Section 3.1.5) – it no longer has the tense and agreement suffixes found in (11). The Evenki main verb and the auxiliary in (11) and (12) take the same basic affixes, although the PAST affix is pronounced rather differently in (12):

- (11) Bi dukuwūn-ma duku-cā-w. (Evenki)
 I letter-ACC write-PAST-1SG
 'I wrote a letter.'
- (12) Bi dukuwūn-ma ə-ə-w duku-ra
 I letter-ACC NEG.AUX-PAST-1SG write-PARTICIPLE
 'I didn't write a letter.'

The English translation in (12) also uses an auxiliary, *didn't*, for the negation. But *do* is not inherently negative, while the Evenki auxiliary is. English expresses negation by using a separate morpheme, *not*, which can optionally be amalgamated with auxiliaries (*isn't*, *shan't*, *won't* etc.). So here we see the third method of expressing a grammatical category associated with verbs: by using an independent morpheme like *not*.

To summarize, this section has shown that the grammatical information associated with verbs is mainly represented in three different ways: with verbal morphology,

with an auxiliary, or by adding an independent word. These alternative means of expressing information (via separate words or via affixes) recur throughout grammars, not just in the verbal systems, and I will indicate other examples from time to time.

Major ways to express grammatical categories for verbs

- Via inflections on the main verb itself. See (2), (3), (9), (10), (11).
- Via a separate word or particle; an independent grammatical word. See (4); also English *not* as described earlier.
- Via an auxiliary. See (6), (8), (12).

3.1.5 Non-finite verbs

NON-FINITE verbs in English are not marked for tense, person/number agreement or any of the other grammatical categories associated with finite verbs, such as aspect or mood. This is very often true of other languages as well, but not all, as we will see. I divide non-finite verbs into the two main types that occur cross-linguistically, INFINITIVES and PARTICIPLES. English has an infinitive plus two different participles.

- Infinitives

It is not easy to provide a satisfactory cross-linguistic definition of the term ‘infinitive’, and forms corresponding to the English infinitive are not particularly common in other languages. Some languages mark the infinitive with special inflections: for instance, French has the suffixes *-er* (as in *dessin-er* ‘to draw’), *-ir* (as in *fin-ir* ‘to finish’) and *-re* (as in *vend-re* ‘to sell’). In English, the infinitive is the bare verb stem, with no inflections: examples are *eat*, *relax*, *sing*, *identify*, *cogitate*. As we’ve already seen in this chapter, however, this property is not sufficient to identify an infinitive in English, since finite verbs in the present tense also have this same form: *I sing*, *you sing* and so on, apart from the third person singular (*sings*).

We can identify English infinitives instead by their distribution. Modal auxiliaries in English require a following infinitive, as in *Kim must ____ (that)*. An infinitive also occurs after *to* in environments such as *I had to ____ then*; *For you to ____ now would be good*. This *to* is an INFINITIVAL MARKER, not to be confused with the entirely different *to* which is a preposition (and which is followed not by a verb, but by a noun phrase).

A distributional test for English infinitives

- Following a modal auxiliary or form of auxiliary *do*, e.g. *must leave*, *could eat that cake*, *can’t relax*, *does love chocolate*.
- Following the infinitival marker *to*: *To err is human*, *We ought to be leaving*, *I have to arrive on time*, *Kim wants Lee to sing*.

- Participles

Participles are widespread cross-linguistically. In Indo-European languages, the term ‘participle’ is generally used to refer to the types of non-finite verbs which primarily co-occur with a finite auxiliary. Such an example is also given from Evenki in (12).

Cross-linguistically, participles are considered to be verb forms that can also be used in positions normally filled by adjectives or nouns. Two examples from German follow. The words in bold are known as ‘present participles’; though derived from verbs, they behave exactly like adjectives in modifying a noun, and in taking the same gender agreement suffixes that adjectives normally take. So in (17), *glaubend* takes the masculine ending *-er*, agreeing with a masculine noun, while in (18), *gehend* takes the feminine ending *-e*, agreeing with a feminine noun:

- (17) ein **glaubend-er** Priester (German)
 a.M believing-M priest.M
 ‘a priest who believes’ (*Literally, ‘a believing priest’*)
- (18) eine **gehend-e** Person
 a.F walking-F person.F
 ‘a person who’s walking’ (*Literally, ‘a walking person’*)

It’s quite common for languages to have a number of distinct participles (e.g. Basque, Armenian and Lezgian), though English has only two different participles. In languages other than English, verbal categories such as tense and aspect are often marked on participles, not just on finite verbs. Some languages, perhaps rather surprisingly from a European perspective, have only a CLOSED class of finite verbs, but an OPEN class of participles. For instance, in the Australian language Wakiman, the finite verb class has only around 35 members, while participles are a genuinely open class of verbs (Cook 1988).

We’ll now consider in a little more detail the two distinct PARTICIPIAL forms in English – the *-ing* form and the *-ed/-en* form. Note that the morphology (each has its own suffix) distinguishes the participles from the English infinitive, which is the bare verb stem.

- The *-ing* participle

What traditional grammars term the ‘present participle’ is the *-ing* form of the verb which, together with auxiliary *be*, gives progressive aspect, as in (19a). But the *-ing* form doesn’t only co-occur with an auxiliary: the verb form *laughing* also appears on its own in the other examples in (19):

- (19) a. Kim was **laughing** loudly.
 b. Kim kept on **laughing**.
 c. **Laughing** loudly, Kim rushed into the room.
 d. I found Kim **laughing** in the corner.

However, not all words with an *-ing* suffix are participles, or indeed verbs of any kind, as the usual distribution tests show. For instance, *boring* is clearly an adjective

in *this very boring film* – it co-occurs with the adjectival modifier *very*. Compare this with a (**very*) *sleeping child*, where *sleeping* is participial (i.e. a verb form), so can't be modified by *very* – remember that the asterisk inside the parentheses means that the example is ungrammatical if that word is included. Another English example is *a burning branch*, where again, the participial form *burning* does not behave like an adjective. Other *-ing* forms can be nouns; *singing* is a noun (a form traditional grammar refers to as a gerund) in such contexts as *Their singing was beautiful*.

- The past participle

The past participle of most English verbs has the *-ed/- (e)n* ending, as in *played, shown, seen, forgotten*. In English, this form of the verb, together with auxiliary *have*, gives the perfect aspect. Some examples (with the past participles in bold) are: *Have you **eaten** the cake?*; *Kim has **had** flu*. There are many irregularities in the form of English past participles. Although some verbs have distinct past participle forms (e.g. *eaten, known*), these are all irregular verbs. Regular verbs have past participles which are identical to their PAST TENSE, such as *worked, left, decided*: they both have an *-ed* ending. It is important that you understand the distinction between past participle and past tense. A simple distribution test can help you to tell which is which:

(20) **Distribution test to distinguish between English past participle and past tense**

- Kim has ____ (that) already.
PAST PARTICIPLE, e.g. *eaten, forgotten, written, left, decided, played*
- Kim ____ (that) yesterday.
PAST TENSE, e.g. *ate, forgot, wrote, left, decided, played*.

In English, as well as a number of other European languages, such as French and German, past participles are also used in the PASSIVE construction (see Section 7.1) as in *This book was **written** last year*, or *It has been **made** into a film*.

English also often uses a past participle to modify a noun, as in a ***boiled** egg*, a ***baked** potato*. These are verbal rather than adjectival; for instance, they don't take any of the typical adjectival modifiers discussed in Section 2.4.

3.1.6 Co-ordination of clauses

So far in this chapter, we have looked mostly at simple sentences: sentences containing only one clause. COMPLEX SENTENCES are sentences that consist of more than one clause. One way that complex sentences are formed is by CO-ORDINATION. In (21), we see three independent clauses. These simple clauses can be joined together, or CO-ORDINATED, to form a complex sentence, as in (22):

- Kim arrived early.
Lee was half an hour late.
Ceri didn't even show up.
- Kim arrived early **and** Lee was half an hour late, **but** Ceri didn't even show up.

The words in bold are CO-ORDINATING CONJUNCTIONS (another in English is *or*), used to CONJOIN (= join together) strings of simple sentences. In clausal co-ordination, each clause could stand alone as an independent clause, and there are no syntactic restrictions on the order of the clauses, though there may be pragmatic restrictions (the sentence may not make good sense if the clauses are re-ordered).

All the clauses in a co-ordination have equal syntactic status – no clause is dependent on any other. As we will see in Section 3.2, this is not the case in complex sentences involving SUBORDINATION.

3.1.7 Summary

Simple sentences consist of only one clause, and most contain a finite verb, although some languages allow sentences with no finite verb, or no verb at all. The finite element may be either a main verb or an auxiliary; a finite auxiliary always co-occurs with a main verb, which is usually non-finite. In English, the finite verb always appears before any non-finite verbs, and if there are any auxiliaries, the main verb always follows them. Cross-linguistically, non-finite verbs fall into two major categories: the infinitive, and participial forms of the verb, which often combine with aspectual auxiliaries to give various categories of verbal aspect. Simple sentences can be conjoined to give a type of complex sentence where all the clauses have an equal syntactic status: this is co-ordination. We turn next to an examination of subordination: a subordinate clause is one that is dependent on another clause in some way.

3.2 INTRODUCTION TO SUBORDINATION

3.2.1 Complement clauses

In (23), the clauses do not all have an equal syntactic status. Each of these examples has two clauses: a MAIN or MATRIX clause, and a SUBORDINATE clause which is embedded within the matrix clause. The subordinate clauses are all in square brackets in (23), and the verbs in the matrix clause are in bold. The subordinate clause is dependent on the matrix clause, as we'll see in a moment:

- (23) a. My friend **claimed** [(that) Ceri liked chips].
 b. **I wondered** [whether/if Lee had gone].
 c. They **want** [to leave before breakfast].

Each of the bracketed subordinate clauses is an obligatory ARGUMENT of the verb in the matrix clause. In other words, these verbs (*claim*, *wonder*, *want*) need a particular kind of syntactic phrase to complete their meaning. We can't have sentences like **My friend claimed* or **I wondered* or **They want* – these wouldn't be complete. In some of these cases, we could just have a direct object as the argument of the verb: for instance, *They want an egg*. Other verbs, though, like *wonder*, in fact require an argument which is an entire clause. The subordinate clauses specify what was

claimed, wondered or wanted. Subordinate clauses that are selected by a verb in this way are known as COMPLEMENT clauses.

You can see from these examples that subordinate clauses have some distinctive properties. First, they are often introduced by a small functional element known as a COMPLEMENTIZER. In (23), *that*, *whether* and *if* are all complementizers. Complementizers can typically be omitted if they don't bear any real meaning, and this is true of English *that* in (23a). But *whether* and *if* couldn't be omitted. In fact, the matrix verb *wonder* selects a clause that starts with a complementizer of this kind, whereas *claim* selects a finite clause introduced (optionally) by *that*. We can't switch these around:

- (24) a. *My friend claimed whether/if Ceri liked chips.
 b. *I wondered that Lee had gone.

The verb in the matrix clause not only selects a subordinate clause, it selects a subordinate clause with specific properties, and often, a specific type of complementizer.

A second property of subordinate clauses concerns finiteness. Independent clauses in English must be finite, as we've seen. The verb in the highest matrix clause must be finite too. But many subordinate clauses contain only a non-finite verb form. This is the case in (23c), where *to leave before breakfast* is a non-finite clause. The verb *want*, then, selects a non-finite clausal complement. From the examples so far, you should be able to see that these subordinate clauses are syntactically dependent on the matrix clause, or more specifically, on the verb in the matrix clause.

A third property of (some) subordinate clauses is also seen in (23c), *They want [to leave before breakfast]*. Here, the non-finite complement clause *to leave before breakfast* has no overt subject; it has only an understood subject, referring back to *they* in the matrix clause. This is a sure signal in English that we are dealing with a subordinate clause. An alternative option to (23c) is a non-finite subordinate clause with an overt subject: *They want [the girls to leave before breakfast]*. But then it is clearly understood that this subject, *the girls*, refers to a different entity from the matrix subject, *they*.

A subordinate clause is part of the matrix clause, and so is said to be EMBEDDED (= contained) within it. We can indicate this embedding as in (25). The innermost square brackets show the subordinate clause, and the outermost brackets show the matrix clause; as you can see, the subordinate clause is entirely contained within the matrix clause.

- (25) [My friend claimed [that Ceri liked chips]].
-

In (26), we see another role that subordinate clauses (underlined) can fulfil:

- (26) a. That Chris liked Lee so much really **surprises** me.
 b. For Mel to act so recklessly **shocked** everyone.

These subordinate clauses are known as CLAUSAL SUBJECTS (or SENTENTIAL SUBJECTS), because they are clauses, but also fulfil the requirement for the matrix verbs (in bold) to have a subject. You can see that they are in the subject position by replacing them with an ordinary noun phrase subject, the phrase in square brackets here:

- (27) a. [This] really surprises me.
 b. [Mel's behaviour] shocked everyone.

Again, the clausal subjects are embedded within the matrix clause, but this time, of course, they are in the subject position:

- (28) [[That Chris liked Lee so much] really pleases me].
 ← subordinate clause →
 ← matrix (or main) clause →

Like the subordinate clauses in (23), these clausal subjects may be either finite (26a) or non-finite (26b). Both of them are also introduced by a complementizer: *that* introduces the finite clause, and *for* the non-finite clause. It's not too surprising that both of these complementizers are obligatory here, because they signal the start of a special kind of subject: an entire subordinate clause. For instance, having complementizer *that* at the start of the finite clausal subject prevents the hearer from assuming incorrectly that the noun phrase *Chris* is just the subject of the matrix clause: **Chris liked Lee so much really pleases me*.

Subordination is not generally restricted to a depth of just one embedded clause. In fact, in most languages (though perhaps not all), complex sentences contain a potentially infinite number of subordinate clauses. For example, (23c) could be extended as *They want to know whether we'd expect to leave before breakfast* or *They want to know whether she thought we'd expect to leave before breakfast* (and so on). Such examples of RECURSION are typical, though recursion may be fairly restricted in some languages. Each subordinate clause is dependent on the clause above it, and contained, or embedded, within the clause 'upstairs'. Each verb selects the following dependent clause as its complement. This means that we have structures of the following kind, where each clause is nested inside the clause directly above it:

- (29) They hope/want [to leave before breakfast].
 (30) They hope [we'd expect [to leave before breakfast]].
 (31) They want [to know [whether we'd expect [to leave before breakfast]]].

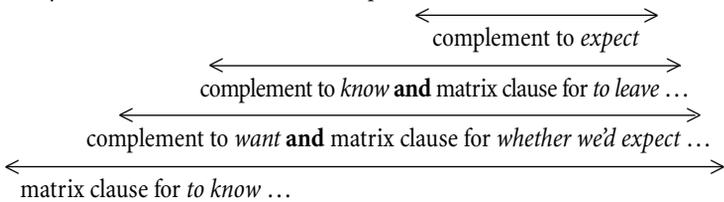
The brackets show the start and end of each clause. So in (30), for example, the *expect* clause doesn't end after *expect* – it can't, because *expect* absolutely requires the presence of the dependent clause *to leave before breakfast*. Instead, the *expect* clause ends after *breakfast*, at which point it is complete. In (30) and (31), the *leave* clause is dependent on the *expect* clause – the verb *expect* selects the non-finite subordinate

clause ‘downstairs’. And the *expect* clause in turn is dependent on the clause above, and again, is selected by the verb in the clause above it (*hope, know*). So the highest clause in (31) (the *want* clause) in fact contains all the other clauses within it:

(32) [They want [to know [whether we’d expect [to leave before breakfast]]]].

What we find, in other words, is not a linear sequence of clauses strung out one after the other, [...] [...], but rather, a HIERARCHICAL structure of clauses embedded within clauses: [... [...]]. The *know, expect* and *leave* clauses here are all complement clauses, since they are required by the verb in the ‘upstairs’ clause. But the *want, know* and *expect* clauses are also all matrix clauses, since they each select as a complement the clause ‘downstairs’. So a clause can be simultaneously a matrix (from the Latin meaning ‘mother’) clause and a complement clause:

(33) [They want [to know [whether we’d expect [to leave before breakfast]]]].



At the moment, we are only talking about English, where complement clauses follow the verb that selects them. Later on we will see that in languages such as Japanese, complement clauses precede the verb that selects them.

The subordinate clauses discussed so far are all complements, because they are essential; they can’t simply be omitted without loss of grammaticality. Clausal subjects, as in (26), are also arguments of the verb, just as much as the embedded clauses in examples like (23). For this reason, clausal subjects are traditionally termed ‘subject complement clauses’.

However, not all subordinate clauses are complements – required arguments of a matrix verb. As we’ll see next, some are optional.

3.2.2 Adjunct or adverbial clauses

Some embedded clauses are not selected by any verb, and instead are just optional modifiers:

- (34) a. Mel will be there [when she’s good and ready].
 b. [If you’re leaving early], please get up quietly.
 c. [Kim having left early], we drank her beer.

The clauses shown in brackets are all ADJUNCTS, to use a term introduced in Chapter 2; this means that they are not obligatory. You can see this for yourself by removing them from (34); all the remaining sentences are fully grammatical. In traditional grammar, these optional subordinate clauses are known as ADVERBIAL CLAUSES. They add a very wide range of additional meanings, including information about time,

Here are the main verbs, in bold:

- (37) a. When Kim **got** on the train, someone **said** she'd left her rucksack in the middle of the platform on a trolley.
 b. Unless we **want to arrive** late, we really **need to be leaving** now.
 c. To **get** to class on time, you should **set** your alarm every Wednesday for about 6.15 in the morning.
 d. To arrive on time **feels** brilliant.
 e. I **promise** to cook the meal while you **sort** the groceries.

And all the complement clauses (including the clausal subject in (37d)) are underlined. Note that the whole of the complex sentence *unless we want to arrive late* is an adjunct to the *need* clause. However, that adjunct itself contains two clauses: a matrix clause with the main verb *want*, and its complement, the subordinate *arrive* clause. The adjuncts are:

- (38) when Kim got on the train
 unless we want to arrive late
 to get to class on time
 while you sort the groceries

At this stage, you will have noticed that I am beginning to use a great deal of the terminology which was introduced and defined in earlier sections and chapters. If you are finding it hard to keep things straight in your mind, you will need to do some revision before reading further, since I will be using the technical terms more often from now on without any reminder of their meaning. In any case, I recommend re-reading this section (3.2) up to this point before moving on.

3.2.4 Root and subordinate clauses: Some distinctions

In a complex sentence, the highest clause in the hierarchical structure is known as the ROOT CLAUSE. It is not uncommon, cross-linguistically, for root clauses to display some special properties that are not shared by embedded clauses. For instance, embedded clauses in English may be finite or non-finite, but the root clause is *always* finite; in other words, it must contain a finite verb. Independent clauses share these properties too, and for that reason may also be termed root clauses.

Another way in which root clauses often differ from subordinate clauses concerns word order. A root clause may have a word order that does not occur in embedded clauses, or vice versa. The Germanic languages are well known for this phenomenon, and later we will see some indications of it in English (which is a Germanic language). The illustration here, though, is from a Celtic language, Breton. Back in Chapter 1, we saw that Welsh, another Celtic language, has verb-initial word order: in other words, the finite verb comes first in the clause. We might expect that the closely related language Breton would be verb-initial too, but it appears from the ungrammaticality of (39) that this is not the case:

- (39) *Lenn ar wazed al levr. (Breton)
 read.PRES the men the book
 ('The men read the book.')

Rather than (39), one grammatical version of this sentence would be (40), where the subject is initial in the clause:

- (40) Ar wazed a lenn al levr.
 the men PRT read.PRES the book
 'The men read the book.' (*Literally, 'It's the men that read the book.'*)

It's also possible for the object of the verb to be initial in the clause:

- (41) Al levr a lenn ar wazed.
 the book PRT read.PRES the men
 'The men read the book.' (*Literally, 'It's the book that the men read.'*)

However, if we make the verb-initial sentence in (39) into a subordinate clause (introduced by a small particle, *e*) then it's perfectly grammatical:

- (42) Int a gav dezho [e lenn ar wazed al levr].
 they PRT think.PRES to.3PL PRT read.PRES the men the book
 'They think that the men read the book.'

So what is going on here? In fact, finite verbs are indeed initial in Breton – but generally, that order is not allowed in root clauses, only in embedded clauses. What happens in Breton (as in German) is that some element must precede the finite verb in a root clause; as we've seen, this could be the subject, the object, or indeed various other elements, such as an adjunct.

In some languages, of which English is generally said to be an example, root clauses do not have complementizers. This is definitely not a universal property; for instance, Swedish is a Germanic language quite closely related to English, and as (43) shows, it does have complementizers in root clauses. One of two complementizers can be chosen here (in bold):

- (43) **Om/att** jag gillar blodpudding. (Swedish)
 if/that I like black.pudding
 'You bet I like black pudding!'

It's true that we don't get English sentences such as **That my friend likes chips* or **Whether/if it will rain today*. On the whole, then, we can agree that English typically lacks root clause complementizers. However, a common usage in some varieties of English has *so* in root clauses:

- (44) Interviewer: Tell us about the new website, then.
 Interviewee: **So** this site has been up and running for around a month.

This is clearly not the *so* of purpose adverbial clauses, as in *I'll stop talking so you can concentrate* – in fact, its only function seems to be to delineate the start of the clause. I consider, then, that *so* here is a root clause complementizer.

English root clauses have two other properties that will help you to distinguish them from subordinate clauses.

- **Only root clauses in English have subject/auxiliary inversion.**

The usual way of asking YES/NO QUESTIONS in English (that is, questions expecting the answer *yes* or *no*) involves what is known as SUBJECT/AUXILIARY INVERSION. The subject of a root clause undergoes inversion (= switching of position) with a finite auxiliary. In a simple sentence, the word order in a statement is *Kim didn't like chips*, while the word order in a question is *Didn't Kim like chips?* *Kim* is the subject, and *didn't* the finite auxiliary, which moves to the left of the subject. Some more instances of this are: *You can speak Italian fluently* (statement) and *Can you speak Italian fluently?* (question); *Lee has been sleeping badly* and *Has Lee been sleeping badly?*

Now let's look at subject/auxiliary inversion in the root clause of some complex sentences:

- (45) a. If you're leaving early, **should you** make sure your alarm works?
 b. **Can Mel** persuade Kim to cook a nice meal?

The inversion test will tell you whether a clause is a root clause or an embedded clause. Obviously, this test can be used only in finite clauses, since only *finite* auxiliaries can be inverted in this way. So we can't apply the inversion test in non-finite clauses: **Having Kim left early, we drank her beer*. But we already know that all non-finite clauses are subordinate clauses anyway. Let's try the test in a complex sentence with a finite subordinate clause: *Your friend claimed that Ceri liked chips*. There are two finite clauses here: first, the *claimed* clause:

- (46) **Did your friend** claim that Ceri liked chips?

This works (with *do*-support, as there's no other auxiliary), so we can be sure that *claim* is the verb of the root clause. But we can't do this in the embedded *like* clause:

- (47) *Your friend claimed that **did Ceri** like chips?

This is ungrammatical, so confirming what we already knew: a clause introduced by complementizer *that* must be a subordinate clause.

Here we should add the proviso that inversion is in fact allowed in embedded clauses that are (or act like) a quotation of someone's words. So we find sentences such as *My friend said, 'Did Lee think that Ceri liked chips?'* and *Ceri asked, could they be a little quieter?*

- **Only root clauses in English can have tag questions.**

Tag questions are usually 'tagged onto' the end of the entire sentence, and they have a pronoun as their subject which matches the subject of the root clause. Since they

also use subject/auxiliary inversion, tag questions too are found only in root clauses or when quoting speech. Example (48) illustrates, with the finite verb/auxiliary of the root clause and its associated tag in bold:

- (48) a. We **persuaded** Kim to cook a nice meal, **didn't we?**
 b. For you to act so hastily **was** unexpected, **wasn't it?**
 c. If you're leaving early, you **should** ensure that your alarm works, **shouldn't you?**
 d. Kim having left early, we **drank** her beer, **didn't we?**

As usual, if there's no finite auxiliary in the root clause, then *do*-support is required, as in (48a) and (d). Note that when the root clause is affirmative, the tag is negative, and vice versa: *She hasn't gone yet, has she?*

Tag questions can't be formed from embedded clauses, even if they're finite, as (49) shows – these sound very odd:

- (49) a. *I wondered whether Lee had gone, hadn't he?
 b. *If you're leaving early, you should ensure that your alarm works, aren't you/doesn't it?

In (49b), there are two subordinate clauses: the *leaving* clause (an adjunct) and the *works* clause (a complement). Forming a tag associated with either of these is impossible. We can only form a grammatical tag question from the root clause, as (48c) shows

There are a few exceptions, so some caution is needed: if the root clause verb is a verb like *think* or *say*, we can, in fact, get embedded tag questions, such as *I think we're leaving soon, aren't we?*

3.2.5 Some cross-linguistic variation in subordination

So far, we have only seen examples of subordinate clauses that *follow* the main verb that selects them, as is the case in English and in European languages generally. The next two examples both have a complement clause which *precedes* the verb that selects it. We will be looking at word orders like this in more detail in Chapter 4. For now, you need to understand that the matrix verbs meaning 'know' and 'want' select an embedded clause, just as in English, but that this clause (bracketed) precedes the verb that selects it:

- (50) ?ah [ce k'ew ew tum-tah] hatiskhi? (Wappo)
 1SG that man fish buy-PAST know
 'I know that man bought fish.'

- (51) ?ah [ce k'ew ew tum-uhk] hak'se?
 1SG that man fish buy-INFIN want
 'I want that man to buy fish.'

In (50), the subordinate clause is finite, as we can tell from the past tense marker on the verb, and in (51), the subordinate clause is infinitival.

In English, verbs such as *try* and *want* select subordinate INFINITIVAL clauses, as in *Kim tries/wants/hopes [to leave before breakfast]*, where the infinitival clause (containing the infinitive *leave*) is bracketed. Rather than having an overt (pronounced) subject, such clauses often have an *understood* subject, referring back to the subject of the main clause – we understand that the person leaving will be *Kim*. In English, many matrix verbs can select either an infinitival clause, or alternatively a finite clause, as their complement. So we can also have *Kim hoped [that she could leave before breakfast]*. But not all languages have infinitives. So what do the embedded clauses selected by the equivalent verbs look like in such languages? The examples in (52) and (53) are from modern Greek, and the embedded clauses are in brackets (SJTIV is a subjunctive marker, used to mark some event that hasn't actually happened yet).

(52) o Sokratis theli [i Afrodhiti na ton filisi] (Greek)
 the Socrates want.3SG the Aphrodite SJTV him kiss.3SG
 'Socrates wants Aphrodite to kiss him.'

(53) i Maria prospathise [na diavasi ena vivlio]
 the Mary tried.3SG SJTV read.3SG one book
 'Mary tried to read a book.'

A literal translation of (52) would be something like 'Socrates wants that Aphrodite kisses him' and of (53), 'Mary tried she reads a book'. In other words, the embedded clauses are both *finite* in Greek: as in the matrix clauses, both verbs in the embedded clauses have a third person singular inflection.

3.2.6 Summary: Properties of subordinate clauses and root clauses

- Complement clauses and adjunct (or adverbial) clauses are both types of subordinate clause. A third major type of subordinate clause has not been discussed in this section: the relative clause. This is the construction underlined here: *I never like the food that they serve in the canteen*. Relative clauses are optional, so are in fact a type of adjunct. We will explore them in detail in Chapter 8.
- Complement clauses serve as arguments of the verb (or other lexical 'head') in the matrix clause. For that reason, they are typically obligatory.
- Adjunct clauses are not arguments, but optional modifying elements. These are traditionally termed adverbial clauses.
- Not all subordinate clauses would be possible as independent clauses. All non-finite clauses are impossible as 'stand-alone' clauses, in English and in many (though not all) languages.
- Both complement and adjunct clauses in English can be finite or non-finite. Some languages have more restrictions on the finiteness of subordinate clauses, though

many do not. Any clause that *only* has a non-finite verb, and no finite element at all, will generally be a subordinate clause of some kind.

- Both complement and adjunct clauses in English may begin with a complementizer. English root clauses typically do not, but root clause complementizers are common cross-linguistically.
- Root clauses often have special properties cross-linguistically. In English, they are identified by their ability to take subject/auxiliary inversion and tag questions. In some other languages, root clauses have a special word order that differs from the word order in subordinate clauses.

3.3 MAJOR CROSS-LINGUISTIC VARIATIONS

The majority of languages have complex sentences of some form, but not all languages share the type of complex sentences found in English. The kind of subordination used in familiar European languages is not universal, although it is also widespread outside Europe. But many languages have strategies which seem to avoid the type of complementation common to European languages. This section examines some of the main cross-linguistic variations in clause types.

3.3.1 The co-ordination strategy

The first alternative strategy is CO-ORDINATION. Compare the Kambera examples in (54) and (55) with their English translations. The gloss CONJ indicates a conjunction; see Section 3.1.6:

- (54) Ku-ita-ya **ba** na-laku la Umalulu. (Kambera)
 1SG.SU-see-3SG.OBJ CONJ 3SG.SU-go to Melolo
 ‘I saw him going to Melolo.’
- (55) Ku-rongu-kau **ba** u-ludu.
 1SG.SU-hear-2SG.OBJ CONJ 2SG.SU-sing
 ‘I heard you sing.’

Starting just for comparison with the English translations, the constructions *I saw him going*, *I heard you sing* are examples of COMPLEMENTATION: the verbs in each matrix clause (*see* and *hear*) select a non-finite subordinate clause which contains the verbs *going* and *sing*. As we have already seen, it’s common in English to have a finite verb in the matrix clause which selects some kind of non-finite subordinate clause – one way to tell that we have a subordinate clause in the English translations in (54) and (55) is the very fact that they are non-finite. (If you are not sure that *sing* really is non-finite here, note that the verb cannot take the *-s* inflection for present tense third person singular: **I heard him sings*.)

But the Kambera equivalents use co-ordination rather than subordination. Literally, the Kambera examples could be translated as ‘I saw him and he went to Melolo’ and ‘I

of *hated*, and in (56b), the subject of *surprised*. These nominal *-ing* constructions are traditionally known as GERUNDS in English.

Now compare this Kambera example, where the nominalized clause is in brackets:

- (57) Nda ku-mbuti-nya [na tàka-mu] (Kambera)
 NEG 1SG.SU-expect-3SG.OBJ the arrive-2SG
 'I did not expect you to arrive.'

Literally, this means '*I didn't expect it, your arrival*', which is definitely not very natural English; but the Kambera is perfectly natural. The verb meaning 'arrive' is clearly nominalized here because it occurs with a determiner, *na* 'the', which is a property of nouns.

Example (58) shows a similar example from a native American language, Comanche, with the nominalized clause again bracketed:

- (58) [u-kima-na] nii supanaʔi-ti (Comanche)
 his-come-NOMINALIZER I know-ASPECT
 'I know that he's come.'

Here, instead of a finite subordinate clause, as in the English translation, we find a nominalization: the verb *kima* is turned into a noun form with a nominalizing suffix *-na*, and then it takes a possessive marker *u-* 'his', rather like the examples in (56) had the possessive marker *-ʔ*.

Nominalizations of this kind are still examples of subordination, because the nominalized clause is dependent on a matrix verb (*Lee's*) *losing his licence* couldn't occur as an independent clause, for instance). The next section discusses a construction that doesn't involve subordination.

3.3.3 Serial verbs

As we have seen from English and other languages, the type of complementation familiar from European languages involves an embedded clause which is subordinate to a matrix clause. This strategy is widespread cross-linguistically, but not all languages make much use of subordination. A different but very common strategy, known as VERB SERIALIZATION, occurs widely in the world's languages, for instance in Chinese, in many African languages, and in many of the languages of New Guinea.

Example (59) illustrates a serial verb construction from Nupe (a language of Nigeria), showing two finite verbs simply following one after the other.

- (59) Musa bé lá èbi. (Nupe)
 Musa came took knife
 'Musa came to take the knife.'

English and other European languages allow only *one* finite verb in each clause – that is, a verb marked for such categories as tense and/or person and number (we don't

get **Musa comes takes the knife*). In English, each clause contains just one main verb. In serial verb constructions, though, two main verbs occur within a single clause. Both are finite. Both belong to a single predicate. In the English translation of (59), there's a matrix clause with a finite verb, *Musa came*, and an embedded clause with an infinitival verb, *to take the knife*, but in the Nupe serial construction the two verbs form a single predicate: there is no subordinate clause.

Let's look at the typical properties of serial verb constructions. First, it's very common that no elements at all are allowed to intervene between the two serial verbs, which is not too surprising if they are closely tied together in a single predicate. This is the case in (59), from Nupe, and it's also true of Bare, an extinct language formerly spoken in Brazil and Venezuela. In the Chinese example that follows, also, the direct object *men* 'door' does not intervene between the serial verbs *la-kai* 'pull open':

- (60) Ta la-kai le men. (Chinese)
 he pull-open PERF door
 'He pulled the door open.'

In some languages, though, if the first of the two serial verbs is transitive, an object noun phrase can occur between them, as in (61). Here, the object of the transitive verb *mú* 'took' (*iwé*, 'book') intervenes in this way between the serial verbs *mú* and *wá*, 'came':

- (61) ó mú iwé wá (Yoruba)
 he took book came
 'He brought the book.'

The same happens in (62), where there's a transitive verb *kpá* 'take', with an object *kíyžèè* 'knife', and this immediately follows the verb:

- (62) ù kpá kíyžèè mòng ówl (Vagala)
 he take knife cut meat
 'He cut the meat with a knife.'

A second property of serialization is that the meanings of the two serial verbs together often make up a single complex event. So in (61), the meaning could literally be seen as 'He took the book and came', which is more or less possible in English, but which instead we denote with *bring* – which means to get something and take it to your destination.

Third, the two finite verbs in a serialization must have the same subject. This is crucial to the claim that they are both part of a single clause. We see this in (59), with the subject *Musa*; it's also shown in the Yoruba example in (61), where there is only one subject, *ó* 'he', but it is shared by the two verbs; and it's shown again in the Vagala example in (62), where the subject *ù* 'he' is shared by the two verbs. Another way that this shared subject is sometimes expressed is shown in (63), from Bare: crucially,

the two verbs must both have the same bound pronominal prefixes showing person/number, here *nu-*, giving the meaning ‘I’:

- (63) **nu**-takasā **nu**-dúmaka (Bare)
 1SG-deceived 1SG-sleep
 ‘I pretended (that) I was asleep.’

Note that once again, the English translation uses a finite subordinate clause, (*that*) *I was asleep*.

Contrast (63) with an example of SUBORDINATION in Bare, (64) (some of the following Bare examples are slightly adapted). This is not a serial construction, but instead is very like the English, with an adjunct clause (in brackets) before the matrix clause:

- (64) [mientre-ke **nu**-nakúda-ka] **i**-mare-d’a kubati (Bare)
 while-SEQUENTIAL 1SG-go-SEQUENTIAL 3SG.M-steal-ASPECT fish
 ‘While I was coming in, he stole the fish.’

Despite the fact that the two verbs *nunakúdaka* and *imared’a* in (64) follow one after the other, we can tell that this isn’t a serial construction because each verb has a different subject. Again, this is shown not by independent pronouns as in the English (*I, he*) but by the two different bound pronominal prefixes, *nu-*, *i-*, on the two verbs, indicating the person and number (and gender) of the two different subjects: the ‘go’ verb has the 1SG subject marker (meaning ‘I’) and the ‘steal’ verb, the 3SG masculine subject marker (meaning ‘he’). The verbs are therefore in separate clauses.

A fourth property of serialization is that there is only one marker of negation for the whole serial verb construction. In (65), this is the negative marker *hena*:

- (65) **hena** **nihiwawaka** **nu-tšereka** nu-yaka-u abi (Bare)
 NEG 1SG.go 1SG-speak 1SG-parent-F with
 ‘I am not going to talk to my mother.’

The two serial verbs, *nihiwawaka* and *nutšereka*, both share the negative marker *hena*.

Compare (65) with (66), which is not a serial verb construction, but instead has two separate finite clauses, each with their own negative marker, *hena*:

- (66) **hena**-ka ini-hisa **hena** in-hiwawaka (Bare)
 NEG-DECLARATIVE 2PL-want NEG 2PL-go
 ‘If you do not want, do not go.’

Fifth, the serial verbs can’t be marked independently for such grammatical categories as tense, aspect or mood, but must share the same tense etc. This is either marked on each verb, or else occurs just once but is shared by both verbs. A good example is the Chinese perfect aspect marker *le*, seen in (60); this occurs only once

for the whole serial construction. Another such category is the Bare ‘sequential’ marker *-ka*; this occurs only on one verb in a serial construction:

- (67) nuni hena nu-kiate-d’áwaka nu-yuwahada-ka (Bare)
 I NEG 1SG-fear-ASPECT 1SG-walk-SEQUENTIAL
 ‘I’m not afraid of walking.’

(Note that we again have the same bound pronominal subject markers *nu-* on each verb in the serial construction, and the single shared negative marker, *hena*.) We can compare (67) to an example of subordination in Bare, where we find that each verb in the two subordinate clauses takes the sequential *-ka* marker. In (68), the ‘roll’ verb is in the root clause, and the two other verbs, meaning ‘see’ and ‘sleep’, are in two subordinate clauses; the English translation is just the same in this respect. *Both* the verbs in the embedded clauses take a *-ka* marker.

- (68) nu-khuruna hnumiye ibeuku nu-yada-ka sepultura tibuku
 1SG-roll 1SG.hammock when 1SG-see-SEQUENTIAL tomb over
 nu-duma-ka
 1SG-sleep-SEQUENTIAL
 ‘I rolled my hammock when I saw that I had slept over a tomb.’

Finally, if the serial verb construction seems exotic, note that something similar was common in sixteenth century English (the time of Shakespeare). An example would be *Come live with me and be my love*; constructions of this type have also survived especially in American English, as in *Let’s go eat!* for example.

3.3.4 Summary

Section 3.3 has shown that languages do not necessarily share the same syntactic strategies as the familiar European languages. Finite and non-finite subordination, where one clause is embedded inside another clause, is widely used in many language families, including non-European ones. But it’s important to realize that it’s not the only possible strategy. The two major alternative constructions are nominalization (a verb converted to a noun, so that the dependent clause takes on the properties of a noun phrase) and serialization, which does not involve any subordination, but instead has two finite verbs within the same predicate.

FURTHER READING

Hurford (1994) is helpful for further illustrations concerning both simple and complex sentences, auxiliaries and main verbs, matrix clauses and embedded clauses. Huddleston and Pullum (2002, 2005) provide comprehensive information about English clauses. On what are termed ‘complementation strategies’, the topic of Section 3.3, Dixon (1995) is good but advanced reading, which should be tackled only

after you've finished this book. Whaley (1997: Chapter 15) covers all types of complex clauses. See also T. Payne (1997, 2006).

EXERCISES

1. This exercise concerns a set of words that are possible candidates as English modal auxiliaries: *dare*, *need*, *ought (to)*, *used (to)* (the last two in their auxiliary uses are often represented by linguists as *oughta*, *useta*). All of these display both auxiliary and main verb syntactic properties. A set of properties is taken to be diagnostic of auxiliaries in English. Four central ones, some of which we've already met in Chapter 3, are the NICE properties:
 - a. Negation – an auxiliary, but not a main verb, can be directly negated by *not*:
 - (i) We *do / should / may* not talk about that. / *We talk not about that.
 - b. Inversion – an auxiliary, but not a main verb, can invert with the subject:
 - (ii) *Can/will/did* Lill bake a cake for me? / *Baked Lill a cake for me?
 - c. Code – an auxiliary, but not a main verb, can be used with an ellipsis (omission):
 - (iii) Lill said she'd bake a cake, and she *did/will/might* too.
*Lill said she'd bake a cake, and she baked too.
 - (iv) Kim should get his work finished on time, and so *should* Lee.
 - d. Emphasis – an auxiliary, but not a main verb, can bear heavy stress for emphasis:
 - (v) You say you might not go, but you *might*.
You don't think he read it, but he *did*.
*You don't think he read it, but he *read*.

Some additional properties are shared by standard English modal auxiliaries. They don't take the third person singular present tense *-s* suffix (**She mays leave* or **Kim wills arrive*), while main verbs do. And the auxiliaries in the standard set are also unlike main verbs in that they don't have an infinitive (**She wants to may*) and don't have an imperative (**May leave!* vs. *Leave!*).

Task: Using the diagnostics just presented, work out (i) the ways in which *dare*, *need*, *oughta* and *useta* behave like modal auxiliaries, and (ii) the ways in which they behave like main verbs. There is no single 'right' answer, in part because different dialects of English have different usages of these words. Below I suggest some data that should get you started, but you'll need to provide some additional data of your own. Make sure you list such data in your response. Organize the answer clearly. **NB!** No grammaticality judgements are provided here, since mine may well differ from yours. Decide for yourself which are grammatical in your dialect and which are not.

- (1) I daren't leave./I don't dare leave./He dares leave./He dare(s) not leave. /He daresn't leave.

- (2) Dare you (to) pick up that spider? / Do you dare pick up that spider?
- (3) I might not dare to pick it up. / Well, Lill dared to pick it up.
- (4) Kim used not to / usen't to take any exercise. / Kim didn't use to take any exercise.
- (5) Used Kim to take any exercise? / Did Kim used to take any exercise?
- (6) She ought to stop eating so much chocolate. / She oughtn't to eat that. / She didn't ought to eat any sweets at all.
- (7) Ought/oughtn't she to stop eating chocolate? / Did she ought to stop eating chocolate?
- (8) I needn't go. / He need not / needn't go. / He doesn't need to go.
- (9) Do you need to leave? / Need you leave?
- (10) Kim needs a holiday.

2. Section 3.2 presented some tests for distinguishing root clauses in English from subordinate clauses. One construction discussed there shows up again in the data in (1) through (7).

Task: (i) What construction is it? (ii) What seems to trigger the appearance of this construction in these examples? (iii) Can you decide what property the examples have in common?

- (1) Not for any money would Lill pick up a spider. Neither will I, actually.
- (2) Rarely have we seen such snow before.
- (3) Never again must those students take the last train to Durham.
- (4) Under no circumstances should you press the red button.
- (5) Seldom can you find a better bargain than at Den's Dealership.
- (6) Not till after the weekend might those who are on strike return to their desks.
- (7) Only after 22.00 will there be another train.

- (iv) What issues are suggested by these additional examples?

- (8) She said that under no circumstances could she learn Irish.
- (9) I knew that not even on Sundays / only on Sundays could my daughter lie in bed till midday.

3. This exercise asks you to consider the possible positions and functions of complementizers cross-linguistically.

Task: Examine the data in (1) to (13) and work out: (i) what kinds of function the complementizers (in bold) appear to have in these examples; and (ii) what appear to be the possible positions that complementizers can take in the clause, cross-linguistically? Discuss each data set separately where necessary.

Hints

- Regarding question (i), the function of a complementizer is often simply to introduce a clause; this is the case for complementizer *that* in English. However, some of the complementizers here do more work than this, i.e. they bear some additional meaning.
- Note that the terms *NOM* (nominative) and *ACC* (accusative) in the Japanese data set are used to case-mark the subject (*NOM*) and the object (*ACC*), as outlined in Chapter 2.
- As was the practice in the text of Chapter 3, square brackets indicate the start and end of an embedded clause in these data.

A. Yaqui (Noonan 1985)

- (1) Tuisi tu?i [ke hu hamut bwika-kai]
 very good COMP the woman sing-COMP
 ‘It’s very good that the woman sings.’
- (2) Tuisi tu?i [ke hu hamut bwika]
 very good COMP the woman sing
 ‘It’s very good that the woman sings.’
- (3) Tuisi tu?i [hu hamut bwika-kai]
 very good the woman sing-COMP
 ‘It’s very good that the woman sings.’
- (4) *Tuisi tu?i [hu hamut bwika]
 very good the woman sing
 (‘It’s very good that the woman sings.’)

B. Japanese (Tsujimura 1996; Kuno 1978)

- (5) a. Hanako-ga susi-o tukurimasita
 Hanako-NOM sushi-ACC made
 ‘Hanako made sushi.’
- b. Hanako-ga susi-o tukurimasita ka
 Hanako-NOM sushi-ACC made COMP
 ‘Did Hanako make sushi?’
- (6) Taroo-ga [Hanako-ga kuru to] itta.
 Taroo-NOM Hanako-NOM come COMP said
 ‘Taroo said that Hanako was coming.’
- (7) Taroo-ga [Hanako-ga oisii susi-o tukutta to] itta
 Taroo-NOM Hanako-NOM delicious sushi-ACC made COMP said
 ‘Taro said that Hanako made delicious sushi.’

C. Irish (McCloskey 1979; Ó Siadhail 1989)

- (8) Deir sé [go dtuigeann sé an scéal].
 say.PRES he COMP understand.PRES he the story
 ‘He says he understands the story.’

- (9) Deir sé [**nach** dtuigeann sé an scéal].
 say.PRES he COMP understand.PRES he the story
 'He says he doesn't understand the story.'
- (10) Deir sé [**gur** thuig sé an scéal].
 say.PRES he COMP understand.PAST he the story
 'He says he understand the story.'
- (11) Deir sé [**nár** thuig sé an scéal].
 say.PRES he COMP understand.PAST he the story
 'He says he didn't understand the story.'
- (12) Ní dheachaidh mé ann.
 COMP go.PAST I there
 'I didn't go there.'
- (13) Chuaigh mé ann.
 go.PAST I there
 'I went there.'

4. In Section 3.1.4 we looked at some ways of forming clausal negation cross-linguistically. The Evenki example in (12) has a special negative auxiliary, while English has an independent negative word, *not* (though this is often optionally attached to auxiliaries, giving forms like *can't* and *shouldn't*). This gives us two of the three major ways of expressing grammatical categories listed in Section 3.1.4: a third strategy would be expressing negation via an inflection on the verb itself. All three strategies are exemplified in the following data (1) to (9).

Task: Work out which strategy – negative auxiliary, negative particle, or verbal inflection – is used for the negation in each example that is negative. (There are some positive examples for comparison.) Make sure you cite clear evidence for each answer. If any cases cannot be decided straightforwardly, or display more than one strategy, explain why. Finally, point out any *relevant* grammatical features or changes in the negative examples which don't occur in corresponding positive examples.

Hint

A negative auxiliary can be distinguished from a negative particle like *not* because an auxiliary expresses some of the grammatical categories associated with verbs generally, such as tense, person and/or number. An independent negative particle is just invariable, so will not be marked for any of these morphosyntactic categories. You can see this by comparing (11) and (12) in Chapter 3, and re-reading the discussion of these examples.

- (1) a. Si ə-tci-si bū-ra (Orok)
 you NEG-PAST-2SG give-PARTICIPLE
 'You didn't give.'

- b. Si ə-tcil bū-rə-si
 you NEG-PAST give-PARTICIPLE-2SG
 'You didn't give.'
 (J. Payne 1985a)
- (2) Anghofia / anghofiwch y caws!
 forget.IMPER.SG / forget.IMPER.PL the cheese
 'Forget the cheese!'
 (Welsh)
- (3) Paid / Peidiwch ag anghofio 'r caws.
 NEG.IMPER.SG / NEG.IMPER.PL with forget.INFIN the cheese
 'Don't forget the cheese.'
- (3) a. Gwall ampart eo va breur.
 very competent be.PRES.3SG my brother
 'My brother is very competent.'
 b. Gwall ampart n' eo ket va breur.
 very competent NEG be.PRES.3SG NEG my brother
 'My brother isn't very competent.'
 (Press 1986)
- (4) a. ama-wa-t b. ta-ka-wa-t
 1SG(SU)-go-PERF NEG-1SG(SU)-go-PERF
 'I went.' 'I didn't go.'
- (5) a. kpa-n amayak b. ama kpa-n tampan
 big-1SG be.1SG 1SG big-1SG be.NEG
 'I'm big.' 'I'm not big.'
 (Foley 1991)
- (6) a. apenim un an kasye
 father TOPIC NEG go
 'Father is not going.'
 b. apenim un ka-ci an-husyey
 father TOPIC go-NOMINALIZER NEG-do
 'Father is not going.'
 (adapted from Sohn 1999)
- (7) a. xola-xa-si b. xola:-si-si
 read-PAST-2SG read-NEG.PAST-2SG
 'You were reading.' 'You weren't reading.'
 (T. Payne 1997)
- (8) a. ́ tē ḱ
 he buy rice
 'He bought rice'
 b. ́ sé ḱ tē
 he NEG buy rice
 'He didn't buy rice.'
 (J. Payne 1985a)

- (9) a. Mae Aled yn darllen y llyfr. (Welsh)
 be.PRES.3SG Aled PROG read.INFIN the book
 'Aled is reading the book.'
- b. Dydy Aled ddim yn darllen y llyfr.
 NEG.be.PRES.3SG Aled NEG PROG read.INFIN the book
 'Aled isn't reading the book.'

5. Examine the English sentences in (1) to (8).

Task: (i) Mark in bold type the main verb (i.e. the lexical verb) in each clause. This will help you find where the clauses are. Now, (ii), decide which is the root clause in each example, and underline its main verb. Also, (iii), give at least one piece of evidence for the root status of each of these root clauses you've picked out, using the tests established in Section 3.2. Next (iv), underline each of the subordinate clauses. Give at least one piece of evidence that each clause you've chosen really is a subordinate clause, using the criteria established in Section 3.2. Then, (v), list the adjunct clauses and (vi) the complement clauses, giving some evidence for your decision in each case. Finally, (vii), say which matrix verb each of the complement clauses is a complement to.

Hint

Here is an example that I've done for you: *Kim has sometimes wondered how to cope with unexpected visitors.*

- (i) Kim has sometimes **wondered** how to **cope** with unexpected visitors.
 (ii) Kim has sometimes **wondered** how to cope with unexpected visitors.
 (iii) The *wondered* clause is the root clause here because (a) it can take a tag question: *Kim has sometimes wondered how to cope with unexpected visitors, hasn't he?* and (b) it can take subject/auxiliary inversion: *Has Kim sometimes wondered how to cope with unexpected visitors?* These tests are not relevant for the *how to cope ...* clause as this is non-finite.
 (iv) There is just one subordinate clause, *how to cope with unexpected visitors*. This can only be a subordinate clause in English, because it's non-finite.
 (v-vii) This is a complement clause; the complement to *wondered*. *Wonder* obligatorily selects an embedded interrogative clause (i.e. a clause introduced by a question word such as *how, why, whether* and so on).
- (1) Lee knows it's illegal but she still photocopied the entire book.
 (2) Kim can't accept that the earth was only formed 5,000 years ago.
 (3) That student with the unbelievably bright red sweatshirt over in the corner often stays in the gym till around 11 pm.

- (4) Since you write so well, we hope to hire you to work on the student newspaper.
- (5) That you could spend so much time with Kim frankly amazes Lee.
- (6) Yesterday evening, both the manager of the bar and the person behind the desk expected to give that part-time job to the guy with the faded jeans.
- (7) Before the sun rose, we'd already run about three miles.
- (8) Meet me in my office for a brief chat after class has finished.
6. The examples in (1) to (6) are from a Melanesian language, Tinrin, first seen in Chapter 1, and are taken from Osumi (1995). They all show verb serialization, so you will need to re-read Section 3.3.3 before starting.

Task: (i) Work out what typical properties of verb serialization these examples show. Compare the grammatical and ungrammatical sentences where shown. Be as explicit as possible in your answer, and use the correct grammatical terminology. Then (ii) decide how and why the serialization in (1) to (4) is different from the serialization shown in (5) and (6). The forms *ri*, *rri*, *nrî*, *u* and *nrâ* are all pronouns.

- (1) a. u nrorri gadhu peci ei toni
 1SG give waste letter to Tony
 'I wasted a letter by giving it to Tony.'
- b. *u nrorri peci ei toni gadhu
 1SG give letter to Tony waste
 ('I wasted a letter by giving it to Tony.')
- (2) a. ri ve fi toni
 1PL.INC take go Tony
 'We took Tony away.'
- b. *ri ve toni fi
 1PL.INC take Tony go
 ('We took Tony away.')
- (3) a. rri ve mê arrò
 3PL take come water
 'They brought water.'
- b. *rri ve arrò mê
 3PL take water come
 ('They brought water.')
- (4) a. nrâ nyôrrò mê ò
 3SG cook come pot
 'She cooked and brought the pot dish.'
- b. *nrâ nyôrrò ò mê
 3SG cook pot come
 ('She cooked and brought the pot dish.')

- (5) ri ve nrî fi
1PL.INC take 3SG go
'We take it/him away.'
- (6) rri ve nrî mê
3PL take 3SG come
'They bring it.'

Heads and their dependents

Section 4.1 looks at head words and their dependents within a phrase. Section 4.2 looks at the positioning of heads within their phrase, examining a major typological division into head-initial and head-final languages. Section 4.3 examines the properties of head-marking and dependent-marking languages, another important typological distinction between languages.

4.1 HEADS AND THEIR DEPENDENTS

This section examines the concept of ‘head of a phrase’, and then moves on to discuss what types of phrases are selected by each class of head as obligatory COMPLEMENTS, and what types of phrases accompany each head as optional modifiers.

4.1.1 What is a head?

In any phrase, we distinguish between the word that is the overall HEAD of the phrase, and other words which are DEPENDENTS to that head. The heads of the phrases in (1) – in bold – are bracketed, and their word class indicated with a subscript: ‘N’ for noun, ‘V’ for verb, ‘A’ for adjective and ‘P’ for preposition. All the other words or phrases are dependents to those heads:

- (1)
- a. very bright [_N **sunflowers**]
 - b. [_V **overflowed**] quite quickly
 - c. very [_A **bright**]
 - d. quite [_{Adv} **quickly**]
 - e. [_P **inside**] the house

The head is the most important word in the phrase, first because it bears the crucial semantic information: it determines the meaning of the entire phrase. So the phrase *very bright sunflowers* is ‘about’ sunflowers; *overflowed quite quickly* is about something overflowing, and so on. To take other examples, *a brass statue* means *a kind of* statue, not a kind of brass, so the head is *statue*; *vegetable stew* is *a kind of* stew, not a kind of vegetable, so the head is *stew*. The word class of the head therefore determines the word class of the entire phrase. Since *very bright sunflowers* in (1a) is headed by a noun, it is a Noun Phrase (NP); *overflowed quite quickly* in (1b) is headed by a verb, so is a Verb Phrase (VP); *very bright* in (1c) is an Adjective Phrase (AP);

quite quickly in (1d) is an Adverb Phrase (AdvP); and in (1e), *inside the house* is a Preposition Phrase (PP) headed by the preposition *inside*.

Second, in all the examples in (1) the head is the only word that has the same DISTRIBUTION as the entire phrase. Wherever the whole phrase can occur, it's possible to substitute just the head. For instance, we could say either *Kim liked very bright sunflowers*, or just *Kim liked sunflowers*; we could say *Go inside the house* or just *Go inside*. We can say *The sunflowers were bright* but not **The sunflowers were very* – therefore, *bright* rather than *very* must be the head of the AP.

It follows that the head can't normally be omitted (setting aside contexts where a head has just been mentioned, and is then omitted, as in *Are you angry?* answered by *Very!*). So the third property of heads is that they are the one obligatory item in the phrase.

There are many contexts, however, in which the dependents to a head can't be omitted either. For instance, in the Verb Phrase *released the hostages*, there's an obligatory dependent Noun Phrase, *the hostages*: we can't just say **The soldiers released*. And the Preposition Phrase *beside the wood* has an obligatory NP too, *the wood*; we don't get **She lives beside*. The reason these dependents can't be omitted is that the heads in each phrase require them to be there: we say that the heads select certain dependents as their COMPLEMENT. Two familiar instances are illustrated in this paragraph: a transitive verb like *release* or *enjoy* requires an object NP, and so does a transitive preposition like *beside* or *into*. The fourth property of heads, then, is that they may select an obligatory dependent, a phrase of a particular class (such as NP) and with specific semantic properties: we can say *She lives beside the wood*, but not **She lives beside the speculation*.

In order to have a 'phrase' of some kind, we minimally require the presence of a head; the phrase may additionally contain some (optional or obligatory) dependents. A Verb Phrase, for instance, must contain a verb and often contains other words too. Knowing this, we can capture certain GENERALIZATIONS (= the simplest and most accurate statement of the facts) about the structure of sentences. For example:

- The subject of a clause is a phrase of one word or more which is headed by a noun (so it's an NP).
- The PREDICATE of a clause (see Section 2.2.1 and Section 3.1.1) is normally a VP; this phrase may contain just a head verb such as *overflowed*, giving us sentences like *The bath overflowed*, or else the VP can contain dependents, as it does in the sentence *The bath overflowed quite quickly*.

4.1.2 The influence of heads on their dependents

Heads play a crucial role in determining certain properties of their dependents. This section examines three kinds of DEPENDENCIES involving a relationship between a head and its dependent(s).

First, in all languages, heads select dependents of a particular WORD CLASS: only dependents of a certain category can occur with each kind of head. For example, in English, a head noun can be modified by an adjective such as *bright* as in (1a), but a

Third, in many languages certain heads require their Noun Phrase dependents to occur in a particular grammatical *CASE* (see Section 6.3 for more details). Case is a property of NPs which indicates their grammatical function in a phrase or a clause: in languages that have case, NPs are marked in different ways depending on what function they fulfil. Specifically, the NP dependents of verbs and prepositions are often required to occur in a special form (see Section 2.3.2 for discussion of English pronouns): the verb or preposition is said to *GOVERN* the case of its dependent. For instance, a transitive verb has two arguments, therefore two dependent NPs: the subject and the object. These two NPs fulfil a different function from each other, and in many languages, the subject and the object also differ in form from each other: they are marked with different cases. So in the Japanese example in (4), the subject and object are marked in distinct ways, showing their different functions: the case markers are affixes on the nouns in Japanese. The NP which is the subject of the verb is in the *NOMINATIVE* case, and the object NP is in the *ACCUSATIVE* case. Nominative can generally be considered ‘the case that subjects have’ and accusative, ‘the case that objects have’.

- (4) Kodomo-ga hon-o yon-da. (Japanese)
 child-NOM book-ACC read-PAST
 ‘The child read the book.’

In this section we have seen various kinds of *DEPENDENCY*: a relationship contracted between elements in a phrase or a sentence. These dependencies are (a) the selection of a specific type of argument by a head; (b) agreement: the copying of features from a head to its dependents; and (c) government by a head.

4.1.3 Summary: The properties of heads

To summarize, the main points made about heads so far in this section are:

- The head bears the central semantic information in the phrase.
- The word class of the head determines the word class of the entire phrase.
- Heads are normally obligatory, while other material in a phrase may be optional.
- Heads select dependent phrases of a particular word class; these phrases are sometimes obligatory, and are known as *COMPLEMENTS*.
- Heads often require their dependents to agree with some or all of the grammatical features of the head, such as gender or number.
- Heads may require their dependent NPs to occur in a particular grammatical case. This is one form of a relationship traditionally known as *GOVERNMENT*: a head is said to govern the case of its dependent.

4.1.4 More about dependents: Adjuncts and complements

The dependents are all the remaining words in a phrase other than the head. Traditionally, dependents are classified into two main types: *ADJUNCTS* and *COMPLEMENTS*.

Adjuncts are always optional, whereas complements are frequently obligatory. The difference between them is that a complement is a phrase which is *selected* by the head, and therefore has an especially close relationship with the head; adjuncts, on the other hand, provide optional, extra information, and don't have a particularly close relationship with the head. Let's first consider some adjuncts. In (5), the heads are again bracketed, and the phrases which are the adjuncts are now in bold:

- (5) a. **very bright** [_N sunflowers]
 b. [_V overflowed] **quite quickly**
 c. [_V talks] **loudly**
 d. [_V sings] **in the bath**

As adjuncts, these phrases in (5) are optional. The adjuncts provide additional information about such things as appearance, location or the manner in which something was done. Adjective Phrases such as *very bright* and Adverb Phrases such as *quite quickly* or *loudly* are typical adjuncts. Preposition Phrases (such as *in the bath*) are often adjuncts too. Evidence that the PP *in the bath* in (5d) is an adjunct comes from the fact that it can be replaced by any number of different PPs, using virtually any head preposition: *before breakfast*, *at the bus-stop*, *on the way to work*, *in the waiting room* and so on. The verb *sing*, then, can have as an optional modifier any PP that makes sense: it doesn't place any syntactic or semantic restrictions on what that PP looks like. Such a PP is a typical adjunct: its form is not constrained by the head verb. Note that *overflow*, *sing* and *talk* in (5) are all intransitive verbs – the presence of an adjunct doesn't affect the transitivity of a verb.

Example (6) shows some heads and their complements, again in bold:

- (6) a. [_V admires] **famous linguists**
 b. [_V wondered] **whether to leave**
 c. [_V resorted] **to the instruction manual**
 d. [_A fond] **of chips**
 e. [_P inside] **the house**

Recall that a verb or a preposition which is TRANSITIVE requires an object NP as its complement. *Admire* in (6a) is transitive: the direct object NP is the complement of a transitive verb. Some verbs are always transitive, such as *release* in *The soldiers released the hostages*: such verbs must have an NP as their complement. Other verbs may be either transitive or intransitive: so *sing*, for instance, can also be transitive, as in *Kim sings folk songs*. The preposition *inside* in (6e) is transitive: it has a complement NP. Like verbs, some prepositions are always transitive (*beside*, *into*) whilst others are sometimes transitive and sometimes not.

The head verbs in (6b) and (6c) aren't transitive, because they don't have objects, but they do nonetheless have complements: the clause selected by *wonder* is its complement, as is the PP selected by *resort*. Compare the PP that is an adjunct in (5d) with the complement PP *to the instruction manual* in (6c). The preposition in the adjunct PP could be almost any preposition (*in*, *on*, *over*, *above*, *beside*, etc.), but

Give is one of a number of verbs in English that have both a direct object NP (*the chips*) and what is sometimes termed an INDIRECT OBJECT (*to Lee*): in English, the indirect object really has no special properties, but is just a PP usually headed by *to* or *for*. As (10) shows, though, there's also an alternative construction with two NP complements. Other verbs that behave like *give* are *send*, *show*, *write* and *buy*. Often, such verbs have an alternative classification as transitive verbs, so we get both *I wrote a letter # to Kim* and *I wrote a letter*.

- Some verbs also take an NP and a PP complement, but don't have an alternation with an NP – NP complement of the kind shown in (10):

- (11) Kim [_{VP} put **the potatoes # into the pan**].
 Kim [_{VP} exchanged **her car # for a new bike**].
 *Kim put the pan the potatoes.
 *Kim exchanged a new bike her car.

- PREPOSITIONAL verbs take a PP complement, shown in bold in (12):

- (12) a. This cake [_{VP} consists **of fruit and nuts**].
 b. I [_{VP} applied **for a new job**].

As noted earlier, the PP complement is headed by a specific preposition, the choice of which is determined by the verb: with a dependent PP, this is the main test for complement status. So you can only *apply for* a job, and not **over* or **against* a job. Some more prepositional verbs are seen in *resort to NP*, *rely on NP*, *glance at NP*, *look after NP* and *long for NP*. Adjunct PPs, however, generally aren't headed by any specific preposition, and crucially, they are optional.

- Some verbs select both a direct object NP and a clausal complement, as in (13). The clausal complement to *persuade* can be either FINITE, *that they should leave early* or INFINITIVAL, *to leave early*.

- (13) Kim [_{VP} persuaded **his friends # that they should leave early / to leave early**].

Verbs like *convince*, *allow*, *encourage*, *force* and *permit* are also in this category, although some of these only select infinitival complement clauses.

- Often, a verb can appear in more than one sub-class. For example, *remember* may take no complement at all: it can be intransitive, as in *I can't remember*. But it can also be a transitive verb, as in (14a), or it can take one of three different kinds of clausal complement, either finite, as in (14b), or non-finite, as in (14c) and (14d). As usual, all the complements (in bold) are contained within the VP headed by *remember*:

- (14) a. Chris couldn't [_{VP} remember **that long shopping list**].
 b. Chris [_{VP} remembered **that they'd left it on the shelf**].
 c. Chris [_{VP} usually remembers **to pick up the list**].
 d. Chris [_{VP} remembered **leaving it on the shelf**].

The finite complement clause in (14b) has an overt subject *they* while the two different types of non-finite complement clause in (14c) and (d) have only an ‘understood’ subject, referring to *Chris*. Because there is no overt subject in these cases, some linguists regard such complements as less than clause-sized phrases, rather than a full clause. Here, I will assume they are clauses.

The non-finite complement in (14c) is an INFINITIVAL clause, containing the infinitive form of the verb *pick up*. In (14d), English has the non-finite *-ing* form of the verb in *leaving it on the shelf*. This is a clause type which Huddleston and Pullum (2002: Chapter 14) refer to as a GERUND-PARTICIPIAL clause: they argue that, contrary to what is normally proposed in traditional grammar, English has no distinction between a ‘gerund’ category and a ‘present participle’ category.

This section does not give a comprehensive list of verb classes, but it illustrates some of the most common sub-classes of verb found not just in English, but cross-linguistically.

4.1.6 Other heads and their complements

Heads other than verbs can also select different complement types. Prepositions, adjectives, adverbs, nouns and complementizers are discussed in this section. Again, their complements are shown in bold type.

- Prepositions have notable variety in their COMPLEMENT STRUCTURE, although less than verbs. We have already seen that some prepositions are always transitive, while others may be intransitive too. There are also prepositions that are only intransitive, such as *nearby*, as in *She lives just nearby*; we don’t get **She lives nearby the bank*. We can tell that *nearby* is truly a preposition by the fact that it co-occurs with the modifiers *just* and *right* (see Section 2.6): *She lives right/just nearby*. A number of prepositions take clausal complements, as *before* does in *Kim left before **the bus arrived***, where *the bus arrived* is an entire clause. And prepositions sometimes take PP complements, as *from* does in *He emerged [_{pp} from **under the blankets**]*.
- Adjectives occasionally take an obligatory complement, but this is rare. For instance, *fond* and *devoid* both take an obligatory PP complement headed by the preposition *of*, as in *fond of **fruit*** and *devoid of **meaning***; hence the ungrammaticality of **This speech is totally devoid*. A much larger number of adjectives take an optional PP complement, again headed by a specific preposition; some examples are *bad/good at **spelling***, *sorry for **your friend*** and *free from **any doubts***. Some adjectives (such as *sorry*, *happy*, *angry*, *glad*, *delighted*) take an optional CLAUSAL complement, as in *Kim felt [_{AP} **sorry that their friends weren’t around**]*. And adverbs sometimes have an optional complement too: [_{AdvP} *unfortunately for **me***], [_{AdvP} *independently from **her parents***].

We’ve seen so far, then, that verbs and prepositions often have an obligatory complement, and adjectives very occasionally do.

- The last major word class is that of nouns. Some complements to N are shown in bold in (15):

- (15) a. J. S. Blogg is [_{NP} a manufacturer **of tyres**].
 b. [_{NP} Lee's belief **in extraterrestrials**] is misguided.
 c. [_{NP} Her assertion **that Martians would land soon**] astounded me.
 d. They repeated [_{NP} their demand **for the library to stay open later**].
 e. [_{NP} Our decision **to leave**] came as no surprise.

Nouns often take optional complements, but not obligatory complements. One exception is the noun *denizen*: you have to be a denizen *of* somewhere, such as *denizens of the local bar*. Complements to N may be PPs, as in (15a), *of tyres*, or *of the local bar*, and (15b), *in extraterrestrials*. The specific preposition within these PP complements is *selected* by the head noun, and this shows that these truly are complements. Some nouns take optional clausal complements, as in (15c) and (15d). (15c) has a finite complement clause – *that Martians would land soon*, and (15d) and (15e) both have infinitival complement clauses – *for the library to stay open later* and *to leave*.

- The final word class in this section is that of COMPLEMENTIZER, a small, closed word class. A complementizer (abbreviated as C) is a word such as *that*, *for*, *whether* which introduces a clause, as we saw in Chapter 3. The clause it introduces is the complement to the head C, and the whole phrase (complementizer plus clause) can be termed CP, a Complementizer Phrase:

- (16) a. Mel said [_{CP} **that she was leaving**].
 b. [_{CP} **For Kim to go too**] would be surprising.
 c. I don't know [_{CP} **whether you should go / whether to go**].

As the examples in (16) show, some complementizers – such as *that*, (16a) – select a finite clause as their complement. Others – such as the prepositional complementizer *for* in (16b) – select a non-finite clause. And some can take either a finite or a non-finite complement clause, such as *whether* in (16c).

4.1.7 Summary: The main properties of complements vs. adjuncts

Here I give a brief summary of a vast topic, in order to help you to keep straight the major distinctions between the two kinds of dependent phrases.

(i) Optional vs. obligatory phrases?

- Adjuncts are always optional phrases. They have a fairly loose relationship with the head that they modify.
- Complements are often obligatory phrases, particularly the complements to verbs and prepositions. They have a close relationship with the head that they modify, and are selected by that head. Complements to adjectives are generally not obligatory, however (*I'm cross **with Lee**, I'm tired **of working***). Complements to nouns are essentially optional (*our hopes **for reconciliation**, the decision **to leave early***).

(ii) Limited vs. unlimited number of dependent phrases?

- A given head may be modified by a potentially unlimited number of adjuncts.
- A given head selects a strictly limited number of complements. Most heads have just one complement (e.g. a transitive verb or transitive preposition each select one object), though two or three complements are also fairly common: (*She put [the book] [on the shelf]*).

(iii) Properties of PP dependents

- PPs that are adjuncts are typified by having a wide range of head prepositions (*Lee danced in the ballroom / on the carpet / under the chandelier / for an hour* etc.).
- PPs that are complements are typified by having a specific head preposition in each of their usages (*We glanced at the clock, She sticks to her diet, They came across a small hut*).

(iv) Word class of complements and adjuncts

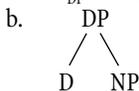
- We can't tell whether a phrase is a complement or an adjunct from its word class. For instance, an NP is most often a complement (to a head verb or preposition), but NPs can also be adjuncts (*He left **last week***). An adverb is most typically an adjunct (*Kim sings loudly*) but can be an obligatory complement, as in *Kim treats Lee **badly***.

4.1.8 Is the noun phrase really a determiner phrase?

In Section 2.3.4, I introduced the closed class of words called DETERMINERS (words like *the, a, some, this, these*) which, I proposed, pair up with nouns to form a noun phrase. In this chapter, we have followed the traditional view that the noun is the head of the NP; under this view, the determiner is one of its dependents. Some linguists consider the determiner to be a particular type of dependent known as a SPECIFIER; we could consider this a kind of adjunct that has a fixed position within the phrase (in English, preceding the head noun). On this view, the other closed class words that pair up with adjectives, adverbs and prepositions respectively (see Chapter 2) are also specifiers: this covers words like *very* in the AP *very happy* and the AdvP *very happily*, and words like *right* and *just* in the PPs *right inside* and *just underneath*.

However, a different view holds that in fact, the determiner is the head of the 'noun phrase', so that this phrase should really be considered a determiner phrase (DP). Under this view, the phrase has a head D, with an NP as its complement, as shown in (17): the head is *this* and its complement NP is in bold in (17a). The 'tree' in (17b) shows the same information as a diagram.

(17) a. [_{DP} this [_{NP} **box of dates**]]



Although the determiner *this* is clearly not the semantic head – the most important element in the phrase in terms of meaning – determiners do fulfil a number of the other criteria for head status outlined in Section 4.1.1. For instance (Section 2.3.4) many determiners can have the same distribution as the entire ‘noun phrase’, as in *I’ll take this/that/these/those/either/some*. The same is not true of the complement NP *box of dates* in (17): **I’ll have box of dates*. This suggests that, indeed, the determiner *this* is the syntactic head. It’s also the one obligatory part of the phrase in (17), which is more evidence for its head status. Furthermore, most determiners specifically select either a singular or a plural NP – *this box of dates* but *these boxes of dates*. So we can say that the head D requires its NP complement to agree with certain properties of the head. It seems, then, that various D elements may indeed select an NP as their complement.

The issue of whether D or N heads the ‘noun phrase’ is not explored further here, and I will continue to refer to a phrase like *this box of dates* as a ‘noun phrase’ without taking a stance on the DP hypothesis. Note, though, that the idea of a closed class word, D, heading a DP has parallels to the less controversial proposal that a closed class word, C (complementizer), heads a clause, which we then term CP.

4.1.9 Phrases within phrases

The dependents of a head are themselves grouped into phrases, and each smaller phrase has its own head which, in turn, has dependents. For instance, in the phrase *very bright sunflowers* in (1a), *very bright* is a dependent – an ADJUNCT to the head *sunflowers*. But in (1c) we see that *bright* is the HEAD of its own phrase, the AP *very bright*. We can indicate this thus: $[[\text{very } [\text{bright}]] \text{ sunflowers}]$.

Linguists often indicate the way a phrase occurs within a larger phrase by enclosing the phrases within square brackets, or by drawing a tree, as I did earlier (Chapter 5 has more discussion). Consider the verb phrase $[_{VP} \text{ sings in the bath}]$, which has the verb *sings* as its overall head. Within the VP there is an adjunct PP *in the bath*, headed by *in*. The brackets indicate the beginning and end of each phrase: $[_{VP} \text{ sings } [_{PP} \text{ in the bath}]]$. Within the PP there’s a dependent NP, *the bath*, which we can also bracket: $[_{VP} \text{ sings } [_{PP} \text{ in } [_{NP} \text{ the bath}]]]$. In this way, we get phrases nested within phrases which, in turn, are nested within phrases. As noted in Chapter 1, this nesting is termed ‘hierarchical structure’, and is a property common to all languages. Each phrase has its own head and its own dependents. So although the PP *in the bath* is a dependent to the head of the whole VP, *sings*, this PP also has its own head and dependents. Within its own phrase a word can’t simultaneously be both a head and a dependent. For instance, the preposition *in* is a DEPENDENT of the verb *sings* within the VP, but within its own phrase – within the PP *in the bath* – *in* is the HEAD.

4.2 WHERE DOES THE HEAD OCCUR IN A PHRASE? HEAD-INITIAL AND HEAD-FINAL LANGUAGES

In this section, I introduce a two-way system of classifying languages which looks at the position of the head in relation to its complements. There is a strong tendency, cross-linguistically, for the head to occur in a fixed position in relation to its complements, and for this order to be the same across all phrases within a language. In HEAD-INITIAL languages the head *precedes* its complements, and in HEAD-FINAL languages the head *follows* its complements. The heads of each phrase are in bold type in this section.

4.2.1 Head-initial languages

English is a head-initial language. Example (18) shows that complements to V, P, A and N all follow the head (which is shown in bold):

- (18) a. [_{VP} **likes** chips]
 b. [_{PP} **into** the water]
 c. [_{AP} **fond** of chips]
 d. [_{NP} **admiration** for Kim]

In (18), both the head verb *likes* and the head preposition *into* precede their complement NPs, while both the head adjective *fond* and the head noun *admiration* precede their complement PPs.

The Celtic languages are all good examples of the head-initial type; I illustrate here with Welsh. As in English, the head P precedes its NP complement:

- (19) [_{PP} **dros** y ffordd] (Welsh)
 over the road

And the verb is also initial within the VP: in (20), *yfed* 'drink' precedes its complement, namely the direct object NP *paned o de* 'a cup of tea'.

- (20) Ddaru Ceri [_{VP} **yfed** paned o de].
 did Ceri drink.INFIN cupful of tea
 'Ceri drank a cup of tea.'

- (21) [_{VP} yfed [_{NP} **paned** o de]]
 drink.INFIN cupful of tea
 '(to) drink a cup of tea'

And in (21), we see that within the VP, there is an object NP which has the head noun preceding its PP complement. The object is *paned o de*, and the head noun *paned* 'cupful' is initial in that NP. You should also be able to see from (21) that the PP *o de* 'of tea' is again head-initial, with the preposition *o* preceding its NP complement.

My final examples of a head-initial language are from the Austronesian language

Tinrin. Example (22) shows that a head verb (in bold) precedes its complement clause within VP, and (23) shows that the head noun (in bold) precedes its complement PP:

- (22) u [_{VP} **tramwâ** mwâ ke maija wake] (Tinrin)
 I know that you much work
 'I know that you work hard.'
- (23) [_{NP} **kò** rugì beebòrrò nrâ mwîê]
 news about drowning POSSESSIVE woman
 'the news of the woman's drowning'

4.2.2 Head-final languages

Examples of clearly HEAD-FINAL languages are Japanese, Turkish, and Lezgian (a language spoken in Daghestan and Azerbaijan). Examples (24) through (27) illustrate from Japanese, with the head word again in bold in each phrase ('dative' is a special case that's often used for recipients, as here):

- (24) Taroo-ga [_{VP} Hanako-ni hana-o **ageta**]. (Japanese)
 Taro-NOM Hanako-DATIVE flower-ACC gave
 'Taro gave Hanako flowers.'
- (25) Taroo-ga [_{VP} tana-ni hon-o **oita**].
 Taro-NOM shelf-at book-ACC put
 'Taro put a book on the shelf.'
- (26) [_{PP} tomodati-**to**]
 friend-with
 'with a friend'
- (27) [_{NP} sono tesuto e no **zisin**]
 that test to POSSESSIVE confidence
 'confidence in that test'

Examples (24) and (25) show that in Japanese verb phrases, the verb is final: in each example the verb has two complements, and these both precede the verb. Example (26) shows that Japanese has a head P *to* 'with' which follows its complement *tomodati* 'friend'. So Japanese is postpositional, not prepositional: see Section 2.6.2. Example (27) shows that the head noun *zisin* 'confidence' follows its complement *sono tesuto e no* 'in that test' (the possessive item *no* is a case marker, showing the relationship between the head noun *zisin* 'confidence' and its complement).

And from Turkish, I illustrate with an adjective phrase: the complement to the adjective (shown in bold) precedes that adjective, as expected in a head-final language. Note the very slightly different use of dative case here, too:

- (28) koca-sin-a **sadık** (Turkish)
 husband-3SG-DATIVE loyal
 'loyal to her husband'

- Ndyuka is a creole language of eastern Suriname. The example in (32) shows an AP (in brackets), *bun fu sama nyan*, with the head adjective *bun* ‘good’ preceding its complement, which is a whole clause *fu sama nyan* ‘for people to eat’. Since the head precedes this complement, this is therefore a head-initial construction.

4.3 HEAD-MARKING AND DEPENDENT-MARKING LANGUAGES

Section 4.2 examined one major cross-linguistic typology, known as head-placement. In this section, we look at another important typological distinction: that between HEAD-MARKING and DEPENDENT-MARKING languages. Section 4.3.1 defines the terms and illustrates the constructions under discussion. Remaining sections give examples from languages of each type, construction by construction, ending by examining the wider picture of typological distinctions between languages.

4.3.1 Definitions and illustrations: Syntactic relationships between heads and dependents

Table 4.1 illustrates four different syntactic relationships – dependencies – between a head and its dependent(s). For ease of exposition, Table 4.1 shows each head before its dependents, but this shouldn’t be taken to imply that only head-initial languages are under discussion; this is not at all the case, as we’ll see.

Table 4.1
Syntactic relationships between a head and its dependent

	Head	Dependent
i.	postposition/preposition	object NP
ii.	verb	arguments of the verb (e.g. subject, object)
iii.	(possessed) noun	possessor NP
iv.	noun	adjective

First, I show these four construction types in English; the relevant heads are given in bold:

- i. **in** [_{NP} the shower] (P + NP)
- ii. Kim **loves** Lee (Su + V + Obj)
- iii. Kim’s **house** (possessor NP + N)
- iv. red **book** (attributive adjective + N)

In this section, we’ll see that languages often mark either the head word or its dependent(s) in some way to signal the syntactic relationship between them. Either the head or the dependent(s) (or sometimes both) will occur in some special form, perhaps taking an affix, or exhibiting some other change in word form. Let’s start with a preliminary illustration. In the noun phrase *Kim’s house*, the HEAD is the noun *house* (because *Kim’s house* IS A house) and the DEPENDENT is the possessor NP *Kim*.

In English, the dependent occurs in a special form here: it has the possessive *-s* affix. The possessed head noun, *house*, however, has no special morphology: it is in its basic form. The *-s* affix shows the possessor NP *Kim* to be a dependent (of a particular kind) to the head *house*. Since it's the dependent that receives the *-s* marking, rather than the head, then *Kim's house* is an example of DEPENDENT-MARKING. In a HEAD-MARKING language, however, the head noun 'house' would occur in some special form. We'll see an example in the discussion of possessive NPs in Section 4.3.4.

The fact that the syntactic relationship between a head and dependent may be marked either on the head or the dependent gives us a broad TYPOLOGICAL distinction (= a division into language types) between HEAD-MARKING and DEPENDENT-MARKING languages. Here's what we expect to find. Typical head-marking languages are those with extensive agreement or CROSS-REFERENCING – heads such as verbs and nouns are marked to agree with grammatical properties of their arguments, for instance number, person and gender. A reminder of such a language, Kambera, can be found in Section 2.2.2.4. For instance, we would expect a head-marking language to have markers on the verb indicating both the subject and the object.

Typical dependent-marking languages, on the other hand, have well-developed case systems: this means that the dependents are marked to show their grammatical relation, say to a head verb or preposition. For instance, subjects and objects themselves appear in a special form which shows that they bear these particular grammatical relations. We've already seen an example of this from Japanese in (4). Subjects bear a special case (nominative), while objects take a different case, known as accusative. English displays a small amount of dependent-marking here too, although it's restricted to the set of first and third person pronouns: see Section 2.3.2. English full NPs don't differ in case depending on grammatical relation, so we get both *My sister saw the girl* and *The girl saw my sister*.

I turn next to examples of the four constructions in Table 4.1 from languages of both types. Dependent-marking languages are more familiar to most readers of this book than are head-marking languages, and so are illustrated first in each section that follows.

4.3.2 Head preposition/postposition and its NP object

English has no marking at all on either head or dependent in this construction: *in the shower*.

4.3.2.1 Dependent-marking in the PP

First, then, I discuss DEPENDENT-MARKING on the object of prepositions within the PP in German. The basic form of the NP meaning 'my friend' is *mein Freund*. If this NP is used as the object of a preposition, then it's a dependent to the head preposition. German prepositions mark their dependent NPs by requiring them to appear in some particular CASE. Example (33) illustrates with two different head prepositions, each requiring a different case:

A second example of head-marking within PP comes from Welsh. Most prepositions in Welsh inflect to agree with their pronominal objects. The basic form of the preposition meaning ‘on’ is *ar*, and three members of its INFLECTIONAL PARADIGM (see Section 1.2.2.3) are shown in (36):

(36)	<i>arna</i>	<i>i</i> ;	<i>arno</i>	<i>fo</i> ;	<i>arni</i>	<i>hi</i>	(Welsh)
	on.1SG	me	on.3.M.SG	him	on.3.F.SG	her	
	‘on me’		‘on him’		‘on her’		

The dependent pronouns in (36) retain their usual, unmarked form (they have no case-marking) while the head preposition *ar* inflects to agree with the pronoun: *arna*, *arno*, *arni*. In the third person singular, the inflection is for person, number and gender. Inflected prepositions are found throughout the Celtic family (of which Welsh is a member) and in a number of other unrelated families, including Semitic (e.g. Arabic, Hebrew).

4.3.3 The clause: A head verb and the arguments of the verb

As noted earlier, English has dependent-marking in the clause only for a subset of pronouns, and not for full NPs at all. English has a tiny amount of head-marking in the clause, as we’ll find later. See if you can figure out what this is before our discussion gets there.

4.3.3.1 *Dependent-marking in the clause*

The main verb in a clause has NP arguments which are its dependents. If we take a simple example of a transitive verb from the DEPENDENT-MARKING language Japanese, we see that the two dependents – subject and object – are each marked with a specific case (by affixes, shown in bold):

(37)	<i>Taroo-ga</i>	<i>tegami-o</i>	<i>kaita</i> .	(Japanese)
	Taroo-NOM	letter-ACC	wrote	
	‘Taroo wrote a letter.’			

The head here, the verb *kaita* ‘wrote’, simply appears in its past tense form, and bears no information about its dependents. Specifically, it has no person or number inflections – no affixes to show who’s doing the writing or what is being written. So there’s no head-marking. But the dependent NPs are case-marked to show their relationship to the head verb: as we saw earlier, the subject of a verb in Japanese bears nominative case, and the object bears accusative case. Again we see that case indicates a dependent-marking construction.

German subjects and objects are also dependent-marked with different cases, again nominative for the subject of a verb and accusative for the object:

(38)	<i>Der</i>	<i>Hund</i>	<i>sah</i>	<i>den</i>	<i>Vogel</i> .	(German)
	the.NOM	dog	saw	the.ACC	bird	
	‘The dog saw the bird.’					

instance, the boy's name – might be mentioned once at the start of the discourse, or else it may be obvious from the context. Full NPs don't need to be explicitly present in most sentences.

Technically, then, there is no grammatical 'agreement' in examples like (40) – the pronominal affixes alone represent the arguments of the verb, and there is no independent subject or object that the verb could 'agree' with here. For this reason, some linguists reserve the term 'agreement' for constructions or languages in which a verb or other head really does agree with another element in the clause, and instead use the term CROSS-REFERENCING for languages like Kambera where pronominal affixes represent the arguments on their own.

Even when the NP dependents of the verb *are* present in the sentence, the head verb is still marked to cross-reference (or agree with) them, as in (41):

- (41) I Ama_S na_S-kei-ya_O na rí muru_O. (Kambera)
 the father 3SG.SU-buy-3SG.OBJ the vegetable green
 'Father buys the green vegetables.'
 (Literally, 'Father he-buys-it the green vegetable'.)

In (41), I've indicated both the subject itself, *I Ama* 'father', and also the subject marker on the verb with a subscript_S (for 'subject'), and I've shown both the object NP *na rí muru* 'the green vegetable' and the object marker on the verb with a subscript_O (for 'object').

Please make sure you understand this kind of head-marking before going further, because it will be vital for understanding later chapters.

4.3.4 Head noun and dependent possessor NP

I turn next to the syntactic relationship between a possessed head noun and the possessor NP which is a dependent to that head. We've already seen one example of dependent-marking in this construction in English – recall the discussion of *Kim's house*. The special pronominal forms *my, your, his, her, our, their* which replace possessive *-s* (we don't say **them's house*) are also examples of dependent-marking.

4.3.4.1 Dependent-marking in the possessive construction

Another example comes from a Papuan language called Mangga Buang (POSS is the 'possessive' marker):

- (42) a. sa-te voov b. yi-te bayêên (Mangga Buang)
 1SG-POSS dog 3SG-POSS village/house
 'my dog' 'his/her village (or house)'

Note that in these examples, there is a possessive marker separate from the person/number marker, whereas English uses special possessive determiners here, *my, your, his, her* etc., which encapsulate both person/number *and* possession in a single form.

4.3.4.2 Head-marking in the possessive construction

Now consider the construction shown in (43), from the HEAD-MARKING language Saliba, an Austronesian language from the island of Saliba, Papua New Guinea:

- (43) sine natu-na (Saliba)
 woman child-3SG
 'the woman's child'

The word order of the (dependent) possessor and the possessed (head) N in (43) is just as in English, but the possessor *sine* 'woman' has no marking, while the head *natu* 'child' bears a third person singular suffix which marks agreement with the possessor, *sine*, 'woman': literally, (43) means 'woman child-her'.

4.3.4.3 Double marking in the possessive construction

It is, in fact, rather common for a language to mark *both* the head and the dependent in the possessive construction: such double marking (i.e. both head- and dependent-marking within a single construction) is illustrated in (44) from a Quechuan language called Ayacucho:

- (44) a. runa-pa wasi-n (Ayacucho)
 man-GENITIVE house-3.POSSESSIVE
 'a person's house'
- b. qam-pa wasi-ki
 you-GENITIVE house-2.POSSESSIVE
 'your house'

GENITIVE is a case marker – like *-s* in *Kim's house* – which shows possession; in other words, it shows the relationship between the possessor and the thing possessed (the head N meaning 'house'). Like all case marking, this is an instance of dependent-marking. The head, though, is also marked in this construction to agree with the possessor: it indicates the PERSON of the possessor, so third person for *runa*, 'man/person', the possessor in (44a), and second person for *qam*, 'you', in (44b).

4.3.5 Head noun and dependent AP

I turn finally to a head noun and a dependent adjective that modifies it. There are no examples from English, since neither noun nor adjective is marked in any way.

4.3.5.1 Dependent-marking in the noun + modifying adjective construction

Dependent-marking means here that the attributive adjective agrees with properties of the head noun, such as gender and number. This occurs in many European languages; (45) illustrates from Spanish:

- The Chadian Arabic examples in (50) are dependent-marking. The adjective meaning ‘tall’ is a dependent on the head noun in each example, and agrees with that noun in gender.

4.3.7 Some typological distinctions between languages

Many languages fall fairly neatly into either the head-marking class or the dependent-marking class. Good examples of HEAD-MARKING languages are Abkhaz (a northwest Caucasian language) and the native American language Navajo. In fact, the indigenous language families of the Americas, and in particular North America, are nearly all head-marking: these families include Mayan (e.g. Jacaltec, Tzotzil), Athabaskan (e.g. Navajo), Iroquoian (e.g. Mohawk, Cherokee), Algonquian (e.g. Cree, Blackfoot), Siouan (e.g. Crow, Lakota) and Salish (e.g. Squamish).

Conversely, many languages from the Indo-European family (to which English belongs) are heavily DEPENDENT-MARKING, including German, Greek, Armenian and the Slavonic languages (e.g. Russian, Polish, Czech, Bulgarian etc.). But dependent-marking languages also predominate among the native Australian languages known as Pama-Nyungan (e.g. Dyirbal, Yidiny); the Northeast Caucasian languages (e.g. Chechen); and the Dravidian languages of southern India (e.g. Malayalam).

Another typological possibility is for the relationship between a head and its dependent not to be formally marked at all. This happens most often in languages which have very little morphology (= variation in the forms of words), such as Chinese, Vietnamese, and indeed English. In Chinese, for example, pronouns (and full noun phrases) have the same form whether they are subjects or objects:

- (51) a. **Wo** changchang jian **ta**. (Chinese)
 I often see he
 ‘I often saw him.’
 b. **Ta** changchang jian **wo**.
 he often see I
 ‘He often saw me.’

Example (51) shows that *wo* translates as either ‘I’ or ‘me’, and *ta* as either ‘he’ or ‘him’ (in fact, *ta* translates both ‘he/him’ and ‘she/her’). So the dependent noun phrases aren’t marked in any way in these Chinese examples. In other words, there is no CASE-MARKING in Chinese: the dependents of a verb are not marked to show their relationship to that verb. And neither is there any head-marking, since the verb doesn’t undergo agreement with either the object or the subject. Note that in such a language, the word order is crucial (as in English) to show who’s doing what to whom. We can conclude that although many languages do have head-marking or dependent-marking, some languages have neither.

English has very little formal marking on either heads or dependents. For example, in an English PP such as *in the shower*, neither the head P *in* nor its dependent NP *the shower* is marked to show the syntactic relationship between them. The same

- b. Wir **sehen** den Vogel.
 we.NOM see.PRES.1PL the.ACC bird
 ‘We see the bird.’

For the most part, German is a typical dependent-marking language: dependent pronouns and full NPs are all case-marked, the subjects as *NOMINATIVE* and the objects as *ACCUSATIVE*. But (53) shows that German also has subject/verb agreement, which is head-marking, and this is much more extensive than in English. So in (53a) we have *sehe*, the first person singular form of the verb, when the subject is *ich* ‘I’, and in (53b) *sehen*, the first person plural form, when the subject is *wir* ‘we’.

In fact, head-marking on verbs in the form of verbal agreement (particularly agreement with subjects) is very prevalent cross-linguistically – even in languages which are otherwise systematically dependent-marking. We can regard this kind of head-marking as a property which is typical of both head- *and* dependent-marking languages, rather than seeing it just as belonging to the head-marking class of languages.

4.3.8 Summary

We have seen in this section that languages divide into various classes in terms of the head-marking versus dependent-marking typology. Some languages rarely mark the syntactic relationships between head and dependent at all; these are languages with very little morphology, such as Chinese. Amongst languages that do mark the relationships, there are two major possibilities: the head may be marked or else the dependent may be marked. Some languages exhibit both head- and dependent-marking constructions. Finally, I noted that the occurrence of verbal agreement, a head-marking pattern, is particularly common, even in languages which are generally dependent-marking.

FURTHER READING

Information on the position of the head within a phrase (head-initial or head-final) can be obtained from Chapter 2 of Song (2001). One approach to heads and their dependents can be seen in Hudson (1984, 1990, 2007); in a different theoretical framework, see Radford (1988). Radford also provides extensive discussion of complements and adjuncts (though from a largely English perspective); see especially Chapters 1 through 5 (I recommend reading my Chapter 5 first). The question of whether or not the determiner heads the noun phrase has generated much interest over the years: two central papers are Zwicky (1985) and Hudson (1987). The seminal reading for Section 4.3 on head-marking and dependent-marking languages is Nichols (1986), though I don’t recommend tackling this until you’ve finished this book.

EXERCISES

1. The examples in (1) through (5) all contain at least one noun phrase.

Task: (i) Pick out all the NPs, and put them in square brackets. Make sure that you get the whole of each NP inside your brackets; i.e. the head noun and all its dependents. In some cases, an NP may have another NP embedded within it. Make sure you bracket these too. (ii) List all the subject NPs, all the direct object NPs, and all the NP predicates.

- (1) My idiot of a neighbour wastes stacks of water on his garden.
 (2) This is a planet that could engulf all the surrounding matter.
 (3) They encountered a bigger problem over the fees rise than they initially anticipated.
 (4) This is too long a story for me to tell you right now.
 (5) The only day currently available for your interview is March 12.

2. You will need to revise Section 4.2 for this exercise.

Task: Examine the data in (1) through (5) and decide, for each of the phrases in brackets, (i) what is the head of the phrase (ii) whether the phrase is HEAD-FINAL or HEAD-INITIAL and (iii) what is the word class of the phrase:

- (1) Ahmet [kitab-ı öğrenci-ler-e sat-tı] (Turkish)
 Ahmet book-ACC student-PL-DATIVE sell-PAST
 'Ahmet sold the book to the students.' (Kornfilt 1997)
- (2) [Ondarrúra áño] (Basque)
 Ondarroa up.to
 'up to Ondarroa' (Hualde *et al.* 1994)
- (3) [jek Petritesko čavo] (Romani)
 a Peter.GENITIVE son
 'a son of Peter's' (Matras 2002)
- (4) [yo-gu numa e-na] (Saliba)
 POSS-my house in-3SG
 'in my house' (Mosel 1994)
- (5) Wnaeth hi [fynd â 'r plant i 'r dre ddoe] (Welsh)
 do.PAST.3SG she go.INFIN with the children to the town yesterday
 'She took the children into town yesterday.'

3. You will need to revise Section 4.3 for this exercise. Examine the data in (1) through (7).

Task: (i) Decide what construction type each phrase illustrates – that is, what is the head and what is the syntactic relationship between head and dependent(s).

(ii) Decide whether each construction is HEAD-MARKING or DEPENDENT-MARKING. (iii) See if you can work out why (4a) and (4b) differ from each other.

Hints

- As an example, here is the answer for (1): this is a possessive construction; the head is the possessed noun *halgan* ‘leg’, and its dependent is the possessor *beje* ‘man’. The head is marked for third person singular possessive, POSS, (agreeing with ‘man’), while the dependent has no marking; this is therefore a head-marking construction.
- In (7), concentrate only on the portions in brackets. DATIVE (which we have met before) and ABLATIVE are two different cases. Ablative typically refers to going *from* some location.

- (1) beje halgan-in (Evenki)
 man leg-3SG.POSS
 ‘the man’s leg’ (Nedjalkov 1997)
- (2) dee-n aaxča (Chechen)
 father-GENITIVE money
 ‘father’s money’ (Nichols 1986)
- (3) Ye-sa-núhwe’-s (Mohawk)
 F.SG.SU-2SG.OBJ-like-HABITUAL
 ‘She likes you.’ (Deering and Delisle 1976, cited in Baker 1996)
- (4) a. xiri-con Xijam (Wari)
 house-3M.SG (male name)
 ‘Xijam’s house’
 b. pije’-nequem Hatem
 child-POSS.3F.SG (female name)
 ‘Hatem’s child’ (Everett and Kern 1997)
- (5) a. le rakles-k-i dej (Romani)
 the.M boy-GEN-F mother
 ‘the boy’s mother’
 b. le rakles-k-e phrala
 the.M boy-GEN-PL brothers
 ‘the boy’s brothers’ (Matras 2002)
- (6) az ember ház-a (Hungarian)
 the man house-3SG
 ‘the man’s house’ (Nichols 1986)
- (7) a. Hasan [köy-e doğru] yürü-dü. (Turkish)
 Hasan village-DATIVE towards walk-PAST
 ‘Hasan walked towards the village.’
 b. Hasan [Ali-den önce] git-ta.
 Hasan Ali-ABLATIVE before go-PAST
 ‘Hasan left a little while before Ali.’ (Kornfilt 1997)

4. The data in this exercise are from a Chadic language called Hdi, spoken in Cameroon, and are taken (slightly adapted) from Frajzyngier (2002).

Task: Examine all the data in (1) through (10), and work out the function of the morpheme *tá*, which I have left un glossed in these examples. Where exactly does it occur?

- (1) ngatsa-f-ngats-i *tá* lfid-a lgut
 have-up-have-1SG *TÁ* new-GEN cloth
 'I have new clothes.'
- (2) tsgha-da-f xaxə̀n *tá* sani
 put.up-away-up they *TÁ* one
 'They sent up one (bag).'
- (3) ghwaghwa-ghwaghwa kri
 bark-bark dog
 'A dog barked.'
- (4) si midu-u
 PAST inside-1DUAL
 'The two of us were inside.'
- (5) skwa-skw-i *tá* plis nda ma na hla
 buy-buy-1SG *TÁ* horse and female DEM cow
 'I bought a horse and a cow.'
- (6) nda ngh-i *tá* pta
 STATIVE see-1SG *TÁ* mat
 'I saw the mat.'
- (7) ta skalu-lu *tá* skalu girvidik
 IMPF dance-SU *TÁ* dance(N) night
 'They danced all night.'
- (8) nda ngh-i *tà* pta
 STATIVE see-1SG on mat
 'I saw (it) on the mat.'
- (9) vra-k-vr-i dzagha ka mbaz-i *tá* mbaza
 return-in-return-1SG home then wash-1SG *TÁ* wash(N)
 'I returned home and washed.'
- (10) ta xanay tsa mndu ya *tá* xani dagala
 IMPF sleep(V) the man DEM *TÁ* sleep(N) large
 'That man sleeps a lot.'

5. Before tackling this exercise, you should revise Section 3.3.3. In (1) through (4), you see some serial verbs in Yimas, a Papuan language of New Guinea. The data are all from Foley (1991), with some small adaptations.

Hints

- There are some elements in the glosses that need a few words of explanation. The gloss ‘A’ is for the ‘agent’ (here, the subject) of a transitive verb. We will see more about this term in Chapter 6. For instance, in (1)–(3), there is a verbal prefix *n-*, glossed as 3SG.A, meaning a third person singular agent; this gives us the subject ‘he’ in the translations. And in (4), there is a prefix *ka-*, glossed as 1SG.A, meaning a first person singular agent; this gives us the subject ‘I’ in the translations. The gloss CONT (for continuous) gives an ongoing event, just like *walking* and *sitting* do in the English translations.
- The serial construction itself in each of these examples is a single grammatical word that comprises a number of distinct morphemes; in other words, no part of it can be split off and stand as an independent word. Each of the examples (1) to (3) contains only *one* grammatical word. Make sure you understand this before moving on. There are two independent words in (4), as well as the serial verb construction. These are the words for ‘water’ and ‘canoe’.
- Verb serialization can be formed in two different ways in Yimas, giving rise to two different interpretations. The two serial verbs in Yimas can be simply juxtaposed, i.e. placed next to each other, as is the case in (1), (2) and (3). This implies that the two events are simultaneous, or are very close (in time and space). Alternatively, the serial verbs can be connected by a morphological marker, most commonly *-mpi*, marked SEQ for ‘sequential’; this construction is used for events that occur one after the other, so are sequential, but where one event did not cause the other. An example is (4).

Tasks: (i) First, examine the serial verb constructions in (1) through (4) and decide which typical properties of verb serialization can be detected in these examples. Be as specific as possible in your answer:

- (1) *impa-n-yakal-kulanaŋ-kanta-k*
3DUAL.OBJ-3SG.A-CONT-walk-follow-TENSE
‘He was walking following those two.’
- (2) *pu-n-yakal-çaŋ-tantaw-malak-ntut*
3PL.OBJ-3SG.A-CONT-with-sit-talk-REMOTE.PAST
‘He was sitting down conversing with them.’
- (3) *ura-n-irm-wampaki-pra-k*
fire.OBJ-3SG.A-stand-throw-toward-TENSE
‘He stood throwing fire toward (them).’
- (4) *arm-n kay i-ka-ak-mpi-wul*
water-in canoe SG.OBJ-1SG.A-push-SEQ-put.down
‘I pushed the canoe down into the water.’

(ii) The serial verb examples in (5) and (6) are variants of (4), but both are ungrammatical; these are not possible constructions. Why not? Which principle of verb serialization do these violate? **NB** There is a certain freedom of word order in Yimas, meaning that independent words such as *kay*, ‘canoe’, can be

found in various positions in the clause. But *in itself*, this is not at all relevant to your answer:

- (5) *kay i-ka-ak-mpi arm-n wul
 canoe SG.OBJ-1SG.A-push-SEQ water-in put.down
 ('I pushed the canoe down into the water.')
- (6) *i-ka-ak-mpi kay wul arm-n
 SG.OBJ-1SG.A-push-SEQ canoe put.down water-in
 ('I pushed the canoe down into the water.')

The example in (7), by way of contrast, is *not* an instance of verb serialization. Instead of having a meaning that refers to a single event, pushing the canoe into the water, there are two *separate* clauses here, as the English translation reflects. Each verb receives markers of its own in this construction.

- (7) kay ak-mpi i-ka-wul arm-n
 canoe push-SEQ SG.OBJ-1SG.A-put.down water-in
 'I pushed the canoe and put it into the water.'

(iii) The examples in (8) and (9) are variants of (7), again showing freedom of word order for independent words; both are fully grammatical. Why are these examples possible, unlike the serial verb constructions in (5) and (6)?

- (8) kay ak-mpi arm-n i-ka-wul
 canoe push-SEQ water-in SG.OBJ-1SG.A-put.down
 'I pushed the canoe and put it into the water.'
- (9) ak-mpi kay i-ka-wul arm-n
 push-SEQ canoe SG.OBJ-1SG.A-put.down water-in
 'I pushed the canoe and put it into the water.'

6. This exercise examines a construction known as 'quantifier float' in a variety of Irish English known as West Ulster English (data and discussion taken from McCloskey 2000). Standard English allows both of the constructions in (1), where (a) is said to have a floating quantifier (*all* or *both*), meaning that it's floating free of the phrase (*they*) that it modifies directly in (b):

- (1) a. They have all/both gone to bed.
 b. They all/both have gone to bed.

Many varieties of English also allow questions of the kind in (2). (If you're not a speaker of such a variety, note that *what all*, *who all* etc. require that you answer with a list, and moreover, a full list; if you met Tom, Jack and Nicky in Derry, answering just *Tom* would not be what was required.)

- (2) a. What all did you get for Christmas?
 b. Who all did you meet when you were in Derry?
 c. Where all did they go for their holidays?

What distinguishes West Ulster English is that it also allows questions of the kind in (3), which have quantifier float. The quantifier is said to be 'stranded' (left behind; not attached to the *wh*-word) in these examples:

- (3) a. What did you get all for Christmas?
 b. Who did you meet all when you were in Derry?
 c. Where did they go all for their holidays?

Some further examples from the same dialect follow.

Task: (i) From (3), (4) and (5), formulate an initial hypothesis about where exactly (in syntactic terms) the floating quantifier is 'stranded' in this dialect; note that (5b) and (c) are ungrammatical, and this must be accounted for in your hypothesis too:

- (4) a. What did you give all to the kids?
 b. What did you put all in the drawer?
 c. Who did you meet all up the town?
- (5) a. What did she buy all in Derry at the weekend?
 b. *What did she buy in Derry at the weekend all?
 c. *What did she buy in Derry all at the weekend?

(ii) Next, consider the data in (6): are these examples consistent with your hypothesis? If so, well done; if not, please formulate a new hypothesis that correctly predicts the grammaticality of (6):

- (6) a. Tell me what you got all for Christmas.
 b. Tell me what you've been reading all.
 c. I don't remember what I said all.

(iii) Next, consider the data in (7), which are ungrammatical; do these affect your hypothesis? If the answer is not at all, well done. If these data are not consistent with your hypothesis, can you formulate a new hypothesis that *is* consistent with all the data seen so far?

- (7) a. *Who did you talk all to?
 b. *What were you laughing all at?

Although the data in (7) are completely impossible, those in (8) are, according to McCloskey, only 'slightly degraded'; in other words, a linguistic analysis would have to account for them as possible data. In what way would these require your hypothesis to be amended?

- (8) a. ?Who did you talk to all (at the party)?
 b. ?Who was he laughing at all?

How do we identify constituents?

This chapter returns to the theme of sentence structure, introduced in Chapter 1. We saw in Chapter 4 that phrases consist of a head word and its complements, plus any optional modifiers to that head. In this chapter, we discover how to identify phrases, and how to distinguish a phrase from a random string of words. The phrases which make up sentences are known as the CONSTITUENTS of a sentence. We will see how constituents are represented in tree diagrams, and start to investigate how languages differ in terms of constituency.

5.1 DISCOVERING THE STRUCTURE OF SENTENCES

Section 5.1.1 demonstrates the existence of syntactic structure, in particular by looking at ambiguous phrases and sentences. Section 5.1.2 introduces three syntactic tests for constituent structure, and Section 5.1.3 examines the ways in which linguists formally represent constituent structure.

5.1.1 Evidence of structure in sentences

One way to show that syntactic structure actually exists is to examine sentences which are syntactically ambiguous; that is, sentences which have more than one meaning. Not all ambiguity is syntactic: some is lexical, as in *Lee went down to the bank*; does this mean ‘the river bank’ or ‘the place where money is kept’? In other cases, though, ambiguity arises because we can’t tell which words group together to form a phrase. This is syntactic ambiguity. For instance, a sentence like the following appeared in a British national newspaper, causing an unforeseen breakdown in communication.

- (1) Black cab drivers went on strike yesterday.

Readers wrote in to say, what did it matter what colour the drivers were? But, of course, the newspaper actually intended *black* to modify *cab*, not to modify *cab drivers*. The two different meanings reflect the fact that the phrase *black cab drivers* has two different STRUCTURES. We can indicate this by using brackets to show which words group together; different bracketings indicate different phrase structures. Example (2) illustrates these different structures: the outraged readers had interpreted the sentence as in (2a), and the newspaper had intended (2b). (For

readers unfamiliar with British culture, a ‘black cab’ is a particular kind of black taxi found in major cities.)

- (2) a. Black [**cab drivers**] went on strike.
 b. [**Black cab**] drivers went on strike.

In both (2a) and (2b), the whole phrase *black cab drivers* is a constituent of the sentence, but, as the brackets show, the words inside that phrase group together in different ways, depending on what *black cab drivers* actually means. In (2a), *cab drivers* is a constituent of the sentence, whereas in (2b), *black cab* is a constituent of this sentence. A constituent is a set of words that forms a phrase in a sentence. If you say aloud the distinct phrases in (2), you’ll probably find that they each have a different intonation pattern; sometimes we show by our intonation which words group together to form constituents.

Occasionally we can discover which words form constituents by looking at inflections, as in the case of English possessive *-’s* (see Section 1.3.3). The affix *-’s* attaches to the end of a phrasal constituent (an NP) giving *Lee’s*, *the boy’s* and so on, so we can use *-’s* to discover whether or not a string of words is an NP (and therefore a constituent). This gives some results that might initially seem surprising, as in (3):

- (3) I’ll be back in [**an hour or so**]’s time.

Here, *an hour or so* must be a constituent, an NP, since *-’s* can attach to the whole phrase.

The *-’s* inflection can itself be the cause of syntactic ambiguity, because we can’t always tell what constituent it’s attached to:

- (4) The boy and the girl’s uncle stayed to dinner.

This, of course, is ambiguous as to whether just one person stayed, or two, as the variants with tag questions make clear:

- (5) a. The boy and the girl’s uncle stayed to dinner, didn’t he?
 b. The boy and the girl’s uncle stayed to dinner, didn’t they?

So in (4) there are two different meanings – or READINGS, to use the technical term – and, as we will see, each of these readings corresponds to a particular CONSTITUENT STRUCTURE, that is, a particular grouping of words.

The ambiguity in (4) lies in the phrase *the boy and the girl’s uncle*. This whole string of words is a constituent of the sentence in both readings, but its internal structure is different in each case. We can’t tell if *-’s* is suffixed to an NP *the boy and the girl*, in which case the uncle is related to both of them, or if *-’s* is just suffixed to an NP *the girl*, in which case the uncle is related to her, but not to the boy. Both options are possible, hence the ambiguity. The structures of the two alternatives are shown in (6), where the brackets mark out the two possible constituents that *-’s* can attach to:

- (6) a. [The boy and the girl]'s uncle stayed. (one person)
 b. The boy and [the girl]'s uncle stayed. (two people)

Only in (6a) is the sequence *the boy and the girl* a whole phrase, a constituent of the sentence.

So, a sequence of words that forms a constituent in one environment need not necessarily do so in another environment. There is absolutely *no* rule of 'once a constituent, always a constituent'. To underline this point, compare the (a) and (b) sentences here:

- (7) a. The students wondered how cheap textbooks could be obtained.
 b. The students wondered how cheap textbooks could be.
 a'. The students wondered how [cheap textbooks] could be obtained.
 b'. The students wondered [how cheap] textbooks could be.

In (7a), there's a constituent *cheap textbooks*, as we can tell by the fact that we can refer to this phrase by the single word *they*: *The students wondered how they could be obtained*. The relevant structure is shown in (7a'). But *cheap textbooks* isn't a constituent in (7b). Instead, *how cheap* forms a phrase in (7b), as you can see from (7b'). Here, *textbooks* is a separate constituent, which can again be replaced by *they*: *The students wondered how cheap they could be*. The examples in (7) show that we can't look at a string of words out of context and decide whether or not they form a constituent. We can find this out only when the string of words appears in a sentence, and when we can manipulate the sentence in various ways to discover its constituent structure. This requires a set of tests for constituency, like the pronoun test we used here: a pronoun such as *they* replaces a whole NP constituent.

5.1.2 Some syntactic tests for constituent structure

We have used the possessive *-s* suffix – which only attaches to NP constituents – as a morphological test for constituency. But to discover all the constituents of a sentence (and not just NPs) we also need syntactic tests. One syntactic test is seen at the end of the previous section: a constituent can often be replaced by a pronoun, but a random string of words cannot. We now go on to examine more syntactic tests.

5.1.2.1 *The sentence fragment test*

The first test in this section utilizes shortened answers to questions. If I ask *Who went on strike?*, a reasonable answer is *Black cab drivers*. Answers like these, which are not full sentences, are called SENTENCE FRAGMENTS, and they provide syntactic evidence about which words group together to form a constituent. A string of words that can be a sentence fragment must be a constituent. So here, *black cab drivers* is confirmed as a constituent of (1). Of course, it is still ambiguous, as its internal structure is not revealed. And if I ask *Who stayed to dinner?*, the answer is *The boy and the girl's uncle*, so this whole phrase is a constituent of (4), whichever internal

structure it has. Both of these particular sentence fragments remain ambiguous, because there is additional constituent structure inside each phrase.

However, this test can often be used to discover more about internal structure. On hearing (4), someone might try to resolve the confusion by asking *But whose uncle stayed to dinner?* A typical answer would be either (8a) or (8b), depending on which reading of the sentence you have in mind:

- (8) a. The boy and the girl's. (one person stays to dinner)
 b. The girl's. (two people stay to dinner)

In (8), the sentence fragment test confirms what we already discovered from (6): the whole sequence *the boy and the girl* is a constituent in the (a) reading, but in the (b) reading, *the girl* doesn't form a constituent with *the boy*. The fact that -s can be attached to either possible sequence in (8) confirms that they are both able to be used as constituents.

We can also use the sentence fragment test for constituent structure to show that in (6b), the sequence *the girl's uncle* is a constituent.¹ Keep in mind the reading where two people stay to dinner. If you didn't hear the speaker too clearly, you might ask *The boy and who stayed to dinner?* The answer is the sentence fragment *The girl's uncle*: this must therefore be a constituent. So we can bracket this phrase too, adding more information about the structure of (6b):

- (9) The boy and [[the girl's uncle] stayed. (two people)

As (9) shows, constituents are in turn built up of smaller constituents. Thus we confirm what we already saw in Chapter 4, namely that phrases contain smaller phrases, with each phrase having its own head and dependent elements. In (9), *uncle* is the head of the phrase *the girl's uncle*, since this phrase is 'about' the uncle.

The sentence fragment test is one of the formal tests for constituent structure. Using such tests, we can discover whether two apparently similar sentences in fact have different structures. Consider the examples in (10) and (11): both contain words of exactly the same syntactic categories or word classes, and in just the same order, as (12) shows (to remind you, D is the category 'determiner').

(10) Kim wrote that book with the blue cover.

(11) Kim bought that book with her first wages.

(12) N V D N P D A N

We might assume, then, that these sentences share a syntactic structure. However, native speakers feel instinctively that (10) and (11) are different; the sentences tend, for instance, to have a different intonation pattern. In (10), *with the blue cover* is a

1 As you can probably tell intuitively, however, the sequence *the girl's uncle* is not a constituent in (6a), where the uncle belongs to both the boy and the girl.

phrase (a PP) modifying the head noun *book* – *a book with a blue cover* is a type of book. So this PP belongs with *that book*, forming a constituent with it in (10). But in (11), the PP *with her first wages* tells us how she bought the book, and not anything about the book itself. So that PP modifies *bought* – it is an adjunct to *bought*. Crucially, the PP in (11) doesn't modify the noun *book*, and so doesn't form a constituent with it.

The sentence fragment test for constituent structure supports these intuitive feelings. In each case, when we ask a question, we get different sentence fragment answers:

- | | |
|--------------------------|---|
| (13) What did she write? | [That book with the blue cover] |
| (14) What did she buy? | a. [That book] |
| What did she buy? | b. * <u>That book with her first wages.</u> |

Example (13) shows that the entire sequence *that book with the blue cover* is a constituent of (10): it can be a sentence fragment. Example (14a) confirms that the sequence *that book* is a constituent of (11). And crucially, (14b) shows that the sequence *that book with her first wages* is not a constituent of (11): it can't be a sentence fragment (remember that the asterisk indicates an ungrammatical example). Contrasting grammatical and ungrammatical examples, as we have done here, is essential: you should use the ungrammatical examples to show that some sequence of words is *not* a constituent of the sentence you are working on.

- Please remember from now on that in the sentence fragment test, the question you ask should always be a *grammatical* one: the test is the answer itself.
- If a string of words from the original sentence can form a grammatical sentence fragment, it is likely that this sequence is a constituent of the original sentence.
- If the string of words is not grammatical as a sentence fragment, it most likely is not a constituent of the original sentence.
- Square brackets are used to show where a constituent begins and ends. Please do not put brackets round a phrase *unless* it is a constituent. You may find it useful to use a wavy underline for a string of words which is *not* a constituent, as I have done in (14b).

5.1.2.2 The echo question test

ECHO QUESTIONS are our next test for constituent structure. These questions are used in English when the speaker doesn't hear part of the sentence, or else is rather incredulous: e.g. *You saw **what**?* We use a *wh*-word (*what, which, who, when, why* and so on, and including *how*) or a *wh*-phrase (*You saw **which film**?*) to replace just the part of the sentence that we want repeated, otherwise 'echoing' the speaker's words. The *wh*-word or phrase doesn't replace a random string of words, but can only stand for a constituent of the sentence:

(15) *Kim wrote **what** with the blue cover?

(16) Kim bought **what** with her first wages?

The grammatical echo question in (16) is fine because *what* replaces *that book*, which is a constituent of (11). But (15) is ungrammatical because *that book* is only part of a larger constituent *that book with the blue cover* in (10); crucially, *that book* itself is not a constituent in this case, and so can't be replaced by a *wh*-word. (We can, however, echo just a head noun on its own: for example, in (10) we can replace *book* with a *wh*-word, giving *Kim lost that what with the blue cover?* The reason for this is that single words are also constituents.)

In fact, we've already used this test earlier: the echo question, *The boy and who stayed to dinner?*, only works when *who* replaces a constituent, such as *the girl's uncle*. So it's the question we'd ask if we were sure that two people stayed to dinner, and that one was the boy, but we weren't sure who the other person was.

To summarize: in the echo question test, a *wh*-word or phrase can replace a constituent; if the resulting question is ungrammatical, though, the string of words which you've replaced is probably not a constituent.

5.1.2.3 *The cleft test*

A further test for constituent structure confirms our findings: the two sentences in (10) and (11) have different structures. In the CLEFT construction illustrated in (17), the string of words in the 'focus' position must be a constituent. So in (17), we can focus on the whole sequence *that book with the blue cover*, showing that this is a constituent:

(17) It was [**that book with the blue cover**] that Kim wrote.

But in (18), the sequence *that book with her first wages* is not a constituent, and so can't occur in the focus position of a cleft sentence. This confirms what we saw in (14b):

(18) *It was that book with her first wages that Kim bought.

Remember that we only bracket a string of words which is a constituent, so we bracket *that book with the blue cover* in (17), but not *that book with her first wages* in (18).

Our original sentence in (11) does, however, contain other word sequences which will fit into the focus position of a cleft sentence. For instance, we can focus on either *that book* or *with her first wages*, showing that both these phrases are separate constituents of (11):

- (19) a. It was [**that book**] that Kim bought with her first wages.
 b. It was [**with her first wages**] that Kim bought that book.

Putting together the information from all three tests, we can show what we've learnt so far about the constituent structure of (10) and (11) by using brackets, as follows:

(20) [Kim wrote [that book with the blue cover]].

(21) [Kim bought [that book] [with her first wages]].

The whole sentence is also in brackets in each case, since both examples occur as independent sentences, and are therefore constituents – if sentence fragments are constituents, then it's not surprising that whole sentences are also constituents. Although there are other constituents in each sentence, the brackets in (20) and (21) show as much information as we have up to now.

Of course, for our constituency tests to have real significance, we must be able to apply them to languages other than English, although not all tests apply equally well in all languages, because certain syntactic constructions may be absent. Cleft constructions occur widely; in the Irish examples that follow, (22a) shows the basic word order, and (22b) is a cleft construction with the noun phrase *an fear* 'the man' in the focus position:

- (22) a. Bhí an fear ag péinteáil cathaoir. (Irish)
 was the man PROG paint chair
 'The man was painting a chair.'
- b. Is é [an fear] a bhí ag péinteáil cathaoir.
 is it the man who was PROG paint chair
 'It's *the man* who was painting a chair.'

Similarly in the next examples, from Lekeitio Basque, (23a) has basic word order, while (23b) is a cleft construction, with focus on the fronted noun phrase *orreri mutillari* 'that boy' (the DATIVE case marking on this NP does the work of the preposition 'to' in English, showing the boy as the recipient):

- (23) a. premižúa orreri mutillari emon-dótze. (Basque)
 prize that.DATIVE boy.DATIVE give-AUX
 'They have given the prize to that boy.'
- b. [orreri mutillari] da premižúa emón dotzé-na.
 that.DATIVE boy.DATIVE is prize give AUX-that
 'It's *to that boy* that they have given the prize.'

From (23b), we can tell that *orreri mutillari* is a constituent of (23a).

5.1.2.4 Displacement and dependency

The constructions in Section 5.1.2 all illustrate an important property of human language: the ability to DISPLACE or MOVE a phrase from its basic position. The hallmark of such displacement is that a phrase is understood semantically as if it were in one position in the clause, but occurs physically (syntactically and audibly) in a different position in the clause. We can illustrate using the cleft examples seen earlier:

- (24) It was [_{NP} that book] that Kim bought ____ with her first wages.
- (25) It was [_{PP} with her first wages] that Kim bought that book ____ .

The gap in these examples shows the ‘original’ position of the displaced phrases. In other words, when you hear an example like (24), you understand it as if the displaced NP *that book* were still in its normal linear position, following the verb *bought*. Importantly, the verb *bought* only has one direct object, and that syntactic fact does not change just because the object is displaced from its basic position in the usual constituent order. The same applies to (25): the displaced PP is understood as if it were in the typical adjunct position, following the direct object.

The displacement of a phrase sets up a DEPENDENCY between the displaced phrase and the ‘empty’ position associated with it: the displaced element provides the physical words we need, but its basic position specifies its syntactic role, for instance the role of ‘direct object’ in (24). The displaced element and the associated gap are of course one and the same entity – moving the object *that book* does not alter the argument structure of the verb *bought*.

It is likely that all languages have instances of displacement of one kind or another. We will see other examples as we go along.

5.1.2.5 Summary

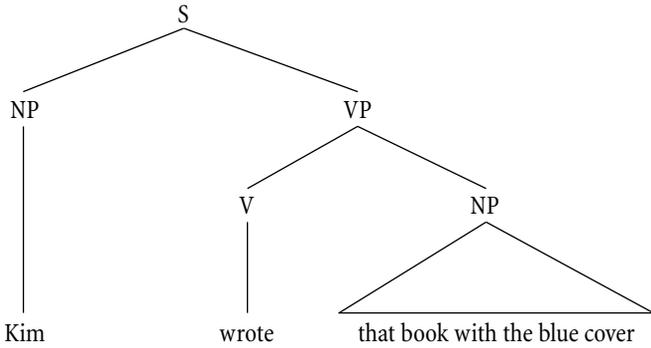
Each of the tests for constituent structure in Section 5.1.2 works by harnessing the intuitions of native speakers of a language. The fact that speakers share GRAMMATICALITY JUDGEMENTS – intuitions about which sentences are possible and which aren’t – shows that we have an unconscious knowledge of the word groupings in a sentence. The tests for constituent structure are just particular syntactic environments which can be filled only by constituents. Whenever we put a string of words that isn’t a constituent into one of these environments, the result sounds impossible to native speakers. This UNGRAMMATICALITY (the technical term for such results) tells us that in such cases, the string of words isn’t a constituent.

So far in Section 5.1, I have introduced these syntactic tests for constituent structure: (1) replacement by a pronoun; (2) sentence fragments; (3) echo questions; and (4) cleft sentences. The information about constituent structure which results from our tests can be represented by using square brackets to mark off the constituents, as I have done so far, or alternatively by using tree diagrams. We turn next to this topic.

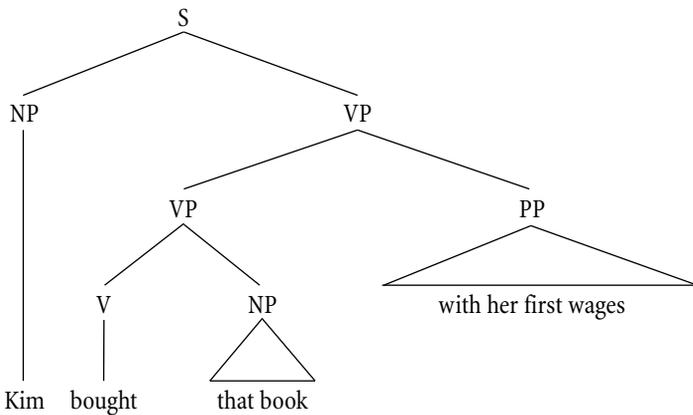
5.1.3 Introduction to constituent structure trees

In (26) and (27), I represent the structure of our two sentences in (10) and (11) by using TREE DIAGRAMS. As you can see, these are upside-down trees, with the root at the top, and branches descending from that root. The root of the tree is labelled ‘S’ for ‘Sentence’, and the clause is divided into two main branches, the subject and the predicate, as discussed in Section 3.1.1.

(26)



(27)



Let's now go through these tree structures. In each case, there is a major split between the subject NP on the left branch and the predicate VP on the right branch. Both trees have a subject, *Kim*. The difference in structure in our two sentences lies within the VP, as we discovered from our tests on constituent structure, and this difference is reflected in the form of the two trees.

In (26), I have suggested that the verb and its object together form a verb phrase (VP). The VP has two branches, V (for verb) containing just the transitive verb *wrote*, and NP, the noun phrase which is the complement of *wrote*. This sequence, *that book with the blue cover*, is shown as a triangle, which indicates that the whole sequence forms a constituent. That doesn't mean that there is no more internal structure within that NP, just that so far, this is all we've discovered.

In (27), we again have a VP consisting of the transitive verb *bought* plus its complement, the object NP *that book*. However, we also have an ADJUNCT here, namely the PP *with her first wages*. Recall from Chapter 4 that an adjunct is a constituent which is syntactically optional, in other words not required in order to make the sentence grammatical: adjuncts are not arguments of the verb, and are therefore non-essential constituents. The structure that I've suggested for (27) reflects this by showing that if we add an adjunct to the VP, we don't get a different

kind of phrase – it's still a verb phrase, but just one that contains more information. The structure is RECURSIVE, in that it has a VP within a larger VP.

In representing VPs in each tree, I have actually shown more structure than I gave in the brackets for each sentence in (20) and (21) – those examples did not include a set of brackets round the verb and its dependents. So we ought to check that the VP really is a constituent in each case. We can do this by using a different test for constituency: the *do so* test. A VP can be replaced by *do so* (or *did so* in the past tense) as follows:

- (28) I thought that Mel [_{VP} wrote that book with the blue cover].
No! Kim [_{VP} *did so*].
- (29) I thought that Mel [_{VP} bought that book with her first wages]
No! Kim [_{VP} *did so*].

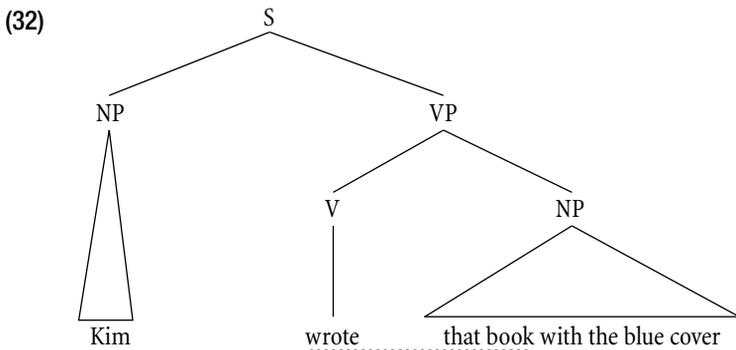
The *do so* test works by replacing the entire VP with something that stands for it, and it only works if the sequence being replaced really is a constituent. In (29), I've replaced the whole larger VP *bought that book with her first wages* with *did so*. But note that (27) also contains a smaller VP, *bought that book*. If our test is to have any validity, this should also be replaceable by *did so*. And indeed it is:

- (30) I thought that Mel [_{VP} bought that book] (with some of her inheritance).
No! Kim [_{VP} *did so*] with her first wages.

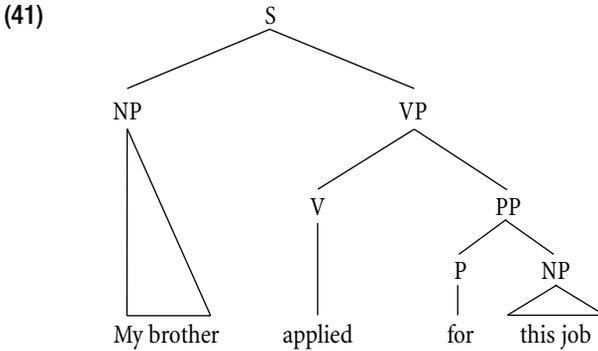
We can also use the *do so* test to confirm that the sequence *wrote that book* on its own does not form a VP constituent in (26). Once again, the contrast in grammaticality demonstrates the difference in structure between the two examples:

- (31) I thought that Mel wrote that book with the blue cover.
No! *Kim did so with the blue cover.

The reason that *wrote that book* does not act like a VP here is because the sequence *that book* is itself not a constituent in this case, but rather is part of the NP *that book with the blue cover*, as we saw in Section 5.1.2. This whole NP is the object of *wrote*, so we can't take part of it and leave the rest behind. The underlining should help you see that *wrote that book* is not a constituent here:



this job is also a constituent in this example: *It was [this job] that my brother applied for*. Now that we know that *this job* is a constituent in this case, we can draw a more detailed tree to represent this: (41) gives more information about the structure than (39) does.



Tree (41) isn't a replacement for (39): it simply gives more information. Both trees would be used by linguists, depending on the level of detail we want to indicate. Tree (41) shows that *this job* is an NP constituent, nested inside a larger constituent *for this job*. If we are using brackets, one constituent is nested inside the other as follows: [*for* [*this job*]].

So, to summarize, exactly how much or how little structure we actually show within the tree diagram or the brackets depends on what we are trying to show. But if we are claiming that two sentences contrast in constituent structure, then the parts that differ must be shown in enough detail to make our claims clear.

5.1.4 Summary

We have so far used the following syntactic tests for constituency: replacement by a pronoun, the echo question test, the sentence fragment test, the cleft test, and the *do so* test for VP status. We indicate which strings of words are constituents of a sentence in two ways: either by placing square brackets round the constituents, or by using tree diagrams. Most importantly, this section shows that we must use contrasting sets of grammatical and ungrammatical examples to argue for a particular constituent structure. Our analysis is valid only if we can show that it also rules out other logically possible analyses. So, as well as using the tests to show what the constituent structure of a phrase or sentence actually is, we use them to rule out any alternative structures.

5.2 RELATIONSHIPS WITHIN THE TREE

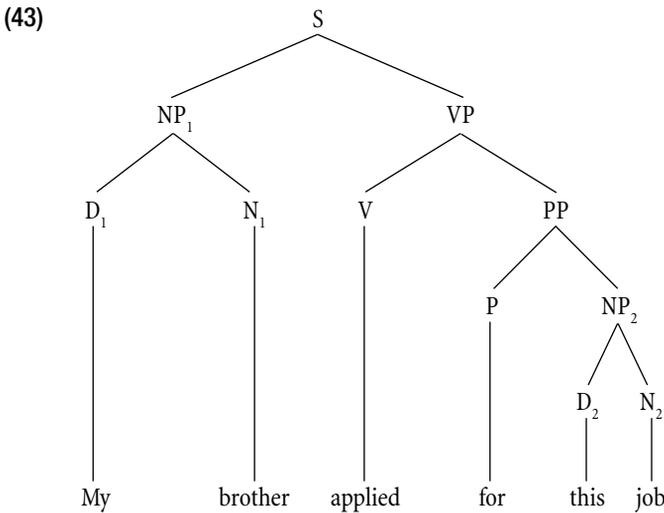
This short section defines the technical terms used by linguists to discuss relationships between words and phrases in a tree diagram. It's common to use LABELLED

BRACKETS or LABELLED TREE DIAGRAMS in which each relevant constituent has a label showing its category. Our trees include word class and phrase class labels such as V, VP, PP, P, NP, and so on, telling us that what's beneath that label is a PP, or a P, or an NP etc. This exact same information can be shown in labelled brackets. For example, the PP *for this job*, which we proved to be a constituent in (37), can be shown as follows:

(42) [_{PP} [_P for] [_{NP} this job]]

These brackets are read like this: the whole constituent is a PP, since this is the label on the outermost brackets (by convention, only the left-hand bracket is labelled). The PP comprises two main constituents, a preposition *for* and an NP *this job*: as we saw, this noun phrase fits into the focus position of a cleft sentence, so must be a constituent. Each individual lexical item (word) is in fact also a constituent, so *for*, *this* and *job* here are constituents, although I haven't labelled or bracketed the last two items here. The words are the smallest constituents of a tree.

Let's now add more information into (41) to give a fully detailed tree diagram, showing *my* and *this* as D (determiners) and *brother* and *job* as Ns:



Using the tree in (43), I now introduce some of the technical terms used in syntax to describe tree structures. Recall that all the lines in the tree are known by the (reasonable!) technical term BRANCHES. Each point in the tree that has a category label or else an actual word attached to it is known as a NODE. In (43) we find PHRASAL nodes NP, VP and PP (nodes denoting the phrases in the tree), and also S. Despite the fact that I've used the label 'S', rather than a 'something phrase', the sentence is, of course, a phrase in its own right, and some linguists reflect this by terming the sentence 'TP', for tense phrase ('tense' in the sense of the tense of the verb). The idea is that a sentence is a phrase that denotes tense, although, as

a new test: ellipsis. Section 5.3.2 introduces another diagnostic for constituency: co-ordination. And Section 5.3.3 considers whether all languages have the same constituents.

5.3.1 Verb classes and constituent structure tests

5.3.1.1 Phrasal verbs and prepositional verbs

Before turning to some detailed tree diagrams which represent sentences of English, I first discuss the differences between two verb classes: phrasal verbs and prepositional verbs. Let's examine two more sentences that may appear superficially similar, but which, in fact, have different constituent structures:

- (45) Those smugglers shook off their pursuers.
 (46) Those smugglers relied on the weather forecast.

There is one clear indication that these two examples are syntactically distinct. In (45), we can take the preposition *off* and place it immediately after the direct object NP *their pursuers*, to give *Those smugglers shook their pursuers off*. The verb *shake off* is a transitive verb – it must have an NP complement, i.e. a direct object. In (46), we don't have a transitive verb, but instead we have a verb *rely* and its PP complement. The verb does not have a direct object at all, and if we attempt to put the preposition *on* after the NP, the result is ungrammatical:

- (47) *Those smugglers relied the weather forecast on.

This test identifies very successfully one particular verb class: transitive PHRASAL VERBS such as *shake off* always allow the preposition to be placed after their object NP. These verbs are single lexical items comprising a V and a P: [_V shake off]. We have, in fact, already met another example of a transitive phrasal verb in (33), namely *write down*, as we can confirm from the fact that we can get *My brother wrote his address down*. English has a vast number of phrasal verbs, both transitive and intransitive. Further examples of the transitive kind include *turn over*, *pull down*, *pick up*, *put out*, *switch on* and *break off*. As (48) shows, the preposition can follow the direct object (in bold) in each case. (For some of these, you may prefer to leave the preposition next to the verb. The point I am making is simply that it may follow the direct object.)

- (48) a. We turned **the place** over.
 b. They pulled **that old farm building with a thatched roof** down last week.
 c. I'd pick **that snake** up.
 d. She broke **her last engagement** off very suddenly.

This test also allows you to identify the full extent of the direct object NP, because the preposition has to be placed immediately *after* that NP (and not in the middle of it).

So for instance, in (48b), the test shows that the whole of the sequence *that old farm building with a thatched roof* comprises the direct object. The preposition *down* can be placed at the end of the direct object, but not elsewhere:

- (49) *They pulled that old farm building **down** with a thatched roof last week.

As noted, English also has phrasal verbs which are – or can be – intransitive, such as *wake up*, *sit down*, *sleep in*, *turn out* (as in *Not many people turned out*), and *break down* (as in *The car broke down*).

Some English grammarians use the term ‘particle’ to refer to the *over*, *down*, *up*, *off*, *out* (etc.) part of the phrasal verb, but we can tell that they are truly prepositions by using the modifier *right*, which we saw in Section 2.6.1 to be a good test for preposition status. So for example, we get *Pull the handle **right** down*, *Break the plastic safety catch **right** off*, *There was a loud bang and I woke **right** up* and so on.

Now let’s compare (46). There, we don’t have a phrasal verb at all. Instead, the verb *rely* takes a PP complement, and this PP must be headed by the preposition *on*: we can only *rely **on*** something, not **rely for*, **rely off*, **rely over* or **rely out*. Verbs that select PP complements are known as PREPOSITIONAL VERBS. Their defining properties are that the PP is obligatory, and is headed by one specific preposition. Further examples of prepositional verbs include *believe in NP*, *hear from NP*, *see to NP*, *glance at NP*, *hope for NP*, *depend on NP* and *look after NP*, among many others. Quite often, the preposition has such a close relationship with the prepositional verb that not even one of the prepositional modifiers, such as *right*, *just* or *straight*, can intervene. These examples give you an idea of the variation that is found; of course, you may not agree with my judgements in each case:

- (50) We rely just/*right on our good fortune.
 They skated right/*just over these damaging issues.
 We looked *just/*right after the children.

Prepositional verbs, then, are a rather special set. On the other hand, if a verb merely has a PP adjunct – in other words, it is modified by an optional PP – the properties are entirely different. The head preposition can be readily changed: *I jumped **on** the wall/off the wall/over the wall/behind the wall* (and so on), and the PP can be omitted entirely, as it is not a complement. The choice of a modifier in the PP is also much freer:

- (51) I jumped just/straight/right over the wall.
 We ran just/straight/right to the end of the beach.
 The vase fell straight/right off the shelf.

The verbs illustrated in (51) are not prepositional, since the PP is an adjunct rather than a complement. We reserve the term ‘prepositional verb’ for a verb with an obligatory PP complement.

A tree is built up from our evidence of what these phrases are, which comes in the form of tests for constituent structure. So let's start with evidence for the two main constituents in each tree: the subject and the VP predicate.

Our tests show that *those smugglers*, the subject in each sentence, is indeed a constituent. Both subjects can be sentence fragments:

- (54) a. Who shook off their pursuers? [_{NP} Those smugglers]
 b. Who relied on the weather forecast? [_{NP} Those smugglers]

Second, both subject NPs can also appear in the focus position of a cleft sentence:

- (55) a. It's [_{NP} those smugglers] who shook off their pursuers.
 b. It's [_{NP} those smugglers] who relied on the weather forecast.

And third, we can also replace both subject NPs with *they*, using a test for NP status introduced in Section 5.1.1. The word *they* is rather badly termed a 'pronoun'. Since it replaces a whole NP it's really a pro-NP: 'pro' means '(stands) for'. The cover term used for all pro-phrases is PROFORM: a proform takes the place of a sequence of words which form a constituent, and so any string of words that can be replaced by an appropriate proform must be a constituent.

A proform test can also prove the existence of the VP constituent, as we saw in Section 5.1.3. We use *do so* (or *did so* in the past tense) to stand for VP, therefore as a 'pro-VP':

- (56) a. Those smugglers [_{VP} shook off their pursuers], and the moonshine
 merchants [_{VP} **did so**] too.
 b. Those smugglers [_{VP} relied on the weather forecast], and these fishermen
 [_{VP} **did so**] too.

Rather than repeating the whole VP, we can replace it with the proform. The *do so* test is a specific test for a VP constituent.

One of the other tests for constituent structure which was given earlier is the cleft construction. However, most dialects of English can't form a cleft using a VP constituent:

- (57) *It's **shake off their pursuers** that those smugglers did.
 *It's **rely on the weather forecast** that those smugglers did.

This does not mean the cleft test is unreliable; it just means that VPs can't be focused like this in English. In some languages, though, such as Irish, it's perfectly OK to focus VPs in the cleft construction. Given a sentence like (58), we can focus the VP to get (59), which is fully grammatical in Irish:

- (58) Bhí an fear ag péinteáil cathaoir. (Irish)
 was the man PROG paint chair
 'The man was painting a chair.'

(63) Sentence fragments

- a. Who did the smugglers shake off? [_{NP} **Their pursuers**]
 b. What did the smugglers rely on? [_{NP} **The weather forecast**]

(64) Clefts

- a. It was [_{NP} **their pursuers**] that the smugglers shook off.
 b. It was [_{NP} **the weather forecast**] which the smugglers relied on.

(65) Proforms

- a. The smugglers shook [_{NP} **them**] off.
 b. The smugglers relied on [_{NP} **it**].

Note also that the pronoun *precedes* the preposition when we have a transitive phrasal verb, as in (65a), but *follows* the preposition when we have a prepositional verb, as in (65b). This is the way each verb class always works in English, and it is a very reliable test. For instance, we can't have **The smugglers shook off them* for a transitive phrasal verb.

So if the NPs which we've tested from (52) and (53) really are constituents, as we've shown, why can't they be omitted in (61)? The reason is that both examples contain a head word which requires these NPs to be present – the NPs are complements, and the sentences are incomplete without these complements. So the transitive phrasal verb *shake off* requires a direct object NP in (61a), and in (61b) the transitive preposition *on* also requires an NP object. Constituents which form the complement to some head, particularly a head verb or preposition, are quite generally unable to be omitted. This means that the ellipsis test can't be used to diagnose the constituent status of such phrases.

Now let's examine the sequence *on the weather forecast*, which is shown as a PP in (53). First we need to confirm the constituent status of this string of words:

(66) Sentence fragment

What did the smugglers rely on? [_{pp} **On the weather forecast**]

(67) Cleft

It was [_{pp} **on the weather forecast**] that the smugglers usually relied.

(Some speakers may not be entirely happy with (67), but the test in (66) confirms that there really is a PP.)

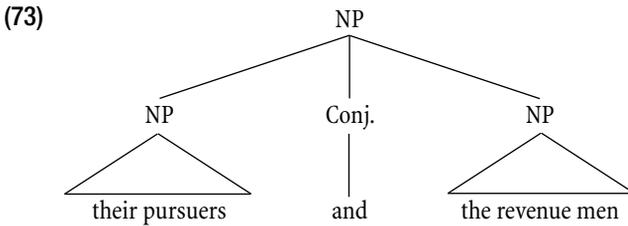
Again, we might expect that if it's a constituent, then the PP could be omitted, but it actually can't be:

- (68)** *Those smugglers didn't rely [_{pp} *on the weather forecast*], but these fishermen did rely [_{pp} *___*] for sure.

Just as with (61), the reason for the ungrammaticality of (68) is that the PP is a complement: prepositional verbs like *rely on NP* require the PP complement to be present, so again, we can't use the ellipsis test for constituent structure in a case like this.

The two NPs in brackets in these examples have been conjoined using *and*, known as a CO-ORDINATING CONJUNCTION. Other such conjunctions in English include *but*, *nor* and *or*.

You may be wondering if the transitive phrasal verb *shook off* in (71) and the preposition *on* in (72) now have two object NPs. No, they do not: when two constituents of the same category are conjoined, they simply make one larger constituent of the same category, as in (73). So there is still only one object for the transitive verb in (71) and the preposition *on* in (72), but this NP may itself contain NPs embedded within it. The node label CONJ means ‘conjunction’.



The co-ordination test can be used to confirm that a phrasal verb and a prepositional verb do have different structures, as we have proposed – look at your tree diagrams for (52) and (53). A prepositional verb contains a PP constituent, according to the tests we’ve seen so far. And indeed, the PP can be conjoined with another PP:

- (74) The smugglers relied [_{pp} on the weather forecast] and (also) [_{pp} on their years of experience].

In (75), on the other hand, we can’t conjoin *off their pursuers* with *off the revenue men*, because these two strings of words are not constituents of any kind:

- (75) *The smugglers shook off their pursuers and off the revenue men.

The preposition of a phrasal verb like *shake off* isn’t attached to the following NP, as we can see from the structure for phrasal verbs in (52): there is no node that dominates just the preposition and *their pursuers*, so this sequence doesn’t form a constituent. Please look at your tree diagram for (52) to confirm this for yourself.

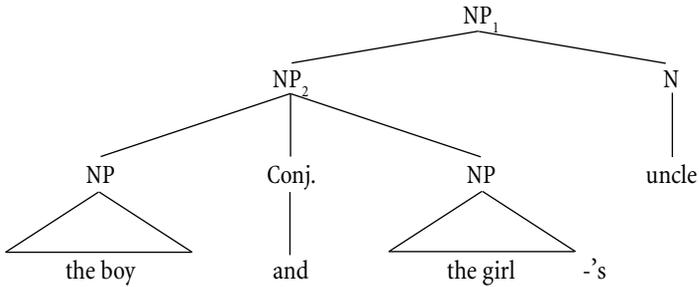
Finally, we can use the co-ordination test to discover more about the structure of one of the ambiguous sentences in Section 5.1: *The boy and the girl’s uncle stayed to dinner*. You’ll see that there’s a conjunction *and* in this sentence. We can now show that there are two possible ways of co-ordinating constituents within the NP *the boy and the girl’s uncle*, which accounts for the ambiguity. First, we can conjoin an NP *the boy* with an NP *the girl’s uncle*, as in (76). This gives the reading in which two people stay to dinner:

- (76) [_{NP} [_{NP} the boy] and [_{NP} the girl’s uncle]]

In (76), the outermost brackets are labelled 'NP': this tells you the category of the whole phrase. Within this large NP, two smaller NPs are embedded, co-ordinated using *and*. The equivalent tree is in fact just the same as that in (73).

In the alternative reading of the phrase, where only one person stays to dinner, we conjoin an NP *the boy* with an NP *the girl*, as in (77): here, it's the uncle to both children who stays to dinner.

(77)



The tree in (77) says that the whole phrase is an NP (NP₁) which has two branches. On the left branch is NP₂, which immediately dominates the two conjoined NPs. Note that this whole phrase in NP₂ effectively replaces a single-word determiner such as *their*. On the right branch is the N *uncle*, the head noun of the *entire* phrase, NP₁.

We've seen in this section that two strings of words can be conjoined if they're constituents, and (normally) of the same syntactic category. Conversely, if a sequence of words which does *not* form a constituent is conjoined with other material, then the result is always ungrammatical, just as in (75). Co-ordination can therefore be added to our set of tests for constituent structure.

5.3.3 Do all languages have the same constituents?

The answer to this question is no, they apparently don't. I illustrate this with VP. Most languages have a clear VP constituent, as can be shown, for example, using VP co-ordination. Examples (78) and (79) show conjoined VPs in Persian and in Malagasy (note that, in Malagasy, the subject – *Rabe*, a name – is at the end of the clause, rather than at the start):

(78) Jân [vp xandid] va [vp dast tekân dâd]. (Persian)
 John smiled and hand sign gave
 'John smiled and waved.'

(79) [vp Misotro taoka] sy [vp mihinam-bary] Rabe. (Malagasy)
 drink alcohol and eat-rice Rabe
 'Rabe is drinking alcohol and eating rice.'

tree. In Section 5.3, we used the tests for constituency to work out the structure of a number of phrases and sentences, and produced labelled brackets and labelled tree diagrams to illustrate the constituent structure that we discovered. The syntactic tests for constituency used in this chapter are as follows: the sentence fragment test; echo questions; cleft sentences; proform tests; ellipsis; and co-ordination.

FURTHER READING

Two introductory texts which concentrate on the grammar and structure of English, and which go significantly beyond what I have done in Chapters 2 through 5, are Börjars and Burridge (2010) and Lobeck (2000). Radford (1988) provides detailed (and relatively introductory) reading on constituent structure, tree diagrams and tests for constituency. At this stage, you may not want to go beyond his Chapter 5. See also Aarts (2008) and Burton-Roberts (2010). For a more detailed account of how to treat the English possessive *-s*, including proposed tree diagrams, I recommend consulting Burton-Roberts (2010).

EXERCISES

1. The two sentences in (1) and (2) below contain words from the same word classes, and in the same order, but they each have different syntactic structures.

Task: (i) Using standard tests for constituency, work out what the constituents of each sentence must be. You should use at least two tests for each putative constituent. Your answers should include contrasting grammatical and ungrammatical examples which reveal the syntactic differences between (1) and (2). Use square brackets to indicate the constituents you find in each example, and remember to bracket constituents only, and not random strings of words; (ii) Next, draw labelled tree diagrams for (1) and (2), taking care that the trees correctly represent the constituent structures you discovered earlier.

- (1) Kim glanced at the actor with a wig.
- (2) Kim glanced at the actor through her binoculars.

2. The two sentences in (1) and (2) below again contain words from the same word classes, and in the same order, but again they each have different syntactic structures.

Task: (i) Using at least three standard tests for constituency, work out what the constituents of each sentence must be. Your answers should include contrasting grammatical and ungrammatical examples which reveal the syntactic differences between (1) and (2). Can you provide any further evidence that (1) and (2) differ in structure? What subclasses of verb does each example contain? (Don't worry if you prefer *stank* in (1); both past tense forms are acceptable!) (ii) Now draw

labelled tree diagrams for (1) and (2), ensuring that the trees correctly represent the constituent structures you discovered earlier:

- (1) The skunk stunk out my garden.
 (2) The skunk slunk out my garden.

3. The data that follow are from Fijian, and are taken from Lynch (1998).

Task: Examine the data and decide what they show about the grammar of possession in Fijian. Describe carefully how the possessive construction is formed. Your answer should account for all the data.

Hint

- The prefix glossed as POSS is a possessive marker. It has three distinct forms, depending on the category of the item possessed. Your answer should note all three forms. However, you are not required to specify what factors determine the appearance of any particular form.

- | | |
|--|--|
| (1) na tina-qu
the mother-my
'my mother' | (2) na me-na niu
the POSS-his coconut
'his coconut' |
| (3) na ke-mu itaba
the POSS-your photo
'your photo'
(i.e. a photo taken of you) | (4) na no-mu itaba
the POSS-your photo
'your photo'
(i.e. a photo you took or have) |
| (5) na yaca-qu
the name-my
'my name' | (6) na ke-mu madrai
the POSS-your bread
'your bread' |
| (7) na me-qu bia
the POSS-my beer
'my beer' | (8) ma ulu-qu
the head-my
'my head' |

4. Welsh is a VERB-INITIAL language: the finite verb or finite auxiliary appears first in the clause in unmarked (= normal) constituent order, as in (1), which has a finite auxiliary (meaning 'was') in initial position. This example also has a non-finite main verb *dweud*, 'tell', lower down in the clause; this construction, then, is rather parallel to English *was telling*:

- (1) Oedd fy ffrind yn dweud ei hanes wrth yr athro y bore 'ma.
 was my friend PROG tell.INFIN her story to the teacher the morning here
 'My friend was telling her story to the teacher this morning.'

Changes in the basic order are used to focus other constituents. Examine the data in (2) through (6) (based loosely on Jones and Thomas 1977: 289).

Task: (i) State how constituents are focused in Welsh (being as precise as you can), and (ii) state precisely what constituent is focused in each example, giving its category. (iii) A translation for (2) is already provided; suggest appropriate translations into English for the remaining examples. (iv) In both (3) and (5) there are additional grammatical changes; can you say what these are?

Hint

- In some cases, your translation into English may not sound very natural. The reason for this is that languages display differences in what constituents may be focused or otherwise manipulated, as we saw in Chapter 5. Provide the best translations you can, while trying to capture the meaning of the source language.

(2) Fy ffrind oedd yn dweud ei hanes wrth yr athro y bore 'ma.
my friend was PROG tell.INFIN her story to the teacher the morning here
'It was *my friend* who was telling her story to the teacher this morning.'

(3) Ei hanes oedd fy ffrind yn ddweud wrth yr athro
her story was my friend PROG it.tell.INFIN to the teacher
y bore 'ma.
the morning here

(4) Wrth yr athro oedd fy ffrind yn dweud ei hanes
to the teacher was my friend PROG tell.INFIN her story
y bore 'ma.
the morning here

(5) Dweud ei hanes wrth yr athro oedd fy ffrind y bore 'ma.
tell.INFIN her story to the teacher was my friend the morning here

(6) Y bore 'ma oedd fy ffrind yn dweud ei hanes wrth yr athro.
the morning here was my friend PROG tell.INFIN her story to the teacher

5. Examine the data in (1) through (6) from Malayalam, a Dravidian language spoken in India. These data (taken from Asher and Kumari 1997) all illustrate one particular construction which manipulates constituents in a certain way which was discussed in Chapter 5. However, I have left one crucial morpheme (part of a word) in the source language both unidentified and un glossed.

Task: (i) Identify what construction is illustrated in the data; (ii) work out *exactly* how this construction is formed in Malayalam; and (iii) work out what syntactic category of constituent (e.g. PP, NP etc.) is being manipulated in each separate Malayalam example. (To remind you, ACC is accusative case, indicating a direct object.)

(1) avan bhaaryayootum makka|ootum kuute taamasikkunnu
he wife.with children.with together.with stay.PRES
'He stays with his wife and children.'

- (2) naan raamaneyum avançe muunnaamate makaneyum kaṇṭu
 I Raman.ACC his third son.ACC see.PAST
 'I saw Raman and his third son.'
- (3) aṽ viiṭṭilum hoostṭalilum taamasikilla
 she house.in hostel.in stay.FUT.NEG
 'She will not stay in the house or the hostel.'
- (4) avan ezuttu v.ṭṭiyaayum vyaktamaayum ezuti
 he letter neatly legibly write.PAST
 'He wrote the letter neatly and legibly.'
- (5) avan kaappi kuṭikkukayum pinṇe vaṭa tiṇṇukayum ceytu
 he coffee drink.INFINITIVE and.then vada eat.INFINITIVE do.PAST
 'He drank coffee and then ate vada.'
- (6) uṇṇiyum baabuvum vanna
 Unni Babu come.PAST
 'Unni and Babu came.'

6. The data in (1) and (2) (from Clamons *et al.* 1999) are from a Cushitic language, Oromo, spoken in Ethiopia, Kenya and Tanzania. In the examples in (1), the subject of the sentence also has a special pragmatic property: it is a TOPIC, which Clamons *et al.* define as 'what the sentence or discourse is about'. The topic property is marked on subjects which are topics using a topic marker *-n*. The subjects in (2) are *not* topics. Subjects in general are marked with a 'subject case' marker (Su).

Task: Discover what grammatical change occurs in the sentence when its subject is also a topic. Articulate it as clearly as you can, using the correct grammatical terminology, and generalizing so that you cover all the data with a single statement. You are looking for a property which is common to all the grammatical data in (1) versus all the grammatical data in (2), but this property is manifested slightly differently from example to example.

Hints

- The subject of the sentence can, of course, consist of one noun phrase conjoined with another noun phrase: for instance, in (1b) the 'girl' noun phrase and the 'boy' noun phrase are co-ordinated in this way to form a subject meaning 'the girl and the boy'. The subject will then have the grammatical properties of the two conjoined phrases together.
- English does not have a special topic construction, but topics are typically associated with a particular emphatic intonation. I've tried to indicate this by using italics in the translations in (1).
- The background information above the examples is there purely to help you see where topics are used in Oromo. You can see from (1) that topics are normally a phrase which has already been mentioned in the discourse; this contrasts with the 'out-of-the-blue' sentence that you might find at the start of

a story, as in (2a), or a sentence in which the participants clearly haven't been mentioned before, as in (2b, c). Note also that the translations in (1) and (2) differ subtly, reflecting the fact that the subjects are topics in (1) but not in (2).

- It will help you to compare (1a) with (2a), (1b) with (2b), and (1c) with (2c). The additional data are there simply to give you more clues.

- (1) a. (i). (In answer to: What did the girl and the boy do?)
 Intal-t-ii-n hoolaa bit-t-e.
 girl-FEM-SU-TOPIC sheep buy-FEM-PAST
 'The girl bought a sheep.'
- (ii). *Intal-t-ii-n hoolaa bit-e.
 girl-FEM-SU-TOPIC sheep buy-PAST
 ('The girl bought a sheep.')
- b. (In answer to: What did the girl and the boy do?)
 Intal-t-ii-n -ifi gurbaá-n wal lol-an
 girl-FEM-SU-TOPIC and boy.SU-TOPIC each.other fight-3PL.PAST
 'The girl and the boy were fighting.'
- c. (In answer to: Where was I when the boy came?)
 Ati -ifi Salma-á-n nyataa godhu tur-tan
 you.SG and Salma-SU-TOPIC food make were-2PL.PAST
 'You and Salma were cooking.'
- (2) a. (Passage at the start of a story, i.e. with no previous context)
 Intala takka-á hoolaa bit-e
 girl one.FEM-SU sheep buy-PAST
 'A girl bought a sheep.'
- b. (In answer to: Who was fighting?)
 Intala -afi gurbaa tokko-ó wal lol-e
 girl -and boy one.MASC-SU each.other fight-PAST
 'Some girl and boy were fighting.'
- c. (In answer to: Who was cooking?)
 Ati -ifi Salma-á nyataa godhu tur-e
 you.SG and Salma-SU food make were-PAST
 'You and Salma were cooking.'
- d. (i). Intala-á dhuf-e.
 girl-SU come-PAST
 'The girl came.'
- (ii). *Intala-á dhuf-t-e.
 girl-SU come-FEM-PAST
 ('The girl came.')

Relationships within the clause

After introducing the ways in which languages can indicate grammatical relations within the clause (Section 6.1), this chapter outlines in detail the major systems: constituent order (6.2); case systems (6.3); and agreement and cross-referencing (6.4). Section 6.5 looks at grammatical relations cross-linguistically, and asks whether they are universal. Section 6.6 is a case study of languages with ‘free’ word order, based on Warlpiri.

6.1 INDICATING GRAMMATICAL RELATIONS IN THE CLAUSE

In this chapter, I investigate the relationships between verbs and their noun phrase arguments within the clause. All languages have intransitive clauses – clauses with a verb and just one NP participant, such as *The dog growled* – and transitive clauses – clauses with a verb and two NP participants, such as *The dog bit my friend*. The NP participants that occur in these basic clause types are known as CORE arguments, and this chapter examines the ways in which the world’s languages distinguish between core arguments.

There are three main ways in which a language may indicate the relationship between core NPs and the verbal predicate. First, each core NP may have a fixed position in the clause: such a system uses CONSTITUENT ORDER to indicate the relationship between NP and verb. In English, both subjects and objects have a fixed position, which is how we determine who killed who in a pair of sentences like *The snake killed the bird* and *The bird killed the snake*.

But core NPs don’t have a fixed position in all languages. Core NPs in Latin can appear quite easily in different positions; both sentences in (1) have the same meaning, although the order of the NPs is different in (1a) and (1b):

- (1) a. Puer-um puella audi-t. (Latin)
 boy-ACC girl.NOM hear-PRES.3SG
 ‘The girl hears the boy.’
 b. Puella puer-um audi-t.
 girl.NOM boy-ACC hear-PRES.3SG
 ‘The girl hears the boy.’

This variation in constituent order is possible because in Latin, the form of the NPs themselves indicates what relationship they have with the verb: this is CASE MARKING. The NOMINATIVE NP (glossed NOM) signifies the subject and the ACCUSATIVE NP (glossed

ACC) signifies the object of the verb. Nominative and accusative are grammatical terms for distinct cases. Latin, then, utilizes the second main way of distinguishing core NPs: by case-marking. Subjects in Latin are not distinguished from objects by their position, but by being specifically marked as subjects or objects.

The third way in which a language can indicate the relationship between NP participants and the predicate is by verb AGREEMENT or CROSS-REFERENCING. Latin and English have a limited amount of agreement: the *-t* suffix on *audit* and the *-s* suffix on *hears* both indicate a third person singular subject (in English, this occurs only in the present tense). But many languages have far more extensive systems to indicate the participants via marking on the verb itself, typically in the form of pronominal affixes. Look back at Section 4.3.3.2: the head-marking language, Kambera, is a typical CROSS-REFERENCING language, where free pronouns are only used for emphasis or disambiguation. So in (2), there are no independent ‘free’ pronouns meaning ‘I’ and ‘you’, and it’s only the subject and object markers on the verb that determine who’s doing the asking.

- (2) Jaka **ku-karai-kai** tiang ... (Kambera)
 if 1SG.SU-ask-2PL.OBJ later
 ‘If I ask you (*plural*) later ...’

In (2), the ‘bound’ pronominal affixes on the verb (shown in bold) are clearly crucial, whereas in Latin and English, verbal agreement markers don’t have much of a function in distinguishing subject and object.

These three systems – order, case and agreement – are not mutually exclusive: most languages use some combination of systems, although it is common for one to predominate. Sections 6.2 through 6.4 examine each system in turn. Section 6.5 then looks at noun phrases in terms of their GRAMMATICAL RELATIONS.

6.2 ORDER OF PHRASES WITHIN THE CLAUSE

6.2.1 Basic and marked orders

As we saw in Chapter 1, linguists often talk about the ‘word order’ of a particular language. In fact, this term refers not to single words but to the order of PHRASES, so a better term is CONSTITUENT ORDER. Here, I concentrate on the order of the three major constituents in a transitive clause: subject, object and verb. In many languages, including English, subjects are distinguished from objects by having a fixed position for each NP, in the ordinary, basic constituent order. Given the three constituents S, O and V, there are six logically possible variations, and indeed all six orders do occur as a basic constituent order amongst the languages of the world:

- (3) Mpša e-lomile ngwana. (Northern Sotho)
S V O
 dog SU-bit child
 ‘The dog bit a/the child.’

- (4) Müdür mektub-u imzala-dı. (Turkish)
 S O V
 director.NOM letter-ACC sign-PAST
 ‘The director signed the letter.’
- (5) Tuigeann Bríd Gaeilge. (Irish)
 V S O
 understands Bridget Irish
 ‘Bridget understands Irish.’
- (6) E kamatea te naeta te moa. (Gilbertese)
 V O S
 3SG kill.3SG the snake the chicken
 ‘The chicken killed the snake.’
- (7) kaikuxi etapa-và toto, papa tomo (Apalai)
 O V S
 jaguar kill-PAST 3PL father 3PL
 ‘They killed a jaguar, father’s group.’
- (8) anana nota apa (Apurinā)
 O S V
 pineapple I fetch
 ‘I fetch pineapple.’

In these examples, the constituent orders shown are all reasonably uncontroversial: they represent the basic order, or one of the basic orders, found in each of the languages. So, for instance, we can say Northern Sotho is an SVO language, and Turkish is an SOV language. However, saying a language has a certain basic constituent order doesn’t mean that it never has any other orders. For instance, English has a basic SVO order, as in *They adore syntax*, but we can also use an object-initial order, as in *Syntax, they adore*, to give particular emphasis to the direct object, in this case *syntax*. An order which is used like this to focus on a constituent is known as a MARKED (= non-basic) order.

In some languages, it is not easy to decide on a basic constituent order. First, two (or more) orders may be UNMARKED – equally neutral. For instance, some verb-initial languages such as Fijian, Tongan and Samoan (all Austronesian languages) are not clearly definable as either VSO or VOS: both orders are frequent. Languages which allow all of the six possible constituent orders are common; which order is actually chosen depends on pragmatic factors such as focus, and which constituent is the topic of the sentence. Some languages with free constituent order have one order which is clearly basic. So for example, the Slavonic languages Polish and Russian are SVO; Mohawk, by way of contrast, has no single basic order. Some languages also have free or very unrestricted WORD ORDER in the most literal sense; we will examine such languages in Section 6.6.

Second, some languages have a different order in root clauses and in subordinate clauses. For instance, a number of Germanic languages, including German and

Dutch, have SOV order in embedded clauses but have unmarked SVO order in root clauses; see Section 3.2.4 for discussion of this phenomenon.

Third, it may not be possible to tell whether there's an unmarked word order because sentences don't typically contain independent subject and object NPs. This is frequently the case in languages which are strongly HEAD-MARKING (see Section 4.3). The verb itself in such languages always has subject and object markers, as in (2), but in natural discourse there are very few clauses containing both a lexical subject and a lexical object NP (like *The bird killed the snake*), so we can't easily say what the order of S, O and V might be.

In instances like all these, the constituent order which is designated 'basic' often depends more on the theoretical allegiances of the linguist than on any properties of the language. The criteria linguists use to determine a basic constituent order include frequency, which means seeing how often each order occurs in a text, and neutrality, which means looking at sentences with no particular focus or emphasis. Native speakers also have strong intuitions about which order(s) are the most neutral, if any, and indeed whether or not word order changes make any difference to the meaning of a sentence.

6.2.2 Statistical patterns

The six basic constituent orders presented in Section 6.2.1 don't all have equal frequency. Statistically, we would expect to find the world's languages split evenly among the six possible orders. But in fact the basic orders SVO and SOV are by far the most frequent, between them covering around 80–90 per cent of the world's languages (roughly equally split). VSO is the only other major group, covering perhaps 9–12 per cent of languages, including Celtic, Semitic (for example, Biblical Hebrew and Classical Arabic) and Polynesian languages (such as Maori). Languages with the basic order VOS are much rarer, covering around 3 per cent of the world's total. As noted in Section 6.2.1, though, many verb-initial languages have both VSO and VOS as basic patterns. Both OSV and OVS were once thought not to exist as basic orders, and in particular the OSV order is extremely rare. But both are attested in the languages of the Amazon basin, as shown in (7) and (8). A certain amount of estimation is unavoidable in any figures given, not least because reliable information on basic constituent order is not always available.

Two major generalizations about constituent order in the world's languages emerge from the statistics. First, the vast majority of languages have subject-initial order (SOV, SVO), and even if subjects are not absolutely clause-initial, they generally precede objects (SOV, SVO, VSO). In one large language sample (Tomlin 1986), 96 per cent of the languages have subjects before objects. Why might this be? Subjects appear to be more salient than objects, which may account for their initial position: subjects typically initiate the action expressed by the verbal predicate, are often agents of that action or at least in control of it, and are often the topic of the clause. By way of contrast, objects are the theme or patient, the entity that is acted on, and are less typical as topics.

Second, the majority of languages place V next to O (in either order): again, over 90 per cent of a typical sample of languages do this. Only two constituent orders lack the

VO/OV grouping – the extremely rare OSV order and the much more frequent VSO order. In VSO languages, though, there are often alternative orders available which do place O and V together. For example, many VSO languages have an SVO alternative order (e.g. Arabic and Berber). And the Celtic languages, though generally considered to be VSO (like the Irish example in (5)), also have a very frequent auxiliary-SVO word order, as in (9). In this order, the subject precedes the other main elements in the clause and a transitive verb and its object are also grouped together into a VP:

- (9) Bhí an fear [VP ag péinteáil cathaoir inné]. (Irish)
Aux S V O
 was the man PROG paint.INFIN chair yesterday
 ‘The man was painting a chair yesterday.’

This grouping of O and V which predominates cross-linguistically gives support to the traditional two-way division of the clause into a subject and a predicate, which in turn contains the verb and its object (see Chapter 5).

Examination of large statistical samples of languages also reveals that the word order *within* constituents correlates with the order of the major constituents themselves (see, for example, Dryer 1991). In Chapter 4, I introduced the idea that languages fall into two basic groups, HEAD-INITIAL and HEAD-FINAL.

Head-initial order

- The verb *precedes* its objects and complement clauses.
- Adpositions are prepositions, giving [P NP] order in PPs.
- Complementizers (such as *that*, *if*, *whether*) precede the clause they select as complement.

Head-final order

- The verb *follows* its objects and complement clauses.
- Adpositions are postpositions, giving [NP P] order in PPs.
- Complementizers follow the clause they select as complement.

It turns out that OV languages (the largest group is SOV) are very generally head-final, while VO languages (SVO plus all verb-initial languages) are characteristically head-initial. For example, OV languages are far more likely to have postpositions than prepositions: in a typical sample (for instance, Dryer 1991), around 96 per cent of verb-final languages are postpositional. Conversely, VO languages are typically prepositional: only around 14 per cent of SVO languages have postpositions, and only 9 per cent of verb-initial languages. Similarly, in VO languages, complementizers such as *if* and *that* virtually always precede their subordinate clause, as in English. But in around 70 per cent of OV languages, the complementizers follow the subordinate clause; see, for example, the Japanese examples in exercise 3 in Chapter 3.

To summarize Section 6.2, for some languages constituent order is the major way to distinguish the grammatical relations (subject, object etc.) in a sentence. We expect such languages to have a fairly rigid constituent order, as is true of English, for example. Other languages have much more freedom of constituent order. These are typically languages which have case marking and/or a well-developed system of verb agreement: both these features allow subjects to be distinguished from objects even if the NPs don't have a fixed position in the sentence. The following two sections look in detail at case-marking and agreement, starting with an examination of case systems.

6.3 CASE SYSTEMS

6.3.1 Ways of dividing core arguments

In Chapter 4, I introduced the concept of a head and its dependents. We saw that the relationship between these elements need not be marked morphologically at all (for instance, it's not indicated in Chinese). But if it *is* indicated, this can be either by marking on the head (head-marking) or on the dependents (dependent-marking). In languages with CASE systems, the noun phrase dependents are marked to show their relationship with the head element in the phrase or clause. This section concentrates on the relationships between a head verb and its NP arguments; case-marking shows, for example, which NP is the subject and which the object.

We've often used the terms 'subject' and 'object'. But do these terms apply equally well to all languages? In this section, we'll see that it is helpful to distinguish between different types of subjects, in order to describe case systems that occur outside the familiar European language families. I will divide the CORE ARGUMENTS of a verb as shown in Table 6.1, and use the abbreviations S, A and O to designate their grammatical relations (Dixon 1972, 1979, 1994).

Table 6.1
The core arguments

Subject of an intransitive verb	S
Subject of a transitive verb	A
Object of a transitive verb	O

For example:

- (10) The snake(S) hissed.
 (11) The chicken(A) bit the snake(O).

You can remember 'S' as 'subject' but, more transparently, as the 'single' argument of an intransitive verb. 'O' is clearly 'object'. And 'A' is for 'agent', which is the prototypical semantic role taken by the subjects of transitive verbs such as 'bite', 'examine' or 'regurgitate'. All languages must have some way of distinguishing the transitive

subject, A, from the object, O, so that we can tell, for example, who gets bitten. In languages like English, fixed constituent order does this work. What, though, if the constituent order is free? One solution is to ensure that A has a different *form* from O: this is the role of case-marking.

A logically possible way of distinguishing the three core arguments would, of course, be to have a different marking for each of them. Such a language would distinguish three different cases, one for S, one for A, one for O; an example is given as (26) in this chapter. However, this is actually an extremely unusual system, cross-linguistically. The reason for this is undoubtedly because a much more economical system is attainable, using just two case distinctions. Only A and O need to be marked differently. There are no clauses with both an S and an A: they can't co-occur, because within any given clause the verb is either transitive or intransitive. Similarly, there are no clauses with both an S and an O: if the verb is intransitive, it just has an S, and not an O. So to achieve the most economical case system possible, there are two equally logical alternatives, both of which require just two case distinctions.

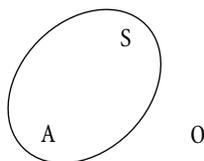
The first system marks S and A in the same way, and O differently. In other words, all subjects receive one case-marking, and objects receive a different case. This is known as the NOMINATIVE/ACCUSATIVE pattern, and it occurs in most European languages (a notable exception is Basque). In modern English, full noun phrases have no formal case-marking, but we can see the relics of a previous nominative/accusative case system in the forms of the first and third person pronouns:

(12) We(S) left.
We(A) like her(O).

(13) She(S) left.
She(A) likes us(O).

We and *she* are NOMINATIVE forms, used for both S and A: in other words, all subjects have the same form. *Her* and *us* are ACCUSATIVE forms, used for O.

Figure 6.1
The nominative/accusative grouping

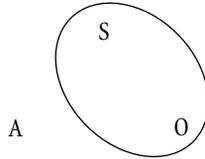


Because this grouping of S and A is so familiar from European languages, you may consider it entirely natural to case-mark all subjects in the same way. But remember that this is only one of the two equally economical ways of dividing the

core arguments. The second system marks S and O in the same way, but marks A differently; this is known as the *ERGATIVE/ABSOLUTIVE* pattern:

Figure 6.2

The ergative/absolutive grouping



ERGATIVE is the case of A – the subject of transitive verbs. *ABSOLUTIVE* is the case of both S and O, the subject of intransitive verbs and the object of transitive verbs.

A summary of the two systems is shown in Table 6.2. You can see that both case systems require only two distinctions. One system groups S with A (since they never co-occur); this is typically known simply as the *accusative* pattern. The other system groups S with O (they, too, never co-occur); this is typically known simply as the *ergative* pattern.

Table 6.2

The major case systems

Accusative system			Ergative system		
A	S	O	A	S	O
Nominative		Accusative	Ergative	Absolutive	

In the following two sections, I move on to an illustration of each of the main case systems in turn.

6.3.2 Nominative/accusative systems

I start with the most familiar system, *NOMINATIVE/ACCUSATIVE* (or just *ACCUSATIVE*). This system has an *AS/O* pattern: A and S are marked the same, O differently. Good examples are Latin, German, Japanese and Turkish, among many other languages. Subjects of both transitive and intransitive verbs are marked in the same way, with *NOMINATIVE* case. Objects of transitive verbs are marked with *ACCUSATIVE* case. This 'alignment' of NPs is sometimes indicated by using the notation $S = A \neq O$.

- (14) Puella veni-t.
 girl.NOM come-PRES.3SG
 'The girl(S) comes.'

(Latin)

- (15) a. Puer-um puella audi-t.
 boy-ACC girl.NOM hear-PRES.3SG
 ‘The girl(A) hears the boy(O).’
 b. Puella puer-um audi-t.
 girl.NOM boy-ACC hear-PRES.3SG
 ‘The girl(A) hears the boy(O).’

Since the A and O arguments of the verb *audit*, ‘hears’, are in different cases, there is no problem determining which is which, despite the free constituent order illustrated in (15a) and (15b).

Case is generally considered to be a property of an entire *noun phrase*, rather than just the head noun itself. In some languages, case is indeed marked on the head noun via changes in its morphology (= changes in its form), as in the Latin examples. But elsewhere, for instance in German, case is typically not marked on the head noun, but is marked instead on the determiners and any adjectives in the noun phrase:

- (16) [Der gross-e Hund] knurrte. (German)
 the.NOM big-NOM dog growled
 ‘The big dog growled.’
 (17) [Der gross-e Hund] biss [den klein-en Mann].
 the.NOM big-NOM dog bit the.ACC small-ACC man
 ‘The big dog bit the small man.’

The (masculine) head nouns *Mann*, ‘man’, and *Hund*, ‘dog’, in (16) and (17) don’t undergo any morphological changes: they’re in their basic form. But we can tell who gets bitten in (17) from the case-marking shown on other elements in the NP, the determiners and the adjectives. For instance, *der* is the nominative form of the definite article (‘the’) for masculine nouns, while *den* is its accusative form. The NP *den kleinen Mann* is thus shown as accusative, so it’s the object, while *der grosse Hund* is nominative, so it’s the subject.

6.3.3 Ergative/absolutive systems

The ERGATIVE/ABSOLUTIVE system (or just ERGATIVE) has an SO/A pattern: S and O are marked the same, and A is marked differently. Lezgian (a Daghestanian language spoken in the Caucasus) is a standard ergative language. The subject (A) of a transitive verb has ergative case, while the object (O) of a transitive verb and the subject (S) of an intransitive verb both have absolutive case. This ALIGNMENT of NPs is sometimes indicated by using the notation $S = O \neq A$. Compare in particular the forms of the first person singular pronouns (‘I/me’ in the English translations) in (18) through (20).

- (18) Za zi balk’an c’ud xipe-q^h ga-na. (Lezgian)
 I.ERG my horse.ABS ten sheep-for give-PAST
 ‘I(A) gave away my horse(O) in exchange for ten sheep.’

- (19) **Zun** ata-na.
 I.ABS come-PAST
 'I(S) came.'
- (20) Aburu **zun** ajib-da.
 they.ERG I.ABS shame-FUT
 'They(A) will shame me(O).'

In the English translations, the first person singular pronouns have the same form, *I*, both as an A and an S, while the O has a different form, *me*. By contrast, in Lezgian the A form (*za*) differs from the S, and instead the S and O forms are identical (*zun*). When the first person singular pronoun is an A – the subject of a transitive verb, as in (18) – it takes the ergative case, giving the form *za*. But when it's either an S (the subject of an intransitive verb) as in (19), or an O (an object) as in (20), it takes the absolutive case, giving *zun*.

Our second example comes from an ergative language spoken in Europe, namely Basque, which is a language isolate (= a language with no known relatives). Examples from the Lekeitio dialect are given in (21) through (23): compare the case-marking of the word for 'man' in each example.

- (21) **Gixona-k** liburua erosi dau. (Basque)
 man-ERG book.ABS buy AUX.3SG
 'The man(A) has bought the book(O).'
- (22) **Gixona** etorri da.
 man.ABS come AUX.3SG
 'The man(S) has come.'
- (23) **Gixona** ikusi dot.
 man.ABS see AUX.1SG
 'I(A) have seen the man(O).'

The NP meaning 'man' has the ergative case suffix *-k* in (21), where it's an A, i.e. the subject of a transitive verb. When this NP is an S or an O, as in (22) and (23), it takes the absolutive case.

If you understand the data, but are having difficulty remembering which NPs group together in the ergative/absolutive system, I recommend the mnemonic 'Abso', for 'A-but-SO' grouping. I hope this helps!

Ergativity is not found in the European language families (Romance, Germanic, Celtic, Greek, Albanian and so on – to which Basque is unrelated), and is also very rare in Africa. However, it is common in Australian languages, and also occurs widely in Tibeto-Burman languages, Mayan languages (Central America), and a number of Papuan languages (New Guinea), among others. In other words, ergative systems are not purely localized, but are spread around the world. Dixon (1994: 10) estimates that perhaps one quarter of the world's languages can be described as ergative languages; clearly, then, the accusative system is far more common.

examples show that in Pitta-Pitta, the two main case-marking systems – the ergative system and the accusative system – partially intersect.

Another kind of split ergative system is illustrated in Section 6.4.4.

6.3.5 Marked and unmarked forms

At this point, we can discover why linguists often just use the terms ‘ergative’ or ‘accusative’ to describe the two systems: it is common for *just this one* member of each system to be the only NP that is overtly case marked, while the other member of each system is unmarked, i.e. has no special inflection for case at all. Instead, we find the ordinary root of the noun or pronoun (the form with no inflections).

In an ergative system, the form lacking overt marking is the absolutive NP, while the ergative NP does have a special inflection. This is true of all the ergative systems illustrated so far, Lezgian, Basque and Dyirbal. Please confirm this by looking at the Dyirbal examples in (24): the absolutive forms are not inflected; the ergative form is.

In an accusative system, the form lacking overt marking is the nominative NP, and the accusative form has a special inflection. This is confirmed by (25): the nominative pronouns are not inflected, while the accusative one is.

In fact, we can make a generalization which works for both case systems: whichever case is used for the S argument (either absolutive or nominative), that will generally (with a few exceptions) be the NP that *lacks* any overt marking (Dixon 1994: 56f). Not only is the case used for S generally formally UNMARKED (= lacking special marking), as in the Dyirbal examples in (24) and (25), it’s also functionally UNMARKED. This means it’s more widespread in occurrence and more basic in terms of usage. For instance, the absolutive or nominative form is typically used as the citation form of a noun, generally the form given in a dictionary.

6.4 AGREEMENT AND CROSS-REFERENCING

6.4.1 What does verb agreement involve?

Case-marking and verb agreement (also termed concord) are in fact two alternative (and sometimes overlapping) ways to represent the same information. Recall that the relationship between a head verb and its dependent NPs can be morphologically indicated either by DEPENDENT-MARKING (case) or HEAD-MARKING (agreement). As we saw in Chapter 4, it’s very common for a language to have both verbal agreement with a subject and also case-marking on the core NPs: see example (1) from Latin. This is an instance of case and agreement overlapping.

In Section 6.4, we see that ‘an accusative system’ doesn’t necessarily imply that the language has nominative/accusative *case*; the same applies to ‘an ergative system’. The relationships between verb and core NPs may instead be shown by verb agreement, either following the accusative pattern or the ergative pattern. In other words, the verb agrees with certain of its dependents and not others.

Agreement, or cross-referencing, means that a head verb is formally marked to reflect various grammatical properties of its NP arguments. To take a simple

example, a verb might be marked for third person singular when its subject is a singular NP, and third person plural when its subject is a plural NP. (English has a verb agreement marker for third person singular subjects, -s, but only in the present tense: *S/he sings*.) Cross-linguistically, the most common categories involved in agreement are PERSON, NUMBER, GENDER (= noun class) and CASE. We will see that verb agreement can follow an accusative or an ergative pattern even when there's no actual case marking on the NPs themselves.

Logically, the options are for a verb to agree (a) with none of its arguments, (b) with some but not others, or (c) with all its arguments, and in fact all of these possibilities occur, as we will see later. There are, then, languages with no verb agreement whatever, for example Swedish, Japanese, Chinese, Maori and Malagasy. Example (27) illustrates this for Chinese:

- (27) a. Wo xihuan ta (Chinese)
 I like he
 'I like him.'
- b. Ta xihuan wo
 he like I
 'He likes me.'

The verb has the same form, *xihuan*, no matter what the person and number of the subject pronoun. In fact, constituent order is the sole way of distinguishing the subject and object in these examples, since there's no case-marking on the NPs either: the third person singular pronoun, for instance, is *ta* whether it's a subject or an object.

We next turn to languages that do have verb agreement.

6.4.2 Nominative/accusative agreement systems

Within the Indo-European family, it is common for the verb to agree only with the subject, as for example in Italian, French, Spanish, German, Dutch and English. Subject-only agreement also occurs in Turkish and other Altaic languages, in Tamil and other Dravidian languages, and in Finnish and other Uralic languages. Examples (28) and (29) illustrate from French:

- (28) Nous avons vu ce film. (French)
 we have.1PL seen this film
 'We(A) have seen this film(O).'
- (29) a. Nous avons décidé.
 we have.1PL decided
 'We(S) have decided.'
- b. Ils ont décidé.
 they have.3PL decided
 'They(S) have decided.'

These three examples show that there is subject/verb agreement in French, expressed on the auxiliary here rather than the lexical verb. So auxiliary *avons* has a first person plural inflection to agree with the 1PL subject pronoun *nous*, and a third person plural inflection to agree with the 3PL subject pronoun *ils* (29b). The subjects of transitive verbs (A) and the subjects of intransitive verbs (S) are both marked on the verb in the same way, while the verb does not agree with the object, *ce film*, in any way in (28). We can therefore say that French has an ACCUSATIVE agreement pattern, or alignment: A and S pattern together, as opposed to O. Again, this alignment can be indicated as $S = A \neq O$. French does not have case-marking on NPs, however: as in English, only pronouns display the relics of an earlier case system.

The other possibility, also common cross-linguistically, is that the verb cross-references more than one of its arguments. So in Kambera, which also has an accusative alignment, the verb cross-references both the subject *and* the object: these markers are shown in bold type in (30). To help you see what refers to what, I've indicated both the independent subject NP and the bound subject marker on the verb with a subscript _{SU}. I also indicate both the independent object NP and the bound object marker on the verb with a subscript _{OBJ}:

- (30) [I Ama]_{SU} **na**_{SU} -kei-**ya**_{OBJ} [na rí muru]_{OBJ}. (Kambera)
 the father 3SG.SU-buy-3SG.OBJ the vegetable green
 'Father buys the green vegetables.'

In Kambera, the subject marker is a prefix (i.e. it precedes the verb stem), and object marker is a suffix (i.e. it follows the verb stem). Example (30) has an overt subject and object, but if these are omitted the sentence is still perfectly grammatical, because the bound pronominals alone serve to indicate both a subject and an object. Such a sentence then simply has the (less specific) meaning 'He/she buys it'. Look back at example (2) to see an instance of this kind. As we saw there, and in Chapter 4, head-marking languages (such as Kambera) often have whole sentences consisting of just the verb. Free pronouns are generally not required, since the pronominal person and number affixes on the verb provide all the information about the verb's arguments: again, see (2).

In some languages, constituent order affects which agreement markers occur. So for example, in Northern Sotho, a Bantu language, the unmarked (= basic, usual) constituent order is SVO, as in (31) and (32):

- (31) Mpša **e**-lomilè ngwana. (Northern Sotho)
 dog SU-bit child
 'The dog bit a/the child.'

- (32) Di-mpša **di**-lomilè ngwana.
 PL-dog SU-bit child
 'The dogs bit a/the child.'

In (31) and (32), there is only a subject marker, a verbal prefix (shown in bold). This prefix agrees with the noun class and number of *mpša*, 'dog': this is a language with

from an Australian language called Biri, the verb cross-references three arguments, all of which are expressed as suffixes on the verb stem, and are shown in bold:

- (35) *nhula manhdha yaba-nha-la-nga-ngu* (Biri)
 3SG.SU food give-FUT-3SG.SU-3SG.OBJ-1.DU.DATIVE
 'He will give food to us two.'

The verb stem in (35) is *yaba*, and this has a future tense marker, followed by three pronominal affixes, or person/number markers: *-la* marks the third person singular subject (and there's also an independent third person subject pronoun *nhula*, 'he', here); *-nga* marks the third person singular object, agreeing with *manhdha* 'food'; and *-ngu* is a marker for first person dual ('us two'), and is also dative. Dative is a case often used to mark a recipient, which gives rise to the meaning of something being handed over to someone here.

In this section, we have seen accusative systems of agreement: the verbs agreed with their subjects, or both with their subjects and objects.

6.4.3 Ergative/absolutive agreement systems

We turn now to systems with ergative alignment. When verb agreement follows the ergative pattern, it marks S (intransitive subjects) and O (all objects) in the same way and A (transitive subjects) differently. So we can say that $S = O \neq A$. ERGATIVE AGREEMENT MARKING occurs in a number of Caucasian languages, and also in Mayan languages (Mexico and Central America). Our examples are from the Northwest Caucasian language Abaza. In (36), we have an intransitive verb, and, in (37), a transitive verb. In all these examples, the data consist simply of a verb with bound pronominal affixes showing the person and number of the participant(s): I have indicated in bold the function of each morpheme:

- (36) a. *d-thád.* (Abaza)
S-V
 3SG-go
 'He/she's gone.'
- b. *h-thád.*
S-V
 1PL-go
 'We've gone.'
- (37) a. *h-l-bád.*
O-A-V
 1PL-3SG.F-see
 'She saw us.'
- b. *h-y-bád.*
O-A-V
 1PL-3SG.M-see
 'He saw us.'

- c. d-h-bád.
O-A-V
 3SG-1PL-see
 ‘We saw him/her.’

All the person/number markers are prefixes on the verb in Abaza: note that they have a fixed order, S-V and O-A-V, so it is always clear who’s doing what. The prefixes show the SO versus A pattern characteristic of ergativity. Throughout, any S and O markers which refer to the same person/number have the same form. First, let’s look at third person singular prefixes. In (36a), we have a 3SG S prefix *d-*, giving a meaning equivalent either to ‘he’ or ‘she’, and the same prefix occurs as the 3SG O prefix in (37c), giving rise to the ‘him/her’ meaning: thus, SO group together. Note that *d-* is only an SO form, and is, of course, not used to mark a third person singular A, since we’re dealing with a grouping of SO vs. A here. Instead, the 3SG A prefixes in (37a) and (37b) occur in an entirely different form, and moreover they’re differentiated according to gender (*l-* for the 3SG feminine, and *y-* for the 3SG masculine), which the SO form isn’t.

Next, let’s look at first person plural prefixes. In (36b), we have a 1PL S prefix *h-*, giving the ‘we’ meaning, and the same prefix occurs as the 1PL O prefix in (37a) and (37b), giving the ‘us’ meaning. The data contain an additional complication which you may have noticed: *h-* also means first person plural (‘we’) in (37c), where it’s an A, rather than S or O. How then do native speakers of Abaza know what’s going on? The answer is that because the order of prefixes is fixed, the data indicate clearly to an Abaza speaker that *h-* really is the A argument in (37c), the subject of the transitive verb, since it follows the O prefix. It is rather common for languages to ‘re-use’ pieces of morphology in this way: as long as they are clear in their context, duplications of this kind don’t appear to cause confusion.

6.4.4 Split systems II

A language with an ergative agreement system may have ergative case marking too (for instance, the Northeast Caucasian language Avar) but it is also possible to have ergative agreement on the verb but no case-marking on NPs – in fact, Abaza (illustrated earlier) falls into this category. There are also languages which have ergative case-marking on NPs, but a nominative/accusative system of cross-referencing on the verb. The Australian language Warlpiri illustrates this system. Lexical noun phrases and independent (i.e. freestanding) pronouns are all marked with ergative/absolutive case (compare the Dyirbal in (24) and (25)). The agreement markers (in bold) are affixed to the auxiliary, the element in second position in (38) and (39):

- (38) Ngaju ka-**rna** wangka-mi. (Warlpiri)
 I.ABS AUX.PRES-1SG.SU speak-NONPAST
 ‘I(S) am speaking.’
- (39) Ngajulu-rlu ka-**rna-ngku** nyuntu nya-nyi.
 I-ERG AUX.PRES-1SG.SU-2.OBJ you.ABS see-NONPAST
 ‘I(A) see you(O).’

Look first at the independent pronouns in (38) and (39), which are marked according to the ergative/absolutive system. The S argument *ngaju* ('I') in (38) is absolutive, as is the O argument *nyuntu* ('you') in (39). Conversely, the A argument *ngajulu-rlu* ('I') in (39) is ergative. SO thus group together in opposition to A, as we anticipate in this system. We clearly see that the pronoun for 'I' has a different case according to whether it's the subject of an intransitive verb (S), as in (38), or a transitive verb (A), as in (39).

Now look in contrast at the verb agreement. This marks *both* instances of first person singular in the same way, with the suffix *-rna* designating any first person singular subject. So the affixes reflect a grouping of all subjects (AS) as opposed to all objects (O), namely a nominative/accusative system. The Warlpiri system is not at all unusual, whereas there are no known languages with accusative case systems but ergative agreement systems. This is, then, another way in which the accusative system predominates cross-linguistically.

6.5 GRAMMATICAL RELATIONS

6.5.1 Investigating core grammatical relations

In this section, we examine the cross-linguistic properties of two major core grammatical relations, SUBJECT and OBJECT. To show that these concepts exist, we need to demonstrate that certain linguistic phenomena are best described in terms of 'subject' or 'object'. For languages in the nominative/accusative class, it's clear that 'subject' and 'object' are valid categories: in the last few sections, we've seen a number of illustrations of both case and verb agreement operating in terms of a subject/object split. The examples seen so far show that certain languages are *morphologically* nominative/accusative. This means that the characteristic AS/O split is indicated by changes in the morphology (form) of the NPs, via case-marking, or in the morphology of the head verb, via agreement, or indeed by marking on both NPs and verbs. However, the AS/O pattern is also pervasive in syntax itself. This means that many languages – including those with no case-marking or even with ergative case-marking – are *syntactically* nominative/accusative. In such languages, there are a number of syntactic processes which revolve around the subject and object relations – in fact, particularly the subject, since this grammatical relation is by far the most important. We'll examine some of these processes in this section, and return to this topic in Chapter 7.

The subject relation is crucial cross-linguistically: subjects tend to control the syntax in a number of ways, as we'll see. However, it's hard to give a satisfactory definition of 'subject', because no single property is shared by all subjects in all languages. Instead, there's a set of properties typical of subjects, and each language is likely to exhibit a subset of these properties. We begin by looking at some of the main cross-linguistic properties of subjects (Section 6.5.2), and then turn to the question of subjecthood in specific languages (Section 6.5.3).

6.5.2 Subjects: Typical cross-linguistic properties

i. Subjects are normally used to express the AGENT of the action, if there is an agent.

- ii. Subjects tend to appear first in the clause in unmarked (basic) constituent order. Recall that up to 90 per cent of languages are either SOV or SVO, therefore subject-initial. But since that leaves 10 per cent or more of the world's languages that are not subject-initial, we can't use this as a defining property.
- iii. Subjects are understood as the missing argument in IMPERATIVE constructions. An imperative is a command such as *Sit!* or *Eat up your greens!* Both intransitive and transitive verbs have an understood (or in some languages, overt) second person subject pronoun ('you') in the imperative.
- iv. Subjects control REFLEXIVE NPs, that is, '-self' forms such as the English *herself*, *themselves*, and also RECIPROCAL NPs such as *each other*. So we get *My sister really admires herself*, where the NP *herself* (feminine singular) refers back to the feminine singular subject, *my sister*, but we don't get **Herself really admires my sister*. Note that we can't simply say that the reflexive must refer to a *preceding* NP. We see this in the Madagascan language Malagasy, which has VOS order, so the subject does not precede the '-self' form. Nonetheless, the subject determines the reference of the '-self' NP; that is, the subject determines which NP the '-self' form refers to.

- (40) a. Manaja tena Rabe. (Malagasy)
 respect self Rabe
 'Rabe respects himself.'
- b. *Manaja an-dRabe tena.
 respect ACC-Rabe self
 '*Himself respects Rabe.'

In (40b), *Rabe* is the object, as we can tell from its accusative case marker; only when *Rabe* is the subject is the reflexive sentence grammatical, as in (40a).

- v. Subjects often control the referential properties of an NP in another clause. For instance, when two clauses are conjoined, as in (41), the subject of the second clause can be omitted because it is CO-REFERENTIAL with the subject of the first clause, *Chris*: I show the omitted NP with \emptyset . But it's only the subject that can be omitted, (41a), not the object, (41b). Moreover, the NP that's omitted has to refer back to the subject of the first clause, *Chris*, and not the object, *Lee*. The subscripts _i and _j here have no meaning of their own, but are simply labels to show which NPs co-refer (= designate the same entity).

- (41) a. [Chris_i phoned Lee_j] and [\emptyset _i met him_j later].
 b. *[Chris_i phoned Lee_j] and [he_i met \emptyset _j later].

Second, in many languages verbs like 'begin' and 'want' take an infinitival complement clause, as in *Kim began [to grate the carrots]*. The 'understood' subject of the 'grate' clause is co-referential with the main clause subject. But only the subject in the infinitival clause – and not the object – can be the 'understood' NP:

- (42) a. Chris_i wants [\emptyset _i to meet this famous film star].
 b. *Chris_i wants [this famous film star to meet \emptyset _i].

vi. Subjects are the most usual target for promotion from other positions. For instance, the PASSIVE construction promotes an NP from direct object position to subject position (see Chapter 7), turning *The students forgot her* into *She was forgotten (by the students)*: the pronoun has the form *her* as an object, but *she* as a subject. Although not all languages have promotion processes, if a language has any promotion processes, then it will have ones that move some constituent into subject position.

6.5.3 An examination of subjects in specific languages

We turn now to an examination of subjects in particular languages. We look first at Icelandic (Section 6.5.3.1), which has nominative/accusative morphology and syntax, and so has a clear SUBJECT relation. Section 6.5.3.2 then turns to a morphologically ergative language, Lezgian, for which the notion of subject is more controversial. Section 6.5.3.3 examines Tagalog, which represents a language type different to both accusative and ergative. Section 6.5.3.4 asks whether there are universal grammatical relations.

6.5.3.1 Icelandic

Icelandic is a standard accusative language – subjects are usually in the NOMINATIVE case and objects in the ACCUSATIVE case:

- (43) Ég sá stúlkuna.
 I.NOM saw.1SG the.girl.ACC
 ‘I saw the girl.’

Verbs in Icelandic agree in person and number with the nominative subject:

- (44) a. Við dönsuðum.
 we.NOM danced.1PL
 ‘We danced.’
 b. Þeir dóu.
 they.NOM died.3PL
 ‘They died.’

However, some verbs are exceptional: their subjects take a case other than nominative. In (45), we have a dative subject, and in (46), an accusative subject, *hana* ‘her’ (the object is also accusative in (46)):

- (45) Henni leiddist.
 her.DATIVE bored
 ‘She was bored.’

- (46) **Hana** vantar peninga.
 her.ACC lacks money.ACC
 ‘She lacks money.’

These subjects with ‘quirky’ case don’t trigger subject/verb agreement. In (47), the subject is a *plural* pronoun *þá*, ‘them’ (accusative), but we find the same form of the verb *vantar*, ‘lacks’, as in (46), where the subject is *singular*. Compare (44b), where the nominative *þeir* ‘they’ triggers agreement, giving a plural form of the verb:

- (47) **Þá** vantar peninga.
 them.ACC lacks money.ACC
 ‘They lack money.’

So if these ‘quirky’ subjects don’t trigger verb agreement, on what grounds can we say they’re subjects? There are, in fact, a number of diagnostics for subjects in Icelandic, and the NPs with quirky case pass all of these tests. First, subjects can undergo subject/verb inversion (see Section 3.2.4 on inversion in English). Example (48) shows that an ordinary nominative subject inverts with the finite verb to form a yes/no question, and in (49), we see that a dative subject also inverts. The subjects are in bold type:

- (48) Hafði **Sigga** aldrei hjálpað Haraldi?
 had Sigga.NOM never helped Harold.DATIVE
 ‘Had Sigga never helped Harold?’
- (49) Hefur **henni** alltaf þótt Ólafur leiðinlegur?
 has her.DATIVE always thought Olaf.NOM boring
 ‘Has she always thought Olaf boring?’

Even though there’s also a nominative NP *Ólafur* in (49), this couldn’t be inverted with the verb *hefur* ‘has’.

Second, when two clauses are conjoined, the subject of the second clause can be omitted when it’s co-referential with the subject of the first clause, just as in English: see (v) in Section 6.5.2. Example (50) illustrates with ordinary nominative subjects:

- (50) Þeir fluttu líkið og (þeir) grófu það.
 they.NOM moved the.corpse and they.NOM buried it
 ‘They moved the corpse and (they) buried it.’

Turning next to a quirky subject, we see in (51) that the verb meaning ‘like’ takes a dative subject:

- (51) **Mér** líkar vel við hana.
 me.DATIVE likes well with her
 ‘I like her.’

And it turns out that this dative subject can undergo this SUBJECT ELLIPSIS (= omission) too: the dative subject pronoun *mér* can be omitted in the second clause in (52):

- (52) Ég sá stúlkuna og (mér) líkaði vel við hana.
 I.NOM saw the.girl.ACC and me.DATIVE liked well with her
 ‘I saw the girl and (I) liked her.’

In fact, not only can a quirky subject *undergo* ellipsis, as in (52), it can also be the NP which *permits* ellipsis of another subject. This, then, is the third test for subjecthood. Example (53) has a dative subject in the first clause, and the nominative subject *þeir* can undergo ellipsis in the second clause:

- (53) Þeim líkar maturinn og (þeir) borða mikið.
 them.DATIVE likes the.food.NOM and they.NOM eats much
 ‘They like the food and (they) eat a lot.’

Note that even though there is a nominative NP in the first clause, this is not the subject, and a missing subject can't refer back to it: the sentence couldn't mean, even jokingly, that the food eats a lot.

In sum, then, these (and other) tests for subjecthood in Icelandic show that subject NPs with quirky case really *are* subjects, despite the fact that they fail to trigger subject/verb agreement.

6.5.3.2 Lezgian

We saw earlier that morphologically ERGATIVE languages (i.e. those with ergative case and/or agreement) may be syntactically ACCUSATIVE. This means that syntactic constructions such as subject ellipsis utilize a grouping of the S and A arguments, as opposed to the O argument ($S = A \neq O$). In fact, it is quite usual for languages which have morphologically ergative alignment to be accusative in terms of their *syntax*, and much rarer for them to have ergative syntax. We will see more on this in Chapter 7.

As we saw in Section 6.3.3, Lezgian is morphologically ergative: the case marking on NPs contrasts absolutive (on S and O noun phrases) with ergative (on A noun phrases, the subjects of transitive verbs): $S = O \neq A$. It will help to review the discussion of (18) through (20) before reading further.

Evidence of syntactic accusativity in Lezgian comes from the fact that it has a SUBJECT grammatical relation (Haspelmath 1993). Let's look first at some basic data. The 'subject' consists of three NP types. The first two types are the A and S arguments, i.e. the two NPs that would constitute the 'subject' relation in an accusative language. To illustrate these two, we have the ergative argument (A) of a transitive verb, as in (54), and the absolutive argument (S) of an intransitive verb, as in (55). The NPs in bold type in (54) to (56) are the putative subjects:

(54) **Ruš-a** gadadi-z cük ga-na.
 girl-ERG boy-DATIVE flower.ABS give-PAST
 ‘The girl gave a flower to the boy.’

(55) **Ruš** elq̄wena q̄’uluq̄^hdi kilig-na.
 girl.ABS turn backward look-PAST
 ‘The girl turned around and looked back.’

The third potential ‘subject’ is the experiencer argument of verbs with meanings such as ‘want’, ‘see’ and ‘be afraid’, which in Lezgian take the DATIVE case, as in (56); cross-linguistically, this use of dative case for the semantic role of experiencer is quite common:

(56) **Ruša-z** ada-q^haj kič’e řa-na-č.
 girl-DATIVE he-of afraid be-PAST-NEG
 ‘The girl wasn’t afraid of him.’

Note that all three putative subject NPs have different cases, so we certainly can’t identify ‘subjects’ by their morphological case in Lezgian. Furthermore, although the NP in bold in (55) is absolutive, not all absolutive NPs are subjects, of course: the O noun phrase *cük* ‘flower’ in (54) isn’t. Similarly, not all dative NPs are subjects: *gadadz* ‘boy’ in (54) isn’t the subject – dative case here identifies the indirect object (see Section 6.5.4). If there is a ‘subject’ grammatical relation in Lezgian, then, it cuts across the morphological case-marking.

So why would anyone think that Lezgian has a ‘subject’ relation? Constituent order provides some indication that all three NP types in bold in (54) through (56) pattern together: all have the same clause-initial position, which, as we know from Section 6.2, is the most common position for subjects cross-linguistically. But position alone won’t uniquely identify subjects in Lezgian, because the constituent order is actually very free, so other NP types can be initial in the clause.

However, we can test for subjects using a construction parallel to that in (42) in Section 6.5.2 – please look back to check on this – in which an embedded infinitival clause has an understood subject that refers back to the main clause subject. Look first at the English translations in (57) through (59) to get the idea of the construction, which is very similar in the two languages: the infinitival clause is the complement of a finite verb ‘wants’ in the matrix clause. The main difference is that in Lezgian, the infinitival clause (shown in square brackets) precedes the finite verb *k’anzawa* ‘wants’, while in English the embedded clause follows *wants*. Crucially, the *understood* subject in Lezgian (marked with \emptyset) can only be one of the three NP types tentatively identified earlier as forming a ‘subject’ category: either an ergative subject (an A), an absolutive subject (an S) or a dative subject:

(57) Nabisata-z [[NP_{Erg} \emptyset]_i] ktab k’el-iz] k’an-zawa.
 Nabisat-DATIVE (Subject) book read-INFIN want-IMP
 ‘Nabisat wants to read a book.’

- (58) Nabisata-z_i [[NP_{Abs} Ø_i] q^hüre-z] k'an-zawa.
 Nabisat-DATIVE (Subject) laugh-INFIN want-IMPF
 'Nabisat wants to laugh.'
- (59) Nabisata-z_i [[NP_{Dative} Ø_i] xwa akwa-z] k'an-zawa.
 Nabisat-DATIVE (Subject) son see-INFIN want-IMPF
 'Nabisat wants to see her son.'

We know what case the understood subject would have in each example by looking at what happens in ordinary finite clauses with overt (= pronounced) subjects: the verb for 'read' takes an ergative subject, the verb for 'laugh' an absolutive subject, and the verb for 'see' a dative subject. Compare (60): here, the understood NP is again absolutive, but (60) is ungrammatical because this absolutive NP is an O, an absolutive *object* (Musa is the one being sent) rather than an S, an absolutive *subject* as in (58):

- (60) *Musa-z_i [didedi [NP_{Abs} Ø_i] šherdi-z raqur-iz] k'an-zawa.
 Musa-DATIVE mother.ERG (Abs.Object) town-DATIVE send-INFIN want-IMPF
 ('Musa wants to be sent to town by his mother.')

In sum, the Lezgian data show that even a morphologically ergative language may display syntactic accusativity, and indeed there does seem to be evidence for a subject relation in Lezgian.

6.5.3.3 Tagalog

In this section, we will examine a language which resists clear classification into either the accusative or the ergative type, and seems in fact to have a totally different marking system for NPs. In Tagalog and other languages of the Philippines, NPs are not case-marked, but they are each preceded by a marker (which we can consider a preposition) that indicates their SEMANTIC ROLE (see Section 2.3.1). The preposition *ng* marks both agent and theme; *sa* (or *mula sa*) marks locative – i.e. indicating location, and glossed as 'from' in (61); and *para sa* marks beneficiary, glossed as 'for' in (61). However, in every sentence *one* of the NP participants must be chosen to be the TOPIC of the clause, and it is marked as such by a special preposition, *ang*, which replaces the marker the NP would have otherwise. The topic is shown in bold in each example. (Note that the topic is always understood to be definite, while the other NPs can be understood as definite or indefinite.)

Also, the verb itself has an affix that marks the semantic role of the NP chosen as topic: I've indicated this role beneath the gloss for the verb in each example. This marking is clearly a kind of verb agreement, but it is different from either the accusative system or the ergative system in that it does not operate in terms of the grammatical relation of the NP arguments. In examples like (61), any one of the NP participants can be marked as the topic – and whichever semantic role the topic has will be indicated on the verb, resulting in a verb marked to agree with one of the properties 'agent', 'theme', 'locative' or 'beneficiary'; this is shown in the different morphology that the verb has in each example:

- (61) a. Kukuha **ang babae** ng bigas sa sako para sa bata.
 FUT.take.out TOPIC woman THEME rice from sack for child
AGENT.TOPIC
 ‘**The woman** will take some rice out of a sack for a/the child.’
- b. Kukunin ng babae **ang bigas** sa sako para sa bata.
 FUT.take.out AGENT woman TOPIC rice from sack for child
THEME.TOPIC
 ‘A/the woman will take **the rice** out of a sack for a/the child.’
- c. **Ang sako** ay kukunan ng bigas ng babae para sa bata.
 TOPIC sack be FUT.take.out THEME rice AGENT woman for child
LOCATIVE.TOPIC
 ‘**The sack** will have rice taken out of it by the woman for the child.’
- d. **Ang bata** ay ikukuha ng bigas ng babae mula sa sako.
 TOPIC child be FUT.take.out THEME rice AGENT woman from sack
BENEFICIARY.TOPIC
 ‘**The child** will have rice taken out of the sack for him/her by the woman.’

It’s clear, then, that Tagalog isn’t morphologically marked in accordance with either the accusative system or the ergative system, either by case-marking or by verbal agreement.

However, as we have already noted, a language may nonetheless be *syntactically* accusative despite not being morphologically accusative. Does Tagalog fit this pattern? Looking at the syntactic behaviour of NPs, it turns out that some processes operate in terms of topics, irrespective of their semantic and syntactic role. But there are also other processes that operate in terms of a grouping of the A and S noun phrases, whether or not they are topics: this is a syntactically accusative pattern, and suggests that there may after all be a ‘subject’ in Tagalog.

Let’s look first at a process that targets topics: the ‘all’ construction. In (62), we see that *lahat*, ‘all’, is understood as modifying whichever NP is the topic. In (62a), the topic is the agent, so *lahat* must modify the A noun phrase, meaning ‘the children’. But in (62b), we have a theme topic, referring to the ‘thing written’, so *lahat* must modify the O noun phrase, meaning ‘the letters’. Note that *lahat* is not even adjacent to this latter phrase, *ang mga liham*, in (62b):

- (62) a. Susulat lahat **ang mga bata** ng mga liham.
 FUT.write all TOPIC PL child THEME PL letter
AGENT.TOPIC
 ‘All the children will write letters.’
- b. Susulatin lahat ng mga bata **ang mga liham**.
 FUT.write all AGENT PL child TOPIC PL letter
THEME.TOPIC
 ‘The/some children will write all the letters.’
not ‘All the children will write letters.’

Since linguists don’t normally consider a grouping of A and O to form any grammatical relation, the ‘all’ construction favours a view of Tagalog as not having subjects.

Now consider a syntactic process in Tagalog that targets S and A noun phrases – the classic ‘subject’ pairing – and not topics. Our examples are complement clauses with an understood subject. You will find it helpful to look again at the discussion of similar examples given earlier on before reading further: see (42) from English and (57) through (59) from Lezgian. In the Tagalog construction, the ‘missing’ subject is always an S or an A, whether or not it’s a topic. In the examples in (63), the topic of each matrix clause (the ‘hesitate’ clause) is the agent, *siya*, meaning ‘he’ (the suffix *-ng*, which I have left un glossed, indicates that an embedded clause follows). In the embedded clause, however, the topic is different in each example:

- (63) a. Nagatubili siya-ng [humiram ng pera sa banko].
 hesitate he.TOPIC-NG borrow THEME money from bank
 AGENT.TOPIC AGENT.TOPIC
 ‘He hesitated to borrow money from a/the bank.’
- b. Nagatubili siya-ng [hiramin **ang pera** sa banko].
 hesitate he.TOPIC-NG borrow TOPIC money from bank
 AGENT.TOPIC THEME.TOPIC
 ‘He hesitated to borrow **the money** from the bank.’

Remember that these ‘borrow’ clauses are embedded clauses with an understood subject. In both embedded clauses in (63a) and (63b), this understood subject is the agent (meaning ‘he’), an A noun phrase. In (63a), the missing agent NP in the ‘borrow’ clause happens also to be the NP chosen as the topic, as we can tell from the form of the verb, which, as you’ll recall, is marked for the semantic role of the topic. In (63a), then, there isn’t an overt *ang*-NP – an overt topic – because this topic is the ‘understood’ NP. But in (63b), the topic of the embedded clause is the theme (thing borrowed), namely *ang pera* ‘the money’, yet the understood subject of that clause is still the A noun phrase (‘he’). So topics are clearly *not* the targets for the ellipsis (the omitted part) in this construction. In fact, this process of NP ellipsis suggests that Tagalog is syntactically accusative, at least in this one construction. In other words, the understood subject can be either an S or an A argument (i.e. any type of ‘subject’), but not an O argument. Any process that treats S and A noun phrases together – and O arguments differently – suggests that the language operates at least part of its syntax in terms of a nominative/accusative alignment. Tagalog may indeed, then, have a SUBJECT grammatical relation consisting of S and A.

6.5.3.4 Language universals?

We are left with the indication that the ‘subject’ relation is very important cross-linguistically, even occurring in some languages which are not otherwise nominative/accusative in their morphology and elsewhere in their syntax. Should it be considered a language universal? Some linguists argue that it should not. For instance, Dryer (1997) argues that there are no universal grammatical relations, discussing data from Dyrbal, Cree, Cebuano and Acehnese which are particularly problematic for a view that ‘subject’ is a universal category. This view is supported by Croft (2001); see

also Haspelmath (2007). We will leave this an open question. However, it is clear that recurring properties in grammatical relations *are* found cross-linguistically, among languages of very different syntactic types and from totally unrelated language stocks. It seems, then, that these properties may reflect some universal features, even if these are not fully understood at present.

Finally, we saw in Section 6.5 that languages which have ergative (or no) case-marking may nonetheless exhibit nominative/accusative syntax. Is the opposite situation ever seen – in other words, can a language be syntactically ergative even if it doesn't have ergative/absolutive case morphology? Linguists used to think not, but more recently it has been shown that this situation does exist (Donohue and Brown 1999). Certainly, though, it is very rare.

6.5.4 Objects

The other major grammatical relation is that of OBJECT, in accusative systems the complement of a two-argument verb. There is plenty of morphological evidence for the existence of an object relation in languages with nominative/accusative morphology, since the O argument is designated by a special case (accusative) and/or verb agreement. This chapter contains examples of a case-marked O from languages as genetically diverse as Latin (1), Turkish (4) and Dyirbal (25). (Recall from Section 6.3.4 that Dyirbal is largely ergative, but its first/second person pronouns have an accusative case system.) Verb agreement with the O argument is shown in several examples: see (30) from Kambera and (39) from Warlpiri.

Syntactic evidence for the O relation is more limited than for subjects, but in many languages only an O can be passivized (see Chapter 1 and also Chapter 7 for a demonstration of this). Recall that in Icelandic, we see certain constructions in which a noun phrase doesn't receive the expected case-marking, but instead gets a 'quirky' case: examples of *subject* NPs with quirky case were given in Section 6.5.3.1. Icelandic also has certain *object* NPs with quirky case, so we can see how these act in terms of typical object behaviour. It turns out that not only do ordinary accusative O arguments undergo passivization, so too do O arguments with quirky case. An example of a quirky object is the NP *mér* in (64): this is not accusative, as objects typically are in Icelandic, but rather it is dative:

- (64) Þeir hjálpuðu mér. (Icelandic)
 they.NOM helped me.DATIVE
 'They helped me.'

Like other objects, however, this O can be promoted to subject position, giving (65). Note, though, that the dative case remains on this NP – it doesn't become nominative – though its position is the standard clause-initial position of the subject in Icelandic:

- (65) Mér var hjálpað. (Icelandic)
 me.DATIVE was helped
 'I was helped.'

Verbs such as ‘give’, ‘send’ and ‘show’, which take three arguments (*X gave Y to Z*), can in some languages be said to distinguish a DIRECT OBJECT from an INDIRECT OBJECT. In accusative languages with extensive case systems, the direct object bears accusative case, while what is traditionally termed the indirect object bears DATIVE case, as in Turkish, German, Greek and Latin. The indirect object is typically the ‘recipient’ or ‘goal’ NP, such as *mir* in (66):

- (66) Mein Freund gab mir sein Fahrrad. (German)
 my.NOM friend gave me.DATIVE his.ACC bicycle
 ‘My friend gave me his bicycle.’

The dative is also used for this same purpose in many ergative languages: see (54) from Lezgian. Cross-linguistically, then, the central use of the dative case is to designate the NP that’s the recipient or the beneficiary or the goal of a three-argument verb.

But this type of NP does not always get a special case. For instance, although in Ancient Greek most three-argument verbs have an accusative direct object and a dative indirect object, the verb for ‘teach’ is exceptional in that both of its complements (the NPs meaning ‘the boy’ and ‘the music’) have accusative case:

- (67) Edidaxan [ton paida] [tēn mousikēn]. (Ancient Greek)
 taught.3PL the.ACC boy.ACC the.ACC music.ACC
 ‘They taught the boy music.’

In fact, in English and many other languages, there is little justification for distinguishing an ‘indirect object’ from any other object. Very often, the recipient NP looks just like a direct object – in what is known as the DOUBLE OBJECT construction, the recipient immediately follows the verb and has the same case-marking as any object, as in *Kim lent me the book*. Alternatively, the recipient appears in an ordinary PP headed by ‘to’ or ‘for’, as in *Kim made a cake for me*. In other words, there’s neither a special case nor any special syntactic behaviour associated with the NP that traditional grammar calls the indirect object.

In Section 6.4, we saw that a ditransitive verb such as ‘give’ may agree with all three of its argument NPs – see (35) from Biri. However, a more common situation is that only two arguments of a three-argument verb are actually marked on the verb. One is always the subject, but languages differ in terms of which other NP the verb agrees with: it can be either the NP with the semantic role of theme (such as ‘thing given’), or else the recipient. Commonly, and perhaps surprisingly for speakers of European languages, the verb agrees with the *recipient* NP, rather than the ‘thing given’ as we might expect. Example (68) illustrates from Warlpiri, which, as we saw in (38) and (39), has ergative case marking but accusative verb agreement:

- (68) Ngaju-ku ka-**npa-ju** karli yi-nyi nyuntulu-rlu.
 me-DATIVE PRES-2SG.SU-1SG.OBJ boomerang give-NONPAST you-ERG
 ‘You are giving me a boomerang.’

There are two agreement suffixes in (68). The first, *-npa*, marks the subject ('you'). There is no agreement marker for *karli*, 'boomerang', on the verb at all, but the second suffix is a 1SG marker for the recipient, the dative NP *ngajuku*, meaning 'me'. What's more, the 1SG agreement suffix *-ju* is the same suffix that is used to mark the first person singular O argument of an ordinary transitive verb (as in 'You saw me'). It appears, then, that in some languages the recipient functions as a kind of object. Perhaps it's not too surprising that the verb marks the recipient here: it is typically human, or at least animate, and thus arguably more important than the inanimate theme NP.

6.6 FREE WORD ORDER: A CASE STUDY

Having examined case, agreement and grammatical relations, we are now in a position to return to the topic of word order. We have already seen that languages with extensive case-marking on noun phrases typically allow much variation in constituent order (see, for example, the German data in Section 4.3.3.1, and also the discussion of Japanese in Section 8.3). The same is true of languages with extensive head-marking on the verb, such as Kambera. The current section shows that some languages also allow extremely free *word* order, in the most literal sense. One such language is Latin; another is Navajo. Our illustrations, though, are from Australian languages, and in particular, Warlpiri. To remind you, Warlpiri exhibits a split-ergative system: it has ergative/absolutive case for independent noun phrases and pronouns, but an accusative system for the pronominal affixes marked on the auxiliary (see Section 6.4.4).

First, we illustrate the fact that Warlpiri has free *constituent* order: the only restriction is that the auxiliary, expressing tense and person/number marking, must be in second position in the clause:

- (69) a. Ngarrka-ngku ka wawirri panti-rni. (Warlpiri)
 man-ERG AUX.PRES kangaroo.ABS spear-NONPAST
 'The man is spearing the kangaroo.'
- b. Wawirri ka panti-rni ngarrka-ngku.
 kangaroo.ABS AUX.PRES spear-NONPAST man-ERG
 'The man is spearing the kangaroo.'
- c. Panti-rni ka ngarrka-ngku wawirri.
 spear-NONPAST AUX.PRES man-ERG kangaroo.ABS
 'The man is spearing the kangaroo.'

These three as well as the other three orders of S, O and V are all possible, with no single basic order. A further Warlpiri example is shown in (70):

- (70) a. Jarntu-jarra-rlu lpa-pala-jana ngaya nya-ngu
 dog-DUAL-ERG AUX.IMPF-3.DUAL.SU-3PL.OBJ cat.ABS see-PAST
 'The two dogs were looking at the cats.'

which doesn't form a constituent *can't* precede the auxiliary; this, then, is one test for constituent structure in Warlpiri. If constituents have a different syntactic status to random strings of words, this again suggests that constituents are real in such languages. Finally, examples like (72) showed that in Warlpiri, only the final element in a continuous NP constituent needs to be case-marked – (72a) vs. (72b/c). This indicates that there truly is an NP in (72a): it acts as a unit, so each of its subparts doesn't need a case affix. It seems clear, then, that constituent structure does play a vital role even in free word order languages.

Interestingly, there is apparently no correlation among Australian languages between the existence of discontinuous NPs and the appearance of free constituent order; for instance, Austin and Bresnan (1996) report that Diyari has discontinuous NPs but prefers a fixed SOV constituent order. Moreover, though Warlpiri has an extensive system of bound pronominal marking in the auxiliary (Section 6.4.4), it also appears that, cross-linguistically, this is not a necessary condition for the appearance of split NPs: case-marking on independent NPs and pronouns is enough to allow for discontinuous constituents. So for instance, the Australian languages Jiwari, Dyirbal and Yidiny all have discontinuous NPs but lack the pronominal cross-referencing affixes that characterize the Warlpiri auxiliary.

I hope to have shown definitively in this final section that if we only looked at English and its close relatives, we'd be missing out on a great deal of knowledge about the potential of the human language faculty. I also hope that by this point, you're feeling more comfortable about analysing examples from 'exotic' languages. More are to come in Chapter 7!

6.7 SUMMARY

This chapter has examined three different ways in which languages represent the relationships between core NPs and the verbal predicate on which they are dependent: constituent order, case-marking and verb agreement. All languages use at least one of these methods, and often more than one. Constituent order may be very free or very fixed. In languages with free constituent order (or word order), it is more likely that there will be some system of either dependent-marking (case) or head-marking (agreement) in order to identify the grammatical relation of each core NP participant. The two main case systems are the accusative and the ergative systems. Some languages, such as Chinese, have neither case nor agreement. But even in languages without morphological case, the need to recognize grammatical relations is evident in the syntax. Syntactic constructions generally follow either an ergative or an accusative pattern, the accusative being by far the most common.

FURTHER READING

Good places to start on the topics of constituent order, case and agreement would be T. Payne (2006) and Whaley (1997), moving on to Comrie (1989: Chapters 4 and

6). The seminal work on constituent order and word order is Greenberg (1966). More recent proposals can be found in Hawkins (1983) and in Tomlin (1986); see also the large body of work by Matthew Dryer, for instance Dryer (1991). On case, see Blake (2001a). On grammatical relations, see Palmer (1994) and the collection of papers in Aikhenvald et al. (2001). On agreement, see Corbett (2006). All of these are textbook treatments and are much recommended. On ergativity, Dixon (1994) is a more advanced read, but absolutely central and very worthwhile. The properties of subjects in Section 6.5.2 are largely taken from Keenan (1976); see also Comrie (1989: Chapter 5). The Warlpiri data are largely taken from the work of Ken Hale, who was a brilliant linguist and who undertook extensive fieldwork on endangered languages.

EXERCISES

1. Examine the following data in (1) through (3) (all taken from Stucky 1983). These are simple sentences from the Bantu language Makua, spoken in Tanzania, and they show that the order of phrases is very free in this language. (Makua marks both subject and object with agreement prefixes on the verb. The APPLIC suffix on the verb is an ‘applicative’ marker; it’s this that gives the sense of preparing porridge for someone, rather than an actual preposition meaning ‘for’, which marks the recipient in the English. This construction is discussed further in Chapter 7.)

- (1) Araarima aho-n-ruw-el-a mwaana isima.
 Araarima SU-OBJ-prepare-APPLIC-PAST child porridge
 ‘Araarima prepared porridge for a child.’
- (2) Isima Araarima aho-n-ruw-el-a mwaana.
 porridge Araarima SU-OBJ-prepare-APPLIC-PAST child
 ‘Araarima prepared porridge for a child.’
- (3) Aho-n-ruw-el-a Araarima mwaana isima.
 SU-OBJ-prepare-APPLIC-PAST Araarima child porridge
 ‘Araarima prepared porridge for a child.’

Each sentence contains four phrases – a subject, a verb, a direct object and an indirect object – but they appear in a different order. In fact, any of the 24 (!) possible orders of the four phrases can be used, given the right context.

Now consider complex sentences: given a subject, a verb and an embedded clause, there are six logically possible orders of these three phrases. However, only three out of the potential six orders are grammatical. The orders actually found in Makua are:

- Subject-verb-embedded clause (4).
- Verb-embedded clause-subject (5).

- and in addition, verb-subject-embedded clause (I haven't illustrated this, but you should be able to reconstruct it).
- (4) Araarima aheew-a [wiira nt'u aho-thek-a iluwani].
 Araarima SU.hear-PAST that someone SU.build-PAST fence
 'Araarima has heard that someone built a fence.'
- (5) Aheew-a [wiira nt'u aho-thek-a iluwani] Araarima.
 SU.hear-PAST that someone SU.build-PAST fence Araarima
 'Araarima has heard that someone built a fence.'

Task: Work out what the three unattested (= non-occurring) phrase orders are and state the generalization about possible phrase orders in Makua. In order to do this, you'll need to look at what the three attested orders and then the three unattested orders have in common. Why might a language have such a restriction, do you think?

2. Examine the data that follow (slightly adapted from Van Valin 1985) from Lakhota (a native American language, specifically a Siouan language of South Dakota, Montana and Manitoba) and answer questions (i) through (iv).
- i. Which argument(s) of the verb, if any, does the verb agree with?
 - ii. How is agreement (or cross-referencing) indicated in Lakhota?
 - iii. Using the data in (1) through (3) as comparison, try to figure out why (4) and (5) are grammatical, but (6) is ungrammatical. The notation '≠' indicates that the Lakhota form is not a possible way of translating the English sentence given.
 - iv. In light of your answer to (iii), why do you think (7) is ungrammatical? What generalization can be made about the grammatical vs. ungrammatical examples?

- (1) wičháša ki mathó wə Ø-Ø-kté
 man the bear a 3SG.OBJ-3SG.SU-kill
 'The man killed a bear.'
- (2) mathó wə wičháša ki Ø-Ø-kté
 bear a man the 3SG.OBJ-3SG.SU-kill
 'A bear killed the man.'
- (3) wičháša ki mathó óta wičhá-Ø-kté
 man the bear many 3PL.OBJ-3SG.SU-kill
 'The man killed many bears.'
- (4) wičháša ki ixʔé óta Ø-yəke
 man the rock many 3SG.SU-see
 'The man saw many rocks.'

- (5) wičháša ki mathó óta wičhá-Ø-yąke
 man the bear many 3PL.OBJ-3SG.SU-see
 'The man saw many bears.'
- (6) *wičháša ki ixʔé óta wičhá-Ø-yąke
 man the rock many 3PL.OBJ-3SG.SU-see
 (≠'The man saw many rocks.')
- (7) *ixʔé ki hená hokšíla wą Ø-pi-phá
 rock the those boy a 3SG.OBJ-3PL.SU-hit
 (≠'Those rocks hit a boy.')

3. In Welsh, the verb agrees with one of its argument NPs, but the conditions on this agreement are somewhat different than in more familiar European languages such as English, French or German. Study the following data, and answer these questions.
- Which NP argument does the verb agree with in Welsh? (Name its GRAMMATICAL RELATION.)
 - What MORPHOSYNTACTIC CATEGORIES of the NP does the verb agree with?
 - What are the restrictions on this agreement?
 - Why are (3), (4) and (8) ungrammatical?
 - How could you change (4) to make it grammatical, while retaining the meaning?

Hints

- All data given here are entirely regular, and no data are missing. You have enough information to answer without having to make guesses.
- Welsh has VSO (verb-subject-object) word order, but this is not relevant to your answer.

- (1) Gwelodd y bachgen ddreigiau.
 see.PAST.3SG the boy dragons
 'The boy saw dragons.'
- (2) Gwelodd y bechgyn ddreigiau.
 see.PAST.3SG the boys dragons
 'The boys saw dragons.'
- (3) *Gwelson y bechgyn ddreigiau.
 see.PAST.3PL the boys dragons
 (≠'The boys saw dragons.')
- (4) *Gwelson ein ffrindiau ddreigiau.
 see.PAST.3PL our friends dragons
 (≠'Our friends saw dragons.')

- (5) Gwelais i ddreigiau.
see.PAST.1SG I dragons
'I saw dragons.'
- (6) Gwelodd hi/o ddreigiau.
see.PAST.3SG she/he dragons
'She/he saw dragons.'
- (7) Gwelson nhw ddreigiau.
see.PAST.3PL they dragons
'They saw dragons.'
- (8) *Gwelodd nhw ddreigiau.
see.PAST.3SG they dragons
(≠ 'They saw dragons.')

4. Examine the data in (1) through (3) that follow (from Blake 1977) and determine what case system is found in the Australian language Yalarnnga: either nominative/accusative alignment ($S = A \neq O$) or ergative/absolutive alignment ($S = O \neq A$). Make clear what the evidence is for your conclusion. I have indicated the different case-markers on the NPs in the gloss by marking one case with **X** and the other with **Y**.

- (1) ngia wakamu
I.X fell
'I fell.'
- (2) kupi-ngku ngia tacamu
fish-Y I.X bit
'A fish bit me.'
- (3) nga-tu kupi-∅ walamu
I-Y fish-X killed
'I killed a fish.'

5. Examine the following data from Swahili in (1) through (4) (from Dixon 1994).

Task: (i) Determine first whether verbal agreement in this language represents a nominative/accusative system or an ergative/absolutive system. Make clear what the evidence is for your conclusion.

Hint:

Remember that the same form may sometimes be used for marking a particular person/number combination in more than one case, as in the Abaza data in (36) and (37) in the text of Chapter 6.

- (1) tu-li-anguka
1PL-PAST-fall
'We fell down.'

- (2) m-li-anguka
2PL-PAST-fall
'You all fell down.'
- (3) m-li-tu-ona
2PL-PAST-1PL-see
'You all saw us.'
- (4) tu-li-wa-ona
1PL-PAST-2PL-see
'We saw you all.'

(ii) Now describe the position of the agreement affixes in Swahili as concisely and accurately as you can, using the correct terminology.

6. Examine the data in (1) through (6) (from Anderson 1976 and Otsuka 2005) and determine what case system is found in Tongan, either nominative/accusative alignment ($S = A \neq O$) or ergative/absolutive alignment ($S = O \neq A$). Make clear exactly what the evidence is for your conclusion. Finally, describe carefully how case is represented in Tongan. In this exercise, I've simply indicated all the case-markers with the *same* gloss, CASE. Of course, they are not all the same, and different markers have different roles in the clause. You will need to work out for yourself which is which, by figuring out the role of each marker! (The character that looks like a quotation mark is a letter of the alphabet in Tongan, and represents a specific consonant, a glottal stop. It has no relevance to the answer here.)

- (1) na'e lea 'a e talavou
PAST speak CASE the young.man
'The young man spoke.'
- (2) na'e ma'u 'e sione 'a e ika
PAST get CASE Sione CASE the fish
'Sione got the fish.'
- (3) na'e alu 'a tevita ki fisi
PAST go CASE David to Fiji
'David went to Fiji.'
- (4) na'e tamate'i 'a kolaiate 'e tevita
PAST kill CASE Goliath CASE David
'David killed Goliath.'
- (5) na'e ma'u 'e siale 'a e me'a'ofa
PAST get CASE Charlie CASE the gift
'Charlie received the gift.'
- (6) na'e kai 'a e ika 'e sione
PAST eat CASE the fish CASE Sione
'Sione ate the fish.'

7. For this exercise, it will help you to revise Sections 6.5.2 and 6.5.3, on subjects. Recall from this section that in Icelandic, the subjects of some verbs take what is known as ‘quirky’ case. Now examine the Icelandic data in (1) through (3) (the data and arguments on which this exercise is based are from Sigurðsson 1991). You will see that the QUANTIFIER in bold type meaning ‘all’ (which ‘quantifies’ the number of boys) agrees in case with the subject of the clause, as well as in number (plural, here) and gender (masculine, here):

- (1) Strákarnir komust **allir** í skóla.
 the.boys.NOM got all.NOM.PL.M to school
 ‘The boys all managed to get to school.’
- (2) Strákana vantaði **alla**.
 the.boys.ACC lacked all.ACC.PL.M
 ‘The boys were all absent.’
- (3) Strákunum leiddist **öllum**.
 the.boys.DATIVE bored all.DATIVE.PL.M
 ‘The boys were all bored.’

Next, examine the data in (4) through (6). These examples are parallel to the construction from Lezgian discussed in Section 6.5.3, and it will help you to revise this particular section.

Tasks: (i) How can we account for the case-marking (as well as the number and gender marking) found on the quantifier meaning ‘all’ in each of the examples in (4) through (6)? (ii) What does the quantifier agree with?

- (4) Strákarnir vonast til [að komast **allir** í skóla.]
 the.boys.NOM hope for to get all.NOM.PL.M to school
 ‘The boys hope to all get to school.’
- (5) Strákarnir vonast til [að vanta ekki **alla** í skólann.]
 the.boys.NOM hope for to lack not all.ACC.PL.M to the.school
 ‘The boys hope to not all be absent from school.’
- (6) Strákarnir vonast til [að leiðast ekki **öllum** í skóla.]
 the.boys.NOM hope for to bore not all.DATIVE.PL.M to school
 ‘The boys hope to not all be bored in school.’

The active (a) sentences all have a transitive verb – a verb that has a subject and a direct object. By contrast, the passive (b) sentences all have only a subject, and no object: they have become intransitive. The NP that was the original subject in the active sentences (*Kim, we, three cups of tea*) has been DEMOTED in the passive: it is no longer a subject, but instead appears inside a PP headed with *by*. This means that it's no longer a core NP. In fact, the original subject of the active sentence doesn't necessarily appear in the passive sentence at all: we can also say simply *Some great photos were taken with that old camera, That Ming vase was stolen yesterday* and *The nurse has been revived*. The NP that was the original object in the active (a) sentences has been PROMOTED in each (b) sentence, becoming the subject of the passive sentence. Finally, the verbs in the passive sentences differ in form from the verbs in the active sentences. The passives all contain the PAST PARTICIPLE form of the verb: *taken, stolen, revived*; and they all contain a form of *be* as an auxiliary (in bold): *were taken, was stolen, has been revived*.

The passive in English can therefore be recognized by the following signs:

- Subject of the active sentence > demoted to a *by*-phrase or deleted; removed from the core.
- Object of the active sentence > promoted to subject of the passive.
- Passive contains auxiliary *be* + past participle of the main verb.

How do we know for sure that the NPs *some great photos, that Ming vase* and *the nurse* really are the subjects of the passive sentences? After all, each of these NPs has the semantic role of THEME (OR PATIENT) – what has been taken, what has been stolen, the person being revived – and this is the role more usually associated with objects. We can tell that these NPs in the (b) sentences nonetheless are subjects because they trigger SUBJECT/VERB AGREEMENT, which, as we saw in Section 2.3.2, is one of the diagnostic properties for subjects in English. In (1b), the subject *some great photos* is plural, so we get **were taken**, while in (2b) the subject *that Ming vase* is singular, so we get **was stolen**. The other test for subjecthood in English discussed in Chapter 2 was pronominal case: first and third person pronouns have a special form (nominative case) when they are subjects: *I, we, he, she, they*. The subject of the active sentence in (2a) is *we*, but in the passive, (2b), *us* does not have nominative case, so is no longer a subject. And the subject of (3b), *the nurse*, could be replaced by the nominative pronoun *he* or *she*, so confirming that this is a subject position.

Although not all languages have a passive construction, it is extremely common in a wide variety of languages. Basic passive constructions in all languages are formed from transitive verbs. There are two hallmarks of the passive. First, the CORE arguments of a transitive verb – its subject and object – both undergo changes in their grammatical relations. Specifically, the object of the active sentence is promoted to be the subject of the passive sentence, while the subject of the active sentence is either removed altogether in the passive (as in *Some great photos were taken with that old camera*) or else is simply demoted. 'Demotion' here means that the NP is still present, but is no longer one of the core arguments of a transitive verb (subject/

object). Instead, the former subject becomes an OBLIQUE argument – for instance, it appears inside a PP, such as the *by*-phrase in English; oblique arguments are never subjects or objects, but instead occur in less prominent positions of the clause. Second, the verb has changed its valency: the number of core arguments that it takes (see Section 2.2.2.3). Verbs signal this by changing their own form in some way. For instance, in English we find *took* becoming *were taken* in the example (1). To summarize, the prototypical passive construction has the following properties cross-linguistically.

The passive construction

- Applies to a transitive clause (the active clause) and forms an intransitive clause.
- Object promoted > subject.
- Former subject demoted > oblique argument, or is deleted; removed from the core.
- Changes occur in the morphology of the verb to signal passivization.

In English, as in numerous other European languages, there is no specifically passive form of the verb: the two distinguishing features of the passive construction, namely auxiliary *be* and the past participle verb form (*seen, stolen, played* etc.) both occur separately in different constructions: for instance, *I was singing; We've stolen them*. So neither auxiliary *be* nor the past participle alone indicates a passive construction in English: only when they occur together do we have a passive.

Examples (4) and (5) illustrate languages which, like English, have an auxiliary-plus-main-verb kind of passive. The (a) sentences are active, the (b) ones passive, and the auxiliary verbs are in bold.

- (4) a. Der Frost verdarb den Apfel. (German)
 the.NOM frost spoil.PAST the.ACC apple.
 'The frost spoils the apple.'
- b. Der Apfel **wurde** vom Frost verdorben.
 the.NOM apple became by.the.DATIVE frost spoil. PAST PARTICIPLE
 'The apple was spoiled by the frost.'
- (5) a. Eglurodd y darlithydd y sefyllfa. (Welsh)
 explain.PAST the lecturer the situation
 'The lecturer explained the situation.'
- b. **Cafodd** y sefyllfa ei egluro (gan y darlithydd).
 get.PAST the situation its explain.INFIN by the lecturer
 'The situation was explained (by the lecturer).'
 (*Literally, 'The situation got its explaining by the lecturer.'*)

As (4b) shows, some other languages also use the past participle form of the verb in the passive construction, but this is by no means universal. Welsh, for instance, has no past participle, and the main verb just has one non-finite form, which is used in

active sentences as well as in passives such as (5b). According to Keenan (1985a), the most common auxiliaries occurring in passive constructions are verbs like ‘be’, ‘become’, ‘get’ and ‘receive’, as illustrated in (4) and (5). In fact, English also has a commonly used *get* passive, as in *My bike got stolen*.

In the German examples, we can tell that the former object of the active clause becomes the subject of the passive clause by the change in its case-marking: *den Apfel* in (4a) is accusative, the case of direct objects in German, while *der Apfel* in (4b) is nominative, the case of subjects.

Instead of the auxiliary-plus-verb kind of passive, many languages have a specifically passive form of the main verb; this is known as a MORPHOLOGICAL PASSIVE. Each language illustrated in (6) through (8) has a special passive marker on the verb, shown in bold in each (b) example. This affix is the only change in the verb form that indicates the passive. As before, all the (a) sentences are active, and the (b) sentences passive.¹

- (6) a. Si Juan ha dulalak si Jose. (Chamorro)
 PN¹ Juan 3SG.SU follow PN Jose
 ‘Juan followed Jose.’
 b. D-**in**-ilalak si Jose as Juan.
 -PASSIVE-follow PN Jose by Juan
 ‘Jose was followed by Juan.’
- (7) a. Neko-ga sakana-o tabeta. (Japanese)
 cat-NOM fish-ACC eat.PAST
 ‘The cat ate the fish.’
 b. Sakana-ga neko-ni tabe-**rare**-ta.
 fish-NOM cat-DATIVE eat-PASSIVE-PAST
 ‘The fish was eaten by the cat.’
- (8) a. E kamate-a te naeta te moa. (Gilbertese)
 it kill-it the snake the chicken
 ‘The chicken killed the snake.’
 b. E kamate-**aki** te naeta (iroun te moa).
 it kill-PASSIVE the snake by the chicken
 ‘The snake was killed (by the chicken).’

Note also here that in Japanese, a language with nominative/accusative case-marking, we again see the changes in case that result from the promotion of the object to the subject position, and the demotion of the erstwhile subject. In the passive in (7b), the ‘fish’ NP *sakana* has become nominative, the case of subjects in Japanese, and the ‘cat’ NP *neko* has been demoted from subject position to an oblique (i.e. non-core) position, marked by dative case.

¹ The abbreviation PN in (6) is for ‘proper noun marker’, that is, it marks names in Chamorro. Note also that the passive marker *-in-* is actually an INFIX on the verb in (6b): it’s inserted into the stem of the verb itself.

In fact, passive constructions occur most typically in languages which, like German or Japanese, are syntactically and morphologically accusative in their alignment. Recall from Chapter 6 that this gives rise to languages which have a definite *subject* grammatical relation, and which generally also have case-marking and/or verbal agreement which patterns according to the nominative/accusative alignment. Thus, accusative systems treat all subjects the same way (A plus S noun phrases), and treat objects differently (O noun phrases): $S = A \neq O$.

But what about the passive in ergative/absolutive languages, which group S and O arguments (the ABSOLUTIVE NPs) in opposition to A arguments (the ERGATIVE NPs): $S = O \neq A$? It will help at this point to revise the discussion in Chapter 6 concerning the different ways in which NPs group together in each system. These tables should help to refresh your memory:

Table 7.1
Accusative and ergative case systems

Accusative system			Ergative system		
A	S	O	A	S	O
Nominative		Accusative	Ergative	Absolutive	

It might seem that ergative languages would not have a passive construction, since the division between all subjects and all objects found in accusative languages is much less evident, or even absent. Indeed, not all ergative languages have passives: for instance, Dyirbal and Lezgian (see Chapter 6) do not. However, a number of ergative languages do have a passive construction, as illustrated in (9) from Inuktitut (Greenlandic), and (10) – slightly adapted – from Tzotzil. As before, the (a) sentences are active, the (b) ones passive, and the passive marker on the verb is in bold. The grammatical relations (A, O, S) are also indicated on the NPs:

- (9) a. angut-ip(A) arnaq(O) taku-vaa (Inuktitut)
 man-ERG woman.ABS see-3SG/3SG
 ‘The man saw the woman.’
 b. arnaq(S) (anguti-mit) taku-**tau**-puq
 woman.ABS man-by see-PASSIVE-3SG
 ‘The woman was seen (by the man).’
- (10) a. S-mil-ox-Ø (O)Xun li Petul-e(A) (Tzotzil)
 3SG.ERG-kill-PAST-3SG.ABS John the Peter-DEF
 ‘Peter killed John.’
 b. Mil-**bil**-Ø ju?un Petul li Xun-e(S)
 kill-PASSIVE-3SG.ABS by Peter the John-DEF
 ‘John was killed by Peter.’

In (9a), ergative/absolutive alignment is indicated in the active sentence via case marking on the NPs, the A argument being ergative, and the O argument absolutive:

in other words, standard ergative case-marking, given a transitive verb. The verb in (9a) also agrees with both its core arguments (both are third person singular). In the passive, (9b), the former ergative NP meaning ‘man’ is demoted, and appears in an optional ‘by’-phrase. Moreover, the verb is now *intransitive*, so agrees only with its remaining core argument, *arnaq*, ‘the woman’, which has become the S argument of the intransitive verb. So just as in accusative languages, the NP *arnaq* has undergone a change in grammatical relation in the passive, from O to S. However, in an ergative language, this doesn’t change the case marking of the promoted NP: the NP *arnaq* remains absolutive, because this is the case used both for O and for S. Of course, it doesn’t become ergative, since this case is reserved for the A argument of a *transitive* verb.

The Mayan language Tzotzil (spoken in Mexico) has no case-marking on the NPs themselves, but has an ergative agreement system, indicated by verbal affixes. In the active sentence in (10a), we see two verbal affixes: an ergative agreement prefix, marking the A argument *Petul*, ‘Peter’, and an absolutive agreement suffix, marking the O argument *Xun*, ‘John’. The passive construction in (10b) shows that the verb has lost the ergative prefix *s-*, since there is no longer an ergative NP for the verb to agree with: the former ergative NP *Petul* is now demoted, again appearing in a ‘by’-phrase. The passive verb has become intransitive, as in the other passives we’ve seen, and so agrees just with its one remaining core argument, the NP *Xun* ‘John’. This agreement marker is still absolutive: *Xun*, ‘John’, has changed from being an absolutive O argument in (10a) to the absolutive S in (10b) – the single argument of an intransitive verb.

Other ergative languages with a passive construction include other Mayan languages, the South Caucasian language Georgian, and the European language isolate Basque.

Since the passive construction in an ergative language doesn’t change the *case* of the original O noun phrase – it’s still absolutive when it becomes an S, as we’ve seen – then why have a passive at all? Perhaps the main effect of the passive in ergative systems is to remove focus from the original A noun phrase, in examples such as (9a) and (10a), by removing it from the core: demoting the NP to a ‘by’-phrase makes it less prominent. In fact, passives also have this same effect of defocusing the agent in accusative languages as well, as shown for English in (1) through (3) – the agent is either demoted or deleted entirely, and so becomes much less prominent. Cross-linguistically, then, passives have a common pragmatic effect: that of removing focus from the agent NP. This function holds for passives both in accusative systems and in ergative systems.

7.1.2 The impersonal construction

In this section, we will see the IMPERSONAL construction. Here, the subject argument is suppressed, which also occurs in the passive; but unlike the passive, the impersonal construction does not create a new subject (Blevins 2003). The passive construction involves verbs that are transitive, as shown in Section 7.1.1. The impersonal construction can occur with intransitive verbs, as illustrated in (11) from German; the (a) sentence is active and the (b) sentence is the impersonal:

- (11) a. Die Leute tanzten. (German)
 the people dance.PAST
 ‘The people danced.’
- b. Es wurde getanzt.
 it became dance.PAST PARTICIPLE
 ‘There was dancing.’
 (*Literally, ‘It became danced.’*)

In (11b), no core NP has been promoted to subject. The construction is thus ‘impersonal’ – in fact, the verb here has no core arguments at all. However, (11b) does have what is often called a ‘dummy’ subject, *es* ‘it’; this fills the otherwise empty subject position, but doesn’t have any intrinsic meaning. So we can still maintain that the impersonal has no true subject.

Turkish also has an impersonal construction, with no constituent in subject position. Our example sentence is formed from an ordinary transitive verb, but it is again an impersonal because – unlike the passive – it does not create a new subject. As before, the (a) sentence is active and the (b) one the impersonal:

- (12) a. Hasan dün bütün gün kitap oku-du (Turkish)
 Hasan yesterday whole day book read-PAST
 ‘Hasan read books all day yesterday.’
- b. Dün bütün gün kitap oku-n-du
 yesterday whole day book read-IMPERSONAL-PAST
 ‘Yesterday books were read all day.’
 (*Literally, ‘Book reading was done all day yesterday.’*)

We can tell in Turkish that the object of the active sentence, *kitap*, ‘book’, has not been promoted to subject position in (12b) because it must remain in the standard direct object position that it occupies in (12a), which is immediately preceding the verb.

As with the passive, an important function of the impersonal is to remove focus from the former agent by demoting or deleting the subject NP. The difference is that no other NP is promoted to subject in an impersonal construction. Impersonals are quite widespread, occurring for instance in Dutch, Latin and – outside Indo-European – Turkish, Shona (Bantu) and Tarahumara (Uto-Aztecan). If a language has an impersonal construction, then it will also have an ordinary ‘personal’ passive construction of the type illustrated in Section 7.1.1, which does involve the creation of a new subject.

7.2 THE ANTIPASSIVE

7.2.1 Basic facts

In Section 7.1.1, we saw that both accusative and ergative languages can have a passive construction, although the passive is certainly found more commonly in accusative

languages than in ergative ones. However, another construction which changes grammatical relations also occurs in ergative languages. This is known as the ANTIPASSIVE, and this does not occur in accusative languages. Like the passive, the antipassive also takes a transitive clause and makes it intransitive via a process of promotion of one NP and demotion of another. I will focus first on the DEMOTION effects of the antipassive. Compare the ordinary active sentence in (13a) with the antipassive version in (13b), both from Inuktitut (Greenlandic); the antipassive marker is in bold:

- (13) a. arna-p(A) niqi(O) niri-vaa (Inuktitut)
 woman-ERG meat.ABS eat-3SG/3SG
 ‘The woman ate the meat.’
- b. arnaq(S) niqi-mik niri-**NNig**-puq
 woman.ABS meat-with eat-ANTIPASSIVE-3SG
 ‘The woman ate some of the meat.’

In (13a), the ‘woman’ NP *arnap* is the A argument of a transitive verb, and is therefore ergative, while the ‘meat’ NP *niqi* is the O argument of the transitive verb, and is therefore absolutive: this is the standard ergative case alignment discussed in Chapter 6. In the antipassive sentence in (13b), the former O argument *niqi* ‘meat’ is now DEMOTED. It is no longer a core argument of the verb, but is instead an oblique NP: the suffix *-mik* in fact indicates what is known as ‘instrumental’ case, which I’ve glossed as ‘with’. The effect of this demotion is to give the ‘meat’ NP a PARTITIVE reading – the woman ate *some* of or *part* of the meat, as indicated in the translation. Since the verb in (13b) is no longer transitive (in Inuktitut), the NP *arnaq* ‘woman’ is the single S argument of an intransitive verb, and so takes the absolutive case. I recommend that you re-read this section up to this point before moving on.

The antipassive construction has a variety of functions in ergative languages, including giving rise to a partitive reading as shown earlier. Consider first the pair of sentences from the Siberian language Chukchee in (14); the (a) sentence is active, the (b) antipassive, with the antipassive marker in bold:

- (14) a. ətləg-e(A) keyng-ən(O) penrə-nen (Chukchee)
 father-ERG bear-ABS attack-3SG/3SG.PAST
 ‘Father attacked the bear.’
- b. ətləg-ən(S) penrə-**tko-g**?e keyng-etə
 father-ABS attack-ANTIPASSIVE-3SG.PAST bear-DATIVE
 ‘Father ran at the bear.’

In the active sentence, (14a), the ‘father’ NP *ətləge* is the A argument of a transitive verb, and hence is marked with ergative case, while the ‘bear’ NP *keyngən* is an O, the object of a transitive verb, and hence is marked with absolutive case. The verb agrees with both these core arguments in (14a): it has a 3SG agreement for each of them (fused into a single marker, along with the past tense morpheme). The antipassive again has the effect of demoting the former object: the ‘bear’ NP *keyng* in (14b) has become dative, and we get the effect of *running at* the bear rather than attacking it. The

The active construction in (16a) has a transitive verb, and the clause has the standard case marking in the ergative alignment: an ergative A noun phrase, *yinydyuun bunyaan* ‘this woman’, and an absolutive O noun phrase, *mayi* ‘vegetables’. The antipassive construction in (16b) has only one argument, *yinu bunya* – the absolutive S argument of what is now an intransitive verb – and the former O noun phrase is simply deleted.

7.2.2 Primary grammatical relations and grammatical pivots

So far, we have considered antipassives in which the main effects of the construction are on the O argument of the active verb: this NP has been demoted so it’s no longer a core argument of the verb, or it’s been deleted entirely. However, another equally important use of the antipassive in ergative languages involves the PROMOTION of the A noun phrase – the ergative ‘subject’ in the transitive clause – to be an S: an absolutive ‘subject’ in an intransitive clause. It may surprise you to think of this as promotion. In the more familiar accusative languages, it’s easy to see how the passive construction, which changes the grammatical relation of an object NP and makes it the subject, is a process of promotion – consider the difference between the active *A crocodile ate my friend* and the passive *My friend was eaten by a crocodile*. Any native speaker of English would agree that the passive focuses on what happened to the friend in a way the active does not – indeed, the active can sound truly callous!

But why is A > S a promotion? Recall from Section 6.3 that in both accusative systems and ergative systems it’s the noun phrase that appears as the S argument which is the most basic in usage. Whether it’s a nominative NP, as in accusative systems, or an absolutive NP, as in ergative systems, the S is generally unmarked in both form (case-marking) and function (syntactic constructions). Following Palmer (1994), we can say that the S is always a PRIMARY grammatical relation. In accusative systems, of course, S groups with A to give SUBJECT as the primary grammatical relation, while in ergative systems, S groups with O to give ABSOLUTIVE as the primary grammatical relation.

Table 7.2

Primary grammatical relations

Primary grammatical relations	
Accusative systems	S + A = Subject NPs
Ergative systems	S + O = Absolutive NPs

The passive construction is mostly found in the accusative alignment, while the antipassive occurs exclusively in ergative systems. Both passive and antipassive constructions have the effect of creating a new S argument. The passive does this by promoting O > S, and the antipassive does it by promoting A > S. So both constructions have the effect of making a NONPRIMARY NP into a primary NP: the nonprimary NPs are O in accusative systems, A in ergative systems. Let’s see now what sort of effects this has in ergative systems.

In the Mayan language Mam (Guatemala and Mexico), the verb is initial in the clause in the basic constituent order, but an NP can be focused by FRONTING it to the start of the clause. However, the only NPs that can undergo fronting are the two absolutive NPs, the S and the O – the two NPs which form the primary grammatical relation in an ergative language. Examples (17) and (18) illustrate this fronting, in an intransitive and a transitive clause respectively. The fronted NP is shown in bold in each sentence. Verbal agreement markers occur in each example: you can tell which NP the ergative and absolutive markers refer to by the fact that ‘the man’ NP, *xiinaq*, is always indicated by a 3SG marker, while ‘the horses’ NP *qacheej* is 3PL. There is no ergative case-marking in this language; the ergativity is shown via verb agreement:

- (17) **xiinaq**(S) s-uul (Mam)
 man ASPECT.3SG.ABS-arrive.here
 ‘The man arrived here.’
- (18) **qa-cheej**(O) x-hi kub’ t-tzyuun xiinaq(A)
 PL-horse ASPECT-3PL.ABS DIRECTION 3SG.ERG-grab man
 ‘The man grabbed the horses.’

Examples (17) and (18) are standard active clauses for an ergative language. In an intransitive clause, the S argument is absolutive, and triggers verb agreement, which in (17) is shown by an affix *s-* on the verb. In the transitive clause in (18), the O argument meaning ‘the horses’ is absolutive, and triggers a verbal agreement marker *-hi*; the A argument *xiinaq* ‘the man’ is ergative, and triggers a verbal agreement marker *t-*. What if a speaker wants to focus on the A noun phrase in (18), *xiinaq* ‘the man’? As (19) shows, it’s not possible to do this by simply fronting *xiinaq* in the ordinary active sentence: the result is ungrammatical, because *xiinaq* is an A, not an S or an O, and so is not a primary grammatical relation:

- (19) ***xiinaq**(A) chi kub’ t-tzyuun qa-cheej(O)
 man 3PL.ABS DIRECTION 3SG.ERG-grab PL-horse
 (≠‘The man grabbed the horses.’)

Instead, the ergative NP *xiinaq* must first be promoted to be absolutive – becoming a primary NP – so it can then be fronted. This promotion from A > S is achieved by using the antipassive construction:

- (20) **xiinaq**(S) x-Ø-kub’ tzyuu-n t-e qa-cheej
 man ASPECT-3SG.ABS-DIRECTION grab-ANTIPASSIVE 3-OBLIQUE PL-horse
 ‘The man grabbed the horses.’

In (20), we find the grammatical version of what (19) was unable to express. The former O argument of the transitive clause, *qacheej* ‘horses’, is demoted in (20): it is no longer an O in the Mam sentence, but has become an oblique NP, as is indicated by the oblique marker that precedes it (like a preposition). This means that *xiinaq*, ‘the man’, is now the single argument of an intransitive verb, an S, and so is absolutive

and can be focused. We can tell from the verb agreement (3SG absolutive, agreeing with *xiinaq*) that the promotion has taken place. Hence, the antipassive serves here to allow an NP to be focused where it otherwise couldn't be.

Next we see a second construction requiring the antipassive to promote an NP from ergative to absolutive, from an Australian language, Dyirbal. This involves the CO-ORDINATION of clauses. First, some reminders of facts from a typical accusative language, English. In Section 6.5.2, we saw that a subject can undergo ellipsis (= omission) in the second of two conjoined clauses. You can look back now for revision of this. The subscript index $_i$ or $_j$ shows which NP in the first clause the omitted NP, designated \emptyset , refers back to:

- (21) a. Chris woke up and (Chris) saw Lee.
 b. Chris_i disturbed Lee and \emptyset_i complained bitterly.
 c. *Chris disturbed Lee_i and \emptyset_i complained bitterly.
 d. Chris_i greeted Lee and then \emptyset_i kissed Mel.
 e. *Chris_i greeted Lee_j and then Mel kissed \emptyset_{ij} .

What these examples show is that in accusative languages like English, the ellipsis revolves around *subjects*. So, for instance, (21b) can only mean that it was Chris who complained, and (21c) cannot mean that Lee complained. The grammatical sentences, (21a), (21b) and (21d), show that a subject can undergo ellipsis in the second clause, but only when it's coreferential with (= refers back to) the subject of the first clause. As for the ungrammatical sentences, (21c) shows that an omitted subject can't refer back to the object of the first clause – which is why (21c) can't mean that Lee complained; and (21e) shows that it's only a subject which is omitted in English, and not an object, so that (21e) is ungrammatical whatever the omitted NP refers back to.

We can say that accusative languages which operate like English does in (21) have a SUBJECT PIVOT, comprising the two primary NPs – those with the grammatical relations S and A. A PIVOT links noun phrases together across different clauses, for instance as seen in (21), by allowing one NP to be omitted providing it can refer back to another NP in the first clause. Not all languages have any syntactic restrictions on the interpretation of NPs across clauses. This means that two clauses can be linked together and *any* NP which is repeated can be omitted. In such languages, the equivalent to any of the examples in (21) should be perfectly grammatical in the appropriate context. Languages of this kind don't have a syntactic pivot. In languages that do have a syntactic pivot, it may operate as in English, revolving around the subject relation, or alternatively, in the case of some ergative languages, the pivot may revolve around *absolutive* NPs.

So if a language has an SA pivot, we expect constructions that link NPs to revolve around the S and the A relations. This is what happens in English. First, both S and A noun phrases – that is, all subjects – undergo ellipsis, as we can see from the fact that both an intransitive verb like *complain* and a transitive verb like *see* or *kiss* allow their subject to be omitted. And second, both the S subject of an intransitive verb like *wake up* and the A subject of a transitive verb like *disturb* or *greet* can be the NP that controls an omitted subject in the second clause. Finally, if we want to indicate what

(21c) attempts to do – namely, that it was Lee who complained – we do it by passivizing the first clause, to give *Lee, was disturbed by Chris and Ø_i, complained bitterly*. This, of course, has the effect of promoting Lee to subject position, which makes it a primary grammatical relation, so that it can now control the omitted NP in the second clause.

In a language like Dyirbal, which is syntactically ergative, however, ellipsis revolves around absolutive NPs. So Dyirbal has an ABSOLUTIVE PIVOT: this comprises the two absolutive grammatical relations, S and O, which together form the primary relation. This means that both the NP in the first clause which controls the ellipsis and the NP which undergoes ellipsis must be one of the absolutive NPs, either S or O. Let's see how this works first when ordinary active clauses are co-ordinated, starting with (22). Before you tackle the examples that follow, here are some hints to help you.

- Case is indicated in Dyirbal via a suffix on the nouns, though the absolutive is in fact unmarked (there is no absolutive inflection), while ergative and other cases such as dative each have a particular suffix.
- You can tell which NP refers to which other NP by looking at the subscripts, _i. So for instance, in (22), the NP that undergoes ellipsis in the second clause is coreferential with *nguma* 'father' in the first clause.
- Read the glosses and translations carefully and try not to let the constituent order worry you: the absolutive NP is initial in each clause, whether it's an S or an O. I've indicated the grammatical relation of the NPs here in the gloss, with a small subscript (S, O or A), and also in the translation.
- Note also that there's no actual word for 'and' in Dyirbal co-ordination. In these examples I have put each co-ordinated clause in square brackets, to help you see the start and end of the clauses.

(22) [nguma_i yabu-nggu bura-n] [Ø_i banaga-n'u] (Dyirbal)
 father.ABS_O mother-ERG_A see-PAST []_S return-PAST
 'Mother(A) saw father(O) and [he](S) returned.'

The NP that's omitted in the second clause in (22) has to refer back to *nguma*, 'father', the absolutive O noun phrase – it can't refer back to *yabunggu*, 'mother', the ergative A noun phrase. In English, this is not a possible construction: *Mother saw father and returned* can only mean that mother returned, not that father did. The only way to get that reading in English is to use a pronoun *he* in the second clause, as I've shown in the translation of (22), but crucially, there is no pronoun in the corresponding Dyirbal sentence.

In (22), the two coreferential NPs are an O in the first clause and an S in the second clause. Both are, of course, absolutive. In (23), the first clause has an S and the second clause omits an O which refers back to that S:

(23) [nguma_i banaga-n'u] [Ø_i yabu-nggu bura-n]
 father.ABS_S return-PAST []_O mother-ERG_A see-PAST
 'Father(S) returned and mother(A) saw [him](O).'

The English translation would again be impossible without the pronoun in the second clause: we don't get **Father returned and mother saw*. But again, there is no pronoun in the corresponding Dyirbal: the O argument can be omitted when it is coreferential with the S of the first clause. Both (22) and (23) show that ellipsis in Dyirbal operates in terms of the absolutive NPs, S and O, rather than with a subject pivot as in English. Dyirbal then has an absolutive pivot.

What happens, though, if a Dyirbal speaker wants to say something that means 'Mother saw father and (mother) returned'? Example (22) does not and could not mean this. Instead, the antipassive construction is used: this promotes the ergative NP meaning 'mother' in a sentence like (22) so that it *becomes* an absolutive, and as an absolutive NP it can be a pivot: it can control the ellipsis of the S in the second clause. Example (24) illustrates; the first clause is the one that's antipassive:

- (24) [yabu_i bural-nga-n'u nguma-gu] [∅_i banaga-n'u]
 mother.ABS_S see-ANTIPASSIVE-PAST father-DATIVE []_S return-PAST
 'Mother(S) saw father and (S) returned.'

In the first clause of (24), what in an ordinary active clause such as (22) would be the O – the object of a transitive verb – has now been demoted: the 'father' NP *ngumagu* is now dative, and the verb is detransitivized with the antipassive suffix. The remaining NP, *yabu* 'mother', is therefore the S argument of an intransitive verb meaning 'see'. As an S, it is absolutive, and so can be a pivot: it allows the omitted NP in the second clause to refer back to it. So the antipassive construction serves to make an NP available as a pivot; here, as the controller of ellipsis.

Second, the antipassive can make an NP into a pivot so it is available to *undergo* ellipsis. This is shown in (25), where this time the second clause has become antipassive, in order to get the reading 'Father returned and saw mother'.

- (25) [nguma_i banaga-n'u] [∅_i bural-nga-n'u yabu-gu]
 father.ABS_S return-PAST []_S see-ANTIPASSIVE-PAST mother-DATIVE
 'Father(S) returned and (S) saw mother.'

In the second clause, the 'mother' NP *yabugu* is not a primary NP but has been demoted, as we can tell by its dative case. The antipassive verb meaning 'see' is again detransitivized: it has only one core argument, the S noun phrase – the single argument of an intransitive verb. As an S, this NP is a possible pivot, so allowed to undergo ellipsis when coreferential with another absolutive NP. So the empty S position in (25) refers back to *nguma*, 'father', in the first clause.

To summarize, the antipassive construction has the following characteristics cross-linguistically:

The antipassive construction

- Applies to a transitive clause (the active clause) and forms an intransitive clause.
- The A argument (ergative) promoted > S argument (absolutive).

- O argument demoted > oblique, or is deleted.
- Changes in the morphology of the verb signal antipassivization.

Both the passive and the antipassive constructions have in common the fact that they change basic grammatical relations by promoting some NPs and demoting others. This results in changes to the valency of the verb. The following two sections introduce two other grammatical relation-changing processes: the applicative and the causative constructions. Like the passive and antipassive, these do not occur in all languages, but are widespread nonetheless.

7.3 THE APPLICATIVE CONSTRUCTION

English has an alternation between the (a) and (b) forms in sentences like (26) and (27). Let's assume that the (a) sentences are the more basic, and the (b) sentences are derived from them by processes of promotion and demotion. (One reason for taking the NP-PP constructions as in (26a) and (27a) to be the more basic is that not all verbs which take NP and *to/for*-PP complements can undergo the alternation: **I dispatched the children the presents* vs. *I dispatched the presents to the children*.)

- (26) a. My brother sold his bike to Sue.
 b. My brother sold Sue his bike.
- (27) a. I baked a cake for Kim.
 b. I baked Kim a cake.

This alternation occurs just with certain three-argument verbs in English. In their basic form, these verbs take a direct object NP (such as *his bike*, *a cake*) plus a PP headed by *to* or *for*, such as *to Sue*, *for Kim*. In the (b) sentences, the NPs *Sue* and *Kim* have been PROMOTED to direct object position – immediately following the verb in English – and the original direct object is DEMOTED to become a second object: there is no longer a PP in the (b) sentences. This construction in English is often known as DATIVE MOVEMENT (although English has no actual dative case-marking) because, in some languages, indirect objects such as ‘to Sue’ are marked dative (see Section 6.5.4).

Now compare the parallel construction found in two completely unrelated languages (unrelated both to one another and to English): an Austronesian language, Indonesian, and a Bantu language, Chichewa (the rather strange-sounding examples from this language are taken from Baker 1988). We examine later the APPLIC (standing for APPLICATIVE) affixes shown in bold type on the verb in the (b) sentences:

- (28) a. Mereka mem-bawa [daging itu] [kepada dia]. (Indonesian)
 they TRANS-bring meat the to him
 ‘They brought the meat to him.’

- b. Mereka mem-bawa-**kan** [dia] [daging itu].
 they TRANS-bring-APPLIC him meat the
 ‘They brought him the meat.’
- (29) a. Mbidzi zi-na-perek-a msampha kwa nkhandwe. (Chichewa)
 zebras SU-PAST-hand-ASPECT trap to fox
 ‘The zebras handed the trap to the fox.’
- b. Mbidzi zi-na-perek-**er**-a nkhandwe msampha.
 zebras SU-PAST-hand-APPLIC-ASPECT fox trap
 ‘The zebras handed the fox the trap.’

These constructions involve the same changes in grammatical relations as those found in English in (26) and (27). In (28), the NP *dia* ‘him’, which is originally part of a ‘to’-PP *kepada dia* in (a), is promoted in (b) to become the direct object – as in English, this immediately follows the verb in Indonesian. The preposition disappears. The NP *daging itu* becomes a second object. In Indonesian, but not in English, there is also a special marker on the verb to indicate the promotion: the suffix *-kan*. This is glossed as *APPLICATIVE*, a traditional grammatical term used both for the verbal marker of promotion and for the construction as a whole.

The Chichewa applicative in (29) is exactly parallel: the ‘fox’ NP *nkhandwe* was an indirect object within a PP in (29a), but is promoted to direct object position in (29b). The original direct object in (29a), *msampha* ‘trap’, is demoted in (29b), becoming a second object, and again there’s an applicative marker on the verb, the suffix *-er*.

The general properties of the applicative construction, including English dative movement, can be summarized as follows:

The applicative construction

- Oblique NP or indirect object > promoted to object.
- Former object > demoted to second object or oblique.
- Changes may occur in the morphology of the verb to signal the applicative construction.

English is fairly restrictive in the type of oblique phrase that can undergo promotion, but cross-linguistically various kinds of oblique phrase can be promoted, including locative expressions (= those involving location, such as ‘on the table’, ‘into the water’), goals (as in *We sent the letter to Mel* > *We sent Mel a letter*), beneficiaries (as in *I baked a cake for Kim* > *I baked Kim a cake*) and instrumental phrases, such as ‘with a stick’, as in the Dyirbal example in (30):

- (30) a. yabu nguma-nggu balga-n yugu-nggu (Dyirbal)
 mother.ABS_O father-ERG_A hit-PAST stick-INSTRUMENTAL
 ‘Father hit mother with a stick.’
- b. yugu nguma-nggu balgal-**ma**-n yabu-gu
 stick.ABS_O father-ERG_A hit-APPLIC-PAST mother-DATIVE
 ‘Father used a stick to hit mother.’

Example (30a) is an ordinary transitive clause in Dyirbal, with an ergative A noun phrase, *ngumanggu*, meaning ‘father’, and an absolutive O noun phrase, *yabu*, meaning ‘mother’. In the English translation, *stick* appears inside a PP headed by *with* – it’s an oblique phrase; in Dyirbal, the ‘stick’ NP *yugunggu* is also oblique, and this is marked by a special INSTRUMENTAL case.² Instrumental NPs don’t undergo dative movement in English, whereas in Dyirbal the ‘stick’ NP can indeed be promoted to become a core argument: it’s the O in (30b). This NP *yugu* now has absolutive case – the case of normal objects in ergative systems – while the former O noun phrase *yabu*, ‘mother’, has been demoted, as shown by its dative case marking: *yabugu*.

Finally, an NP which has been promoted by the applicative construction to become a direct object can generally undergo a second promotion by the passive construction, thus becoming a subject. In fact, we have already seen an example of this in Section 1.1.1, in the discussion comparing English and Indonesian. The examples in (31) and (32) are again from Chichewa (English speakers may not find the translation of (31b) grammatical):

- (31) a. Kalulu a-na-gul-ir-a mbidzi nsapato (Chichewa)
 hare SU-PAST-buy-APPLIC-ASPECT zebras shoes
 ‘The hare bought shoes for the zebras.’
 (more literally, ‘The hare bought the zebras shoes.’)
- b. Mbidzi zi-na-gul-ir-idw-a nsapato (ndi kalulu)
 zebras SU-PAST-buy-APPLIC-PASSIVE-ASPECT shoes by hare
 ‘The zebras were bought shoes by the hare.’

In (31a), *mbidzi* ‘zebras’ has already undergone promotion by the applicative construction, and has become the direct object: as in English, the direct object immediately follows the verb. Once promoted to direct object position, the NP *mbidzi* can undergo a further promotion in the passive construction, (31b): it becomes the subject. The former subject *kalulu* ‘hare’ is demoted to an optional *ndi* (‘by’)-phrase. Crucially, the ‘shoes’ NP in (31a), *nsapato*, cannot undergo promotion to subject by the passive construction, because it’s not the direct object but a second object. We can tell that *nsapato* is not a direct object by the fact that it doesn’t immediately follow the verb. If we try to promote the second object in a passive construction, the result is ungrammatical, as in (32):

- (32) *Nsapato zi-na-gul-ir-idw-a mbidzi (ndi kalulu)
 shoes SU-PAST-buy-APPLIC-PASSIVE-ASPECT zebras by hare
 ‘*Shoes were bought the zebras by the hare.’

² The instrumental case in Dyirbal, in fact, has the same suffix as the ergative case, *-nggu*, but there are good reasons to consider the two cases to be syntactically distinct. Dixon (1994: 170, fn. 22) notes that instrumental NPs and ergative NPs have different syntactic behaviour. In the antipassive construction, an ergative NP is promoted from A to S – see (24) and (25) above – but an instrumental NP doesn’t get promoted. In the applicative construction, an instrumental NP gets promoted to O whereas an ergative NP undergoes no promotion.

So in Chichewa – and in English – only an NP which is, *or has become*, a direct object can undergo promotion by the passive. Although this restriction is very common cross-linguistically, it's not universal: in some languages both the direct object and the second object of an applicative construction behave like a prototypical object. In Kinyarwanda – another Bantu language – for instance, either type of object can be promoted to subject by the passive construction (see Palmer 1994: Chapter 6.6).

Cross-linguistically, it is usual to find that the applicative (or dative movement) construction feeds into the passive construction, as illustrated for Chichewa in (31b) and for English by the translation of this example. In other words, the applicative creates new direct objects that can then be promoted to subject. However, not all languages have an applicative construction. French, for example, has no construction parallel to English dative movement; so in French (33a) cannot become (33b), with promotion of *Pierre* to direct object position:

- (33) a. Marie a donné un cadeau à Pierre. (French)
 Marie has give.PAST PARTICIPLE a present to Pierre
 'Marie has given a present to Pierre.'
- b. *Marie a donné Pierre un cadeau.
 Marie has give.PAST PARTICIPLE Pierre a present
 (≠ 'Marie has given Pierre a present.')

In turn, this means that the 'dative movement' construction in (33b) is unavailable as input to the passive construction. The passive version of (33a) is (34a), which is fine – the original direct object *un cadeau*, 'a present', has been promoted to subject position. But since *Pierre* is not a possible direct object in (33b), then we'd predict that this sentence won't be a possible input to the passive construction, since the passive in French only promotes direct objects. And, indeed, the passive version of (33b), with *Pierre* promoted to subject position, is ungrammatical in French as predicted, as in (34b):

- (34) a. Un cadeau a été donné à Pierre par Marie.
 a present has been give.PAST PARTICIPLE to Pierre by Marie
 'A present has been given to Pierre by Marie.'
- b. *Pierre a été donné un cadeau par Marie.
 Pierre has been give.PAST PARTICIPLE a present by Marie
 (≠ 'Pierre has been given a present by Marie.')

So cross-linguistically, we find a continuum, which, at one extreme, allows no applicative constructions, as in French, and, at the other extreme, is very free in the kinds of prepositional objects and other oblique NPs that can be promoted to subject position. Chichewa lies at the latter end of the spectrum, as does Dyirbal; see (30). English falls somewhere in the middle, having dative movement with a restricted set of verbs.

7.4 THE CAUSATIVE CONSTRUCTION

So far in this chapter, we have examined constructions that change grammatical relations by promotion and demotion processes, but which don't introduce any new NP arguments. The passive and antipassive either have the same number of arguments as their active counterparts, or they may reduce that number; the *by*-phrases are optional in (1) through (3), for instance. And the applicative/dative movement construction doesn't change the number of arguments in the construction, but simply promotes one to be a core argument and demotes another. In this section, I introduce the last major construction type which changes grammatical relations: the causative. This differs from the constructions seen so far in that it always increases the verb's valency by introducing a new argument – the causative agent – and it often introduces an entire new causative predicate as well. I illustrate first from English.

In English, the main way of expressing the idea of causing someone else to do something is by using a verb such as *make*, *let*, *cause* or *have*. So we get pairs of sentences like those in (35) and (36):

- (35) a. The students left.
 b. We **made/let** the students leave.
- (36) a. The students read the book.
 b. We **had** the students read the book.

In both examples, the (a) sentences are basic, simple clauses; (35a) is intransitive, (36a) transitive. The (b) examples in each case are CAUSATIVE constructions. In both, *the students* has been DEMOTED from its original position as the subject of the simple clause, and a new subject, *we*, has been introduced. Note that this new subject hasn't been promoted from anywhere, since it doesn't exist in the (a) sentences; it arises from the causative construction. These two properties are common to causative constructions cross-linguistically: the original subject is demoted and a new subject is introduced.

The causative construction in English introduces a new subject and a new predicate – *We made/let/had* in (35) and (36) – so creating a whole new clause. This means that the causative construction turns the simple sentences (with just one clause) in (35a) and (36a) into complex sentences in (35b) and (36b).

This same kind of causative construction with a 'make' or 'cause' verb plus the basic verb also occurs in many other languages. In (37), from Korean, (37a) is the basic clause with *ku sayka* 'the bird' as subject: it has nominative case. And (37b) is the causative, with the causative verb in bold. This has a newly introduced subject, the causative agent *Yonghoka* 'Yongho', which is nominative. It also has a new predicate, glossed 'do'. (The gloss INDIC stands for INDICATIVE, a 'mood' of the verb which is used to refer to real rather than hypothetical events.)

- (37) a. ku say-ka cwuk-ess-ta (Korean)
 the bird-NOM die-PAST-INDIC
 'The bird died.'

- b. Yongho-ka [ku say-lul cwuk-key] **hay-ss-ta**
 Yongho-NOM the bird-ACC die-COMP do-PAST-INDIC
 ‘Yongho caused the bird to die.’

As in English, Korean causatives are complex sentences, containing two clauses. The embedded clause is in brackets, and contains a complementizer, *-key*, ‘(so) that’. Since Korean is head-final, the complementizer *-key* is final in the embedded clause, and the whole complement clause precedes the verb that selects it, *hayssta*. Literally, (37b) means ‘Yongho [that the bird died] caused’. The matrix clause is the ‘cause’ clause with the predicate *ha(y)* ‘do, make, cause’.

French causatives also use a ‘make’ or ‘do’ predicate of causation, the verb *faire*. In (38), (a) is again the basic sentence and (b) the causative, with the causative verb in bold:

- (38) a. Jean a lu ce livre. (French)
 Jean has.3SG read.PAST PARTICIPLE this book
 ‘Jean has read this book.’
- b. Nous avons **fait** lire ce livre à Jean
 we have.1PL make.PAST PARTICIPLE read.INFIN this book to Jean
 ‘We made Jean read this book.’

However, in French, unlike in Korean or English, the causative does not produce a biclausal construction. Although (38b) does contain two independent lexical verbs, the ‘make’ verb of causation and the ‘read’ verb, in fact the two verbs behave generally as a single verbal unit and not as predicates in separate clauses. For instance, unlike in English, the two verbs can’t be separated by the NP *Jean*, as (39) shows:

- (39) *Nous avons fait Jean lire ce livre.
 we have.1PL make.PAST PARTICIPLE Jean read.INFIN this book
 (≠ ‘We made Jean read this book.’)

So *Jean* doesn’t behave like the subject of an embedded clause. In the French, the two lexical verbs are actually both inside a single clause, and share a single set of arguments rather than each having its own arguments as they do in English or in Korean; this should remind you of the verb serialization which we discussed in Section 3.3.3.

One kind of typological variation in causatives, then, concerns whether or not the addition of a causative verb gives rise to an additional clause. However, not all causatives are formed by using an actual causative verb. In Korean, the most productive type of causative is that shown in (37b), but there is another type known as a MORPHOLOGICAL CAUSATIVE, illustrated in (40):

- (40) Yongho-ka ku say-lul cwuk-y-ess-ta (Korean)
 Yongho-NOM the bird-ACC die-CAUS-PAST-INDIC
 ‘Yongho killed the bird.’

The example in (40) contains only a single clause, and instead of a separate causative verb it has causative morphology: an affix *-y* (glossed as CAUS) on the 'die' verb. Note that *ku saylul* 'the bird' has been demoted to object in this clause: it has accusative case.

Many languages (though not English) also have a causative affix on the verb rather than using a separate causative verb. This situation parallels the one discussed in Section 7.1.1, where we saw that some languages have a special passive affix – see (6) through (8) for instance. Other examples of languages with a morphological causative are shown in (41) and (42): the basic sentence types are shown in each (a) example, the causatives in (b), and the causative affixes are in bold:

- (41) a. Mtsuko u-na-gw-a (Chichewa)
 waterpot SU-PAST-fall-ASPECT
 'The waterpot fell.'
- b. Mtsikana a-na-u-gw-**ets**-a mtsuko
 girl SU-PAST-OBJ-fall-CAUS-ASPECT waterpot
 'The girl made the waterpot fall.'
- (42) a. Müdür mektub-u imzala-dı (Turkish)
 director.NOM letter-ACC sign-PAST
 'The director signed the letter.'
- b. Dişçi mektub-u müdür-e imzala-**t**-tı
 dentist.NOM letter-ACC director-DATIVE sign-CAUS-PAST
 'The dentist made the director sign the letter.'

In the Chichewa examples, the causative (41b) differs from the basic sentence in various ways. Example (41a) is intransitive, while (41b) is transitive. The original subject, *mtsuko*, has been demoted to object in (41b): we can tell because there's an object agreement marker *u-* on the verb, agreeing with *mtsuko* 'waterpot' (in gender, though this isn't shown by the gloss). Also, the verb has a new subject agreement marker *a-* in (41b), and this agrees in gender with *mtsikana* 'girl' (rather than *mtsuko*). Finally, there's a CAUSATIVE suffix *-ets* on the verb in (41b).

In the Turkish examples, there's once again a new subject, *dişçi*, introduced into the causative construction in (42b). The former subject, *müdür*, 'director' is demoted to the position of indirect object in (42b), marked by the dative case; since there's already a direct object, *mektub* 'the letter', it can't take that position.

So far in this section, we have seen two types of causative: first, the 'cause'-verb plus 'effect'-verb type, and, second, the morphological causative, as in (41b) and (42b). Although English has no morphological causative (just as it has no morphological passive), it does illustrate a third type of causative construction, the lexical causative. For instance, some verbs can be used either intransitively, so that no causation is expressed, or transitively, so that they include a causer as their subject: *The bottle broke/I broke the bottle* (also *melt*, *sink*, *smash* and many other verbs). A few intransitive verbs have a closely related causative transitive verb, such as *sit/seat* and *rise/raise*, as in *The wreck rose to the surface/We raised the wreck to the surface*. Another example of a lexical causative is shown from Greek in (43):

- (43) a. pijéno (Greek)
 go.1SG
 'I go.'
- b. pijéno to peđí s to sxoló
 go.1SG the child.ACC to the school.ACC
 'I take the child to school.'

Example (43b) is causative, but there's no marker of this at all – the same verb meaning 'go' is used in both (43a) and (43b). Note that the English translation here also uses a lexical causative, but of a different kind, since *go* is replaced in English with a causative verb *take* (= 'cause to go').

As the examples in this section illustrate, causatives can generally be derived from either a basic intransitive verb or a basic transitive verb. The cross-linguistic properties of the construction are as follows:

The causative construction

- \emptyset > subject (i.e. a new subject is introduced).
- Former subject demoted > object; or to an oblique argument; or is deleted.
- Verb adding causation is introduced ('make', 'have', etc.), or else the main verb has causative morphology.

An example illustrating the deletion of the original subject in a causative construction is given in (44). Songhai (or Sonrai) is a Nilo-Saharan language of Mali, Burkino Faso and Niger: the basic sentence is in (44a), the causative in (44b), and the causative affix is in bold:

- (44) a. Garba nga tasu di. (Songhai)
 Garba eat rice the
 'Garba ate the rice.'
- b. Ali nga-**ndi** tasu di
 Ali eat-CAUS rice the
 'Ali got someone to eat the rice.'/'Ali caused the rice to be eaten.'

The original subject of the basic clause, *Garba*, is simply deleted in (44b), while a new subject of the causative verb is added, *Ali*.

Finally, recall from Section 7.3 that the applicative construction can feed into the passive construction by creating new object NPs, and these new objects can then be further promoted to subject. Similarly, the causative construction can create new objects by demoting the former subject, and these new objects are then available to be passivized. So the causative often feeds into the passive construction as well. Example (45) illustrates from Chichewa; in (45a) we have the causative construction, and in (45b) the passive formed from it:

- (45) a. Buluzi a-na-wa-sek-ets-a ana. (Chichewa)
 lizard SU-PAST-OBJ-laugh-CAUS-ASPECT children
 ‘The lizard made the children laugh.’
- b. Ana a-na-sek-ets-edw-a ndi buluzi.
 children SU-PAST-laugh-CAUS-PASSIVE-ASPECT by lizard
 ‘The children were made to laugh by the lizard.’

The NP *ana* ‘children’ in (45a) is a direct object: it triggers object agreement on the verb, so the object marker *wa-* agrees with the *ana* NP (in gender, though again not directly shown by the gloss). In the passive, (45b), this former object *ana* has undergone promotion to the subject position of the whole verbal complex: as in English, subjects are initial in the clause. And the object marker, *wa-*, has now disappeared from the verb, since passivized verbs are, of course, intransitive and hence have no object to agree with.

We can conclude, then, that it is quite general for processes that change the grammatical relations of noun phrases to interact with one another, creating further promotions and demotions.

7.5 SUMMARY

‘Valency’ refers to the number of core arguments that a verb has. We have seen in this chapter that languages typically have one or more valency-changing operation. These may increase the number of core arguments, as, for instance, with the applicative and causative constructions in Sections 7.3 and 7.4. Or alternatively, valency-changing may involve a decrease in the number of core arguments, for instance in the passive and antipassive (Sections 7.1 and 7.2). The processes we’ve seen also involve promotion and demotion of core arguments, foregrounding some NPs and backgrounding others – removing them from the ‘core’ – for various pragmatic purposes. We have also seen that these processes interact with one another, for instance by producing a new core argument that can be further promoted.

FURTHER READING

Palmer (1994) will be very useful for many of the issues covered in this chapter, especially passives and antipassives, syntactic pivots, causatives and applicatives. See also Keenan (1985a), Foley and Van Valin (1985) on the passive, and Comrie (1989: Chapter 8; 1985b) and Song (1996) on the causative. Some of the data on processes that change grammatical relations come from Baker (1988), a very advanced work which you should probably only tackle (as opposed to browsing for interesting data) after a course in theoretical syntax. Dixon and Aikhenvald (2000) is an edited collection of papers which all focus on valency-changing processes, and from which I’ve taken some of the data in this chapter.

EXERCISES

1. In Section 7.3, we considered the type of applicative construction known in English as DATIVE MOVEMENT, an alternation which gives rise to pairs such as *Kim gave the book to Lee*/*Kim gave Lee the book*. As noted earlier, not all verbs which take an NP and a *to*-PP complement can undergo the alternation. Your task is to work out what factors condition the application of dative movement. I have given a few examples, but you will need to find others, to get a fuller picture. I have also suggested grammaticality judgements which accord with my own intuitions, but you should feel free to disagree with them, and to find or make up other examples to support your case. Given that judgements may vary, the ‘correct answer’ here is a rather fluid concept!
- (1) a. Lee donated the prize money to her favourite charity.
b. *Lee donated her favourite charity the prize money.
 - (2) a. The shopkeeper refunded the money to me.
b. The shopkeeper refunded me the money.
 - (3) a. Kim passed the ball to Lee.
b. Kim passed Lee the ball.
 - (4) a. I transferred the money to Lee.
b. *I transferred Lee the money.
 - (5) a. We showed/sent/forwarded/texted that message to all our friends.
b. We showed/sent/?texted/?forwarded all our friends that message.
 - (6) a. Kim dispatched that letter to his lawyer.
b. *Kim dispatched his lawyer that letter.
 - (7) a. I faxed my answer to him straight away.
b. I faxed him my answer straight away.
 - (8) a. I handed/delivered the parcel to the publishers.
b. I handed/*delivered the publishers the parcel.
 - (9) a. I awarded/presented fantastic prizes to the best students.
b. I awarded/*presented the best students fantastic prizes.
 - (10) a. I recommended/introduced *Knowledge of Language* to the students.
b. *I recommended/introduced the students *Knowledge of Language*.
2. Study the data in (1) through (10) (from Klammer 1994) and then say exactly how the causative construction is formed in the Malayo-Polynesian language Kambera. Note that one crucial affix in the Kambera is left unidentified and un glossed. What is it?
- (1) Na pakanabu-ta weling la ài.
he fall-1PL.OBJ move from tree
‘He made us fall from the tree.’

- (2) Da rara hàmu da pàu.
they be.red be.good the.PL mango
'The mangoes are nice and ripe.'
- (3) Na lui du ...
it melt EMPHASIS
'It should dissolve ...'
- (4) Na palui-ya na liling.
he melt-3SG.OBJ the.SG candle
'He melts the candle.'
- (5) Da pakatuda-ya na anakeda.
they sleep-3SG.OBJ the.SG child
'They put the child to sleep.'
- (6) Napa jàka u kabeli ...
later if you return
'Later, if you (sg.) return ...'
- (7) Parara-ya na pàu.
be.red-3SG.OBJ the.SG mango
'Let the mango ripen.'
- (8) Da kawàra katuda.
they both sleep
'They both sleep.'
- (9) Ta pakabeli-ha da tentara.
we return-3PL.OBJ the.PL soldier
'We get the soldiers to return.'
- (10) Ambu ta kanabu.
NEG we fall
'Let's not fall.'

3. The examples in this exercise are from an Australian language, Kalkatungu, and are taken from Blake (2001b).

Tasks: (i) Example (1) shows a basic clause. Work out what construction is illustrated by the data in (2) and (3). (ii) A crucial grammatical morpheme, *ntjama*, is left unglossed. What is its function? (iii) What other changes are seen in (2) and (3) as compared with (1)? Make sure you use the correct grammatical terminology in describing them, as far as possible.

- (1) Kalpin-tu intji-mi nga-tji utjan
man-ERG chop-FUT me-DAT firewood
'The man will chop my firewood.'/'The man will chop the firewood for me.'
- (2) Kalpin-tu intji-**ntjama**-mi ngayi utjan
man-ERG chop-???-FUT me.OBJ firewood
'The man will chop me firewood.'

- (3) Kalpin-tu intji-**ntjama**-mi-ngi utjan
 man-ERG chop-???-FUT-me firewood
 ‘The man will chop me firewood.’

4. Southern Tiwa, a native American language from the Tanoan family of New Mexico, has a passive construction illustrated in examples (1) through (10). However, in Southern Tiwa this construction has an important restriction which doesn’t occur in English or the other languages seen so far.

Task:

- i. What is the syntactic restriction on the passive in Southern Tiwa?
- ii. Why are the examples in (4), (6) and (10) ungrammatical?
- iii. Finally, do you have any ideas about why a language might have such a restriction on the passive? Think again about PERSON and about what effect the passive has on a subject: compare (3) with (4) and (5) with (6).

Hints

- Note that in examples like (3), (5), (7) and (8) there are no independent pronouns in the Southern Tiwa sources. Instead, the verb *mu* meaning ‘see’ has bound pronominal prefixes showing the PERSON and NUMBER of the subject and the object (though not all verbs have an object). These prefixes occur in (3) through (9), and specify all the information that in the English translations is realized by independent pronouns (such as *You saw me*).
- When the verb in Southern Tiwa has both a subject and an object, these markers are fused together to form a single prefix: see (3) and (5), where the gloss indicates these fused forms with /. In (3), the prefix *bey-* means 2SG(SU) and 1SG(OBJ), i.e. it shows simultaneously that the subject is second person singular and the object is first person singular. In (5), the prefix *i-* means that the subject is first person singular and the object is second person singular. Obviously, the fused forms only occur if the verb has both a subject *and* an object. *The answer has nothing whatever to do with the fusion of subject and object markers, or with the appearance or non-appearance of independent pronouns.*
- Read through all the data first. Then go through it step by step, and formulate a hypothesis at each stage about the restriction on the passive. Amend your hypothesis to account for new data as necessary. Compare (4) with (7) and (6) with (8); then compare (6) with (9).
- I’ve used the notation \neq in the English translations to indicate what the ungrammatical forms in Southern Tiwa would mean if they were grammatical.

- (1) seuanide liora-mu-ban
 man lady-see-PAST
 ‘The man saw the lady.’
- (2) liora mu-che-ban seuanide-ba
 lady see-PASSIVE-PAST man-by
 ‘The lady was seen by the man.’

- (3) bey-mu-ban
2SG(SU)/1SG(OBJ)-see-PAST
'You saw me.'
- (4) *te-mu-che-ban ̄i-ba
1SG(SU)-see-PASSIVE-PAST you-by
(≠ 'I was seen by you.')
- (5) i-mu-ban
1SG(SU)/2SG(OBJ)-see-PAST
'I saw you.'
- (6) *a-mu-che-ban na-ba
2SG(SU)-see-PASSIVE-PAST me-by
(≠ 'You were seen by me.')
- (7) te-mu-che-ban seuanide-ba
1SG(SU)-see-PASSIVE-PAST man-by
'I was seen by the man.'
- (8) a-mu-che-ban seuanide-ba
2SG(SU)-see-PASSIVE-PAST man-by
'You were seen by the man.'
- (9) a-mu-che-ban awa-ba
2SG(SU)-see-PASSIVE-PAST him-by
'You were seen by him.'
- (10) *seuanide mu-che-ban na-ba
man see-PASSIVE-PAST me-by
(≠ 'The man was seen by me.')

The data in this exercise are mostly from Allen and Frantz (1983) – modified slightly – with additional data courtesy of Don Frantz.

5. The data in (1) through (3) in this exercise (taken from Nedjalkov 1997) are from the Tungusic language Evenki, spoken in eastern Siberia.

Task: (i) Examine each pair, and figure out what is the function of the verbal suffix marked in bold in each (b) sentence – I have glossed it simply as SUFFIX, rather than showing its meaning. (ii) Identify exactly what kind of construction arises in the (b) sentences. (iii) What other grammatical changes occur in the (b) sentences? Why do they occur?

Hints

- Different verbs take different forms of the suffix in question, but the function of the suffix is the same in each instance.
- It will help to consider what arguments the verbs have in each pair of examples.
- You will need to concentrate especially on the glosses in each example, rather than on the English translations.

- (1) a. Asatkan suru-re-n.
 girl go.away-PAST-3SG
 ‘The girl went away.’
- b. Atyrkan asatkan-me suru-pken-e-n.
 old.woman girl-ACC go.away-SUFFIX-PAST-3SG
 ‘The old woman made the girl go away.’
- (2) a. Beje eme-re-n.
 man come-PAST-3SG
 ‘The man came.’
- b. Beje moo-l-va eme-v-re-n.
 man tree-PL-ACC come-SUFFIX-PAST-3SG
 ‘The man brought firewood.’
- (3) a. Tyge d’alup-ta-n.
 cup become.full-PAST-3SG
 ‘The cup became full.’/‘The cup filled.’
- b. Asatkan tyge-ve d’alup-ki-ra-n.
 girl cup-ACC become.full-SUFFIX-PAST-3SG
 ‘The girl filled the cup.’

6. In Section 7.2, we introduced the idea that syntactically ergative languages can have a pivot which operates in terms of the *absolute* NPs, whilst syntactically accusative languages can have a pivot which operates in terms of *subject* NPs. (You might like to revise Section 7.2 before tackling this exercise.) The data sets that follow are from two unrelated languages: A. is from Bare, an extinct language of the North Arawak family, from Brazil and Venezuela (the data are from Aikhenvald 1995) and B. is from Guugu Yimidhirr, a native language of Australia (the data are taken from Haviland 1979). Both data sets illustrate co-ordinate clauses with ellipsis of one grammatical relation in the second clause. Each clause is bracketed, and neither language uses actual conjunctions such as ‘and’. You will need to look at the index on each NP in order to see which NP in the first clause the omitted NP refers back to.

Task: Examine each data set, and figure out whether each language is syntactically ERGATIVE or syntactically ACCUSATIVE. State your evidence clearly, using the correct grammatical terminology.

Hints

- I haven’t labelled the NPs with A, S and O so you will need to work out for yourself which NP is the A, the S and the O in these examples.
- There is no actual case-marking on the NPs in Bare, so you won’t be able to tell from the form of the noun phrases whether or not Bare is morphologically ERGATIVE.
- A language that is morphologically ergative may or may not also be syntactically ergative.

A. Bare

- (1) a. [kwati_i i-karuka tšinu_j] [Ø_j i-baraka]
 jaguar 3F.SG-bite dog 3F.SG-run
 'A jaguar_i bit the dog_j and [it]_j ran.'
- b. [da kwati_i i-d'áwika] [mawaya_j a-kharuka Ø_i]
 the jaguar 3F.SG-die snake INDEF-bite
 'The jaguar_i died (because) a snake_j; bit [it]_i.'

B. Guugu Yimidhirr

- (2) a. [Nyulu yarrga_i gada-y] [Ø_i mayi_j buda-y].
 3SG boy.ABS come-PAST food.ABS eat-PAST
 'The boy_i came and [he]_i ate the food.'
- b. [Nyulu yarrga-a_i mayi_j buda-y] [Ø_i gada-y]
 3SG boy-ERG food.ABS eat-PAST come-PAST
 'The boy_i ate the food_j and then [he]_i came.'

7. The data in this exercise (slightly adapted from Chung 1976) are from Indonesian, a syntactically accusative language. The usual constituent order is seen in (1). You have five tasks to complete. (i) Examine these data in (1) first and state what is the unmarked order of the verb, subject, object and indirect object or oblique NP.

- (1) a. Monjet men-gigit saja.
 monkey TRANS-bite I
 'A monkey bit me.'
- b. Saja mem-bawa surat itu kepada Ali.
 I TRANS-bring letter the to Ali
 'I brought the letter to Ali.'
- c. Mereka ber-lajar ke Amerika.
 they INTRANS-sail to America
 'They sailed to America.'

The next set of data illustrate a fronting process in Indonesian. (ii) Examine the sentences in (2) and (3) and figure out what GRAMMATICAL RELATION the fronted constituent must bear. Your answer should account both for the grammatical data in (2) and the ungrammatical examples in (3). (The English translations are deliberately neutral here, so you will need to study the original Indonesian carefully.) Then (iii) say what other grammatical changes occur when the constituent is fronted.

- (2) a. Ikan merah itu dia sudah tangkap.
 fish red the he PERF catch
 'He already caught the red fish.'
- b. Itu dapat kita lihat pada mata-nja.
 that can we see in eye-its
 'We can see that in its eyes.'

- (3) a. *Polisi itu saja serahkan sendjata saja kepada.
 police the I surrender weapon I to
 ('I surrendered my gun to the police.')
- b. *Danau itu sedang mereka be-renang di.
 lake the PROG they INTRANS-SWIM in
 ('They were swimming in the lake.')

The next data set illustrates a construction in Indonesian which alters grammatical relations, changing a basic sentence such as (4a) into (4b). (iv) What syntactic processes does this involve? Discuss them in terms of PROMOTION and/or DEMOTION and state the effects of the construction on the grammatical relations.

- (4) a. Saya meng-kirim surat itu kepada wanita itu.
 I TRANS-send letter the to woman the
 'I sent the letter to the woman.'
- b. Saya meng-kirim-i wanita itu surat itu.
 I TRANS-send-APPLIC woman the letter the
 'I sent the woman the letter.'

If the fronting construction you identified in connection with (2) and (3) applies to the examples in (4), the results are as follows: (5a) is ungrammatical but (5b) is grammatical. (v) In light of your answers concerning (4), account for this difference in grammaticality. You will need to say why the constituent can be fronted in (5b) but not in (5a).

- (5) a. *Wanita itu saja kirim surat itu (kepada)
 woman the I send letter the to
 (≠ 'I sent the woman the letter.')
- b. Wanita itu saja kirim-i surat itu.
 woman the I send-APPLIC letter the
 'I sent the woman the letter.'

8. In each of the following three data sets, A. to C., the (b)/(c) sentences show a CAUSATIVE construction derived from the corresponding (a) sentences.

Task: State how the causative is formed in each of the three languages illustrated. Your answer should include:

- (i) an explicit statement of how the causative is expressed in each of the languages
- (ii) an indication of and explanation for any additional grammatical changes in each example, especially in the verbal morphology, and in the position and morphology of any NP arguments of the verb
- (iii) an attempt to explain the reason for the ungrammaticality in examples (8c) and (9c) in the Japanese data set.

Hints

- Don't worry unduly about the actual form of the verbal morphology in these examples. In some cases, there are alternations or irregularities in the morphology, but these need not concern us here.
- You will find it helpful to consider at the start whether the language in each data set is nominative/accusative or ergative/absolutive in its morphology.

A. K'iche' (data from Campbell 2000)

- (1) a. š-e:-kam-ik
ASP-3PL.ABS-die-INTRANS
'They died.'
- b. š-e:-qa-kam-isa:-x
ASP-3PL.ABS-1PL.ERG-die-CAUS-TRANS
'We killed them.'
- (2) a. š-Ø-atin-ik
ASP-3SG.ABS-bathe-INTRANS
'He bathed.'
- b. š-Ø-r-atin-isa:-x
ASP-3SG.ABS-3SG.ERG-bathe-CAUS-TRANS
'She washed him.'

B. Amharic (data from Amberber 2000)

- (3) a. k'ibe-w k'allət'ə
butter-DEF melt.PERF.3M.SU
'The butter melted.'
- b. aster k'ibe-w-in a-k'allət'ə-čč
Aster(female name) butter-DEF-ACC CAUS-melt.PERF-3F.SU
'Aster melted the butter.'
- (4) a. liǰ-u dabbo bəlla
child-DEF bread eat.PERF.3M.SU
'The child ate some bread.'
- b. aster liǰ-u-n dabbo a-bəlla-čč-iw
Aster(female name) child-DEF-ACC bread CAUS-eat.PERF-3F.SU-3M.OBJ
'Aster fed the child some bread.'
- (5) a. aster č'əffərə-čč
Aster(female name) dance.PERF-3F.SU
'Aster danced.'
- b. ləmma aster-in as-č'əffər-at
Lemma(male name) Aster-ACC CAUS-dance.PERF.3M.SU-3F.OBJ
'Lemma made Aster dance.'

C. Japanese (data from Dixon 2000 and Tsujimura 1996)

- (6) a. Taroo-ga konsaato-e it-ta
 Taro-NOM concert-to go-PAST
 'Taro went to a concert.'
- b. Ryooshin-ga Taroo-o konsaato-e ik-ase-ta
 parents-NOM Taro-ACC concert-to go-CAUS-PAST
 'His parents made Taro go to a concert.'
- c. Ryooshin-ga Taroo-ni konsaato-e ik-ase-ta
 parents-NOM Taro-DATIVE concert-to go-CAUS-PAST
 'His parents let Taro go to a concert.'
- (7) a. Hanako-ga aruita
 Hanako-NOM walk.PAST
 'Hanako walked.'
- b. Taroo-ga Hanako-o aruk-ase-ta
 Taro-NOM Hanako-ACC walk-CAUS-PAST
 'Taro made Hanako walk.'
- c. Taroo-ga Hanako-ni aruk-ase-ta
 Taro-NOM Hanako-DATIVE walk-CAUS-PAST
 'Taro had/let Hanako walk.'
- (8) a. Hana-ga migotoni saita
 flower-NOM beautifully bloom.PAST
 'The flowers bloomed beautifully.'
- b. Taroo-ga hana-o migotoni sak-ase-ta
 Taro-NOM flower-ACC beautifully bloom-CAUS-PAST
 'Taro made the flowers bloom beautifully.'
- c. *Taroo-ga hana-ni migotoni sak-ase-ta
 Taro-NOM flower-DATIVE beautifully bloom-CAUS-PAST
 (≠ 'Taro had the flowers bloom beautifully.')
- (9) a. Hanako-ga kizetu-sita
 Hanako-NOM faint.PAST
 'Hanako fainted.'
- b. Taroo-ga Hanako-o kizetu-sase-ta
 Taro-NOM Hanako-ACC faint-CAUS-PAST
 'Taro made Hanako faint.'
- c. *Taroo-ga Hanako-ni kizetu-sase-ta
 Taro-NOM Hanako-DATIVE faint-CAUS-PAST
 (≠ 'Taro had Hanako faint.')

Wh-constructions: Questions and relative clauses

Chapter 7 introduced processes of promotion and demotion: we looked at ways in which languages change the argument structure of verbs by changing their grammatical relations. As we saw, this led to changes in the core arguments of verbs – for instance, objects may be promoted to become subjects, and subjects may be demoted to an oblique phrase, or even deleted. In this chapter, we will see that languages also have ways of moving phrases around within the clause without changing their grammatical relations. I concentrate particularly on two types of construction: *wh*-questions (Section 8.1) and relative clauses (8.2). We also look at focus and other movement constructions (Section 8.3).

8.1 WH-QUESTIONS

8.1.1 Languages with *wh*-movement

Wh-questions are so called because, in English, they begin with a *wh*-word such as *what*, *who*, *where*, *which*, *when*, *why* and also *how*. (1) and (2) illustrate:

- (1) a. Lee saw [that girl with the long scarf] at the bus-stop yesterday.
 b. [Who] did Lee see __ at the bus-stop yesterday?
- (2) a. Lee saw that girl with the long scarf [at the bus-stop] yesterday.
 b. [Where] did Lee see that girl with the long scarf __ yesterday?

Note that the sequence of words which is being questioned must be a constituent – in fact, this was one test for constituent structure in Chapter 5. I have indicated in square brackets the constituent being questioned in the (a) sentences, and as well as the *wh*-word which replaces it in the (b) sentences, since this is also a constituent. The gap shows the position that the questioned phrase formerly occupied. In English, and in many other languages, a *wh*-phrase is ‘fronted’: it occurs to the left of the clause: this is known as *wh*-movement.

Wh-questions are constructed as follows. The phrase that we’re asking a question about is first replaced by a suitable *wh*-word or *wh*-phrase, such as *which girl*. What constitutes a suitable *wh*-word depends on the category and properties of our original phrase. An NP such as *that girl with the long scarf* is replaced by *who*, or *which girl*; an NP headed by an inanimate noun, such as *that wonderful hand-built bike*, or a non-human noun, such as *that dreadful dog*, would be replaced by *what*, or *which X*.

Many other languages, even those entirely unrelated to English, also move *wh*-words and *wh*-phrases leftwards, to a similar initial position; this position is outside the main body of the clause, since it's not a position associated with any grammatical relation. In other words, it's not movement to a subject position, as we've seen, or indeed an object position or any other position occupied by the arguments of a verb. Some further examples of *wh*-phrases in the same position are shown in (6), from Koromfe, a Gur or Voltaic language of Burkina Faso:

- (6) a. **alama** pa vaga koŋ a mũĩ (Koromfe)
 who.PL give dog the ART rice
 'Who (pl.) gave the dog rice?'
 b. **sefu** də na a mane hẽŋ
 when he see ART money the
 'When did he find the money?'
 c. **ase** a kēō hoŋ pane a vaga koŋ
 what ART woman the give.PAST ART dog the
 'What did the woman give to the dog?'
 d. **nde** də na mə sundu koŋ
 where he see my horse the
 'Where did he see my horse?'

In languages with *wh*-movement to an initial position, the *wh*-expression precedes the material that normally occurs at the start of the clause. So for instance, in Welsh, the normal constituent order is VSO – that is, the finite verb or finite auxiliary is initial in the clause in a statement. But the *wh*-expression precedes the finite element in a *wh*-question. Examples (7) and (8) show some statements and the related *wh*-questions, with the gap corresponding to the original position of the moved expression underlined. The finite element is in *italics*, and the *wh*-phrase is in **bold type**:

- (7) a. *Enillodd* y myfyrwyr y wobr ddoe. (Welsh)
 win.PAST.3SG the students the prize yesterday
 'The students won the prize yesterday.'
 b. **Beth** *enillodd* y myfyrwyr _____ ddoe?
 what win.PAST.3SG the students yesterday
 'What did the students win yesterday?'
 (8) a. *Mae* wyau 'n dod o ieir.
 be.PRES.3SG eggs PROG come.INFIN from hens
 'Eggs come from hens.'
 b. **O ba** *greadur* *mae* wyau 'n dod _____?
 from what creature be.PRES.3SG eggs PROG come.INFIN
 'What creature do eggs come from?'

In fact, there is evidence from a variety of languages that the initial position to which the *wh*-phrase moves is actually the position immediately before the clause-

The gap in (10a) is the subject of the embedded clause – it's parallel to a sentence like *Mel left the cake out in the rain*. The gap in (10b) is the object of the embedded clause – compare *The students had read **all the books on the reading list** over the vacation*. And the gap in (10c) is a PP adjunct to the verb *stop*, as in *The bus will stop **at the market place***.

- (c) The major syntactic difference in English between embedded *wh*-questions and *wh*-questions in root clauses is that subject/auxiliary inversion generally only applies in root clauses, as we saw first in Chapter 3. So in (1), we get *Who **did** Lee see?*, but in an embedded clause we'd normally get *He asked [who Lee saw]*, rather than **He asked [who did Lee see]* (though some dialects find this grammatical). Also, as noted in Chapter 3, if the embedded clause is taken to be a quotation of direct speech, then inversion is typically acceptable.

8.1.2 Languages with *wh*-in-situ *wh*-questions

In Section 8.1.1, we saw that one common way of forming *wh*-questions cross-linguistically is to move a *wh*-expression to a special, pre-clause initial position: this is known as *wh*-fronting. However, not all languages form *wh*-questions by moving the *wh*-expression at all. Recall from Chapter 5 the ECHO QUESTION construction, which is illustrated again in (11):

- (11) a. Lee bought **how many copies** of that wonderful book?
 b. Kim took 300 pictures of **which mountain range** with her new camera?
 c. You've fallen in love with **who**?

The main characteristic of examples such as these is that the *wh*-phrase remains in the usual position occupied in the clause by the phrase that is being questioned. So for (11a), for instance, we find a related statement such as *Lee bought **four copies** of that wonderful book*. English generally has the option of asking a *wh*-question in this way; it typically conveys incredulity, or else is used when the addressee didn't hear a portion of the statement.

In some languages, however, the counterparts to (11) form the *only* way of asking *wh*-questions. In such languages, there is no *wh*-fronting, but instead the *wh*-word simply replaces a constituent in its normal position without moving, just as in echo questions in English. The technical term for this construction when the *wh*-phrase does not move is *WH-IN-SITU* – the Latin phrase means that the phrase stays in position.

Chinese and Japanese are both good examples of *wh*-in-situ languages. The first example is from Chinese, with the statement in (12a), and the question, showing *wh*-in-situ, in (12b):

- (12) a. Ni kanjian-le Zhangsan. (Chinese)
 you see-ASP Zhangsan
 'You saw Zhangsan.'
 b. Ni kanjian-le shei?
 you see-ASP who
 'Who did you see ___?'

The Chinese statement in (12a) has SVO order (as in English), so when the direct object is questioned, (12b), the interrogative (question) phrase remains immediately after the verb, in the normal position for the object.

In (13) and (14), we illustrate from Japanese: (13) is a statement, and (14) shows two different *wh*-questions formed from it. The *wh*-phrases are again shown in bold:

- (13) Hanako-ga kinoo [tomodati-to] [susi-o] tukurimasita. (Japanese)
 Hanako-NOM yesterday friend-with sushi-ACC made
 'Hanako made sushi with her friends yesterday.'
- (14) a. Hanako-ga kinoo [**dare-to**] [susi-o] tukurimasita ka?
 Hanako-NOM yesterday who-with sushi-ACC made QU
 'Who did Hanako make sushi with __ yesterday?'
- b. Hanako-ga kinoo [tomodati-to] [**nani-o**] tukurimasita ka?
 Hanako-NOM yesterday friend-with what-ACC made QU
 'What did Hanako make __ with her friends yesterday?'

In (14a), the position questioned is the object of the postposition *to* 'with' – note that the NP object *precedes* the P in Japanese, since this is a head-final language. In (14b), the position questioned is the object of the verb *tukurimasita* 'made', and the object again precedes the verb. Note that there is also a particle *ka* in (14), showing that these are questions; many languages have something similar.

You should now be able to see that just as in an echo question in English, the *wh*-phrase does not move in these Chinese and Japanese examples, but always remains in the normal position of the phrase being questioned.

In some languages, ordinary questions (rather than echo questions) can be formed either by *wh*-movement or by *wh*-in-situ: in other words, it appears that such languages employ both of the available strategies. In (15) and (16), I illustrate from French: the statement is in (15), and the two methods of forming a question (in informal French) are shown in (16):

- (15) Tu vois Pierre ce soir. (French)
 you see.PRES.2SG Pierre this evening
 'You're seeing Pierre tonight.'
- (16) a. **Qui** tu vois ___ ce soir ?
 who you see.PRES.2SG this evening
 'Who are you seeing __ tonight?'
- b. Tu vois **qui** ce soir?
 you see.PRES.2SG who this evening
 'Who are you seeing __ tonight?'

In (16a), we have *wh*-fronting, as in English, but in (16b), the *wh*-word *qui* 'who' is in exactly the same position – the object position – as the ordinary object NP, *Pierre*, in (15). It seems, then, that some languages are 'mixed' in terms of their methods for forming *wh*-questions.

8.1.3 Multiple *wh*-questions

Sections 8.1.1 and 8.1.2 discussed the two main alternatives available cross-linguistically for forming *wh*-questions, and also showed that some languages appear to employ both strategies. In this final section on questions, we illustrate the strategies that are employed when more than one constituent is questioned in a single clause.

English, of course, is a language with *wh*-fronting. However, if more than one constituent is questioned, then only one of the resulting *wh*-expressions can move leftwards to the initial position, and the remaining *wh*-phrase(s) must remain in-situ:

- (17) a. [Kim] saw [that stray dog] last night.
 b. [**Who**] saw [**what**] last night?
 c. *Who what saw last night?

In (17b), we see the only grammatical option for asking a MULTIPLE *WH*-QUESTION in English; (17c) shows that if we attempt to front all of the *wh*-phrases in such a question, the result is completely ungrammatical.

So what happens in other languages? In *wh*-in-situ languages, multiple *wh*-questions also occur, but since there is no *wh*-fronting, then all the questioned phrases must appear in-situ. Examples (18) and (19) illustrate from Japanese:

- (18) Taroo-ga [Yosiko-ni] [hon-o ni-satu] ageta. (Japanese)
 Taroo-NOM Yoshiko-DATIVE book-ACC two-CLASSIFIER gave
 'Taroo gave two books to Yoshiko.'
- (19) Taroo-ga [**dare-ni**] [**nani-o**] ageta no?
 Taroo-NOM who-DATIVE what-ACC gave QU
 'Who did Taroo give what?'

In (18), we see a statement, and, in (19), two of the constituents in that clause have been questioned: both the indirect object (the dative 'recipient' NP, *Yosiko-ni*) and the direct object (the accusative 'theme' NP, *hon-o ni-satu* 'two books'). The *wh*-phrases replacing these two constituents each remain in-situ, and, as (19) shows, each bear the usual case-marking appropriate for the grammatical relations which they hold in the clause.

So far, then, we have seen that multiple *wh*-questions may be formed as in English, by fronting one *wh*-phrase and leaving any others in-situ, or as in Japanese, by leaving all *wh*-phrases in-situ. There is, however, a third option, namely to front *all* the *wh*-phrases in a multiple *wh*-question. This strategy, known as MULTIPLE *WH*-FRONTING, occurs, for instance, in some of the Slavonic languages, such as Bulgarian and Serbo-Croatian. I illustrate first from Bulgarian: (20) through (23) show that all the *wh*-phrases are fronted in multiple *wh*-questions, even if this means fronting three *wh*-expressions:

- (20) **Kogo** vižda John? (Bulgarian)
 who sees John
 ‘Who does John see?’
- (21) **Koj kogo** vidjal?
 who whom saw
 ‘Who saw whom?’
- (22) **Kogo kakvo** e pital Ivan?
 whom what is asked Ivan
 ‘Who did Ivan ask what?’
- (23) **Koj kogo kakvo** e pital?
 who whom what is asked
 ‘Who asked whom what?’

In Bulgarian, the fronted phrases have to occur in a fixed order, as illustrated in these examples. In some languages with multiple *wh*-fronting, however, the *wh*-expressions can occur freely in any order. A closely related language, known by the cover term Serbo-Croatian (comprising Bosnian, Serbian and Croatian), allows *both* of the orders in (24), and both have the same meaning:

- (24) a. **Ko koga** voli? (Serbo-Croatian)
 who whom loves
 ‘Who loves whom?’
- b. **Koga ko** voli?
 whom who loves
 ‘Who loves whom?’

Since *wh*-expressions show case-marking just like ordinary NPs in this language, it is possible to tell which *wh*-phrase represents the subject and which represents the object – formal English also has a relic of a parallel case-marking, as indicated by the *who/whom* distinction.

Finally, just as we saw in Section 8.1.2 that some languages (such as French) may employ both the *wh*-movement and the *wh*-in-situ strategies for forming ordinary *wh*-questions, there are also languages that allow different options in multiple *wh*-questions. In Malagasy, which has the basic constituent order VOS, three constructions occur as alternatives, subject to some syntactic restrictions. The three possibilities are as follows: (a) like English, one *wh*-phrase fronts to the pre-clause initial position and remaining *wh*-phrases remain in-situ: this is shown in (25); or (b), like Japanese, all *wh*-phrases remain in-situ: this is shown in (26); or (c), like Bulgarian and Serbo-Croatian, all *wh*-phrases front to the initial position: (27) illustrates:

- (25) a. **Iza** no nividy **inona**? (Malagasy)
 who PRT bought what
 ‘Who bought what?’

- b. **Inona** no novidin' **iza**?
 what PRT bought who
 'Who bought what?'

(26) Anasan' **iza** **inona** ny savony?
 washes who what the soap
 'Who washes what with the soap?'

(27) a. **Aiza** **iza** no mividy ny vary?
 where who PRT buys the rice
 'Where does who buy the rice?'

- b. **Aiza** **inona** no vidinao?
 where what PRT buy.2
 'Where do you buy what?'

8.2 RELATIVE CLAUSES

8.2.1 Relative clauses in English

The next major *wh*-construction is the RELATIVE CLAUSE. This construction probably occurs in all languages in one form or another. Some typical examples from English are given in (28), where the relative clauses are in brackets:

- (28) a. She snarled at the **students** [who hadn't read the book].
 b. The **paper** [(which) we discuss next week] looks really interesting.
 c. I expect the **film** [(that) we're going to tonight] will be fantastic.
 d. They wrote a review of that **concert** [they heard in Newcastle].

First, note that we are dealing with COMPLEX SENTENCES here (see Section 3.2 for a reminder of these). We can tell that these examples are all complex sentences because they each contain more than one main verb: *snarled* and *read* in (28a), *discuss* and *looks* in (28b), and so on.

The relative clause itself is a type of subordinate clause which modifies (= says something about) a HEAD NOUN in the matrix clause: the head nouns are in bold type in (28). As you can see right away, these embedded clauses – *who hadn't read the book* and so on – couldn't be independent clauses of English, since they are all incomplete in some way, even if we take away the *who*, *which* and so on at the start of these clauses.

The function of the relative clause is to restrict the possible set of students, papers, films and concerts to just the subset that the speaker wants to talk about.¹ For example, in (28a), she didn't snarl at *all* the students, she snarled at a specific

¹ Some languages distinguish RESTRICTIVE relative clauses from NON-RESTRICTIVE ones, which don't serve to delimit a subset of items but are more like parenthetical comments. Examples of the latter from English are *Kim, who you met last night, is my sister's friend* or *Students – who never have any money – often take poorly paid work*. English non-restrictive relative clauses have a special intonation, as the commas or dashes indicate in the written form.

subset of students – only the ones who hadn't read the book. Relative clauses in other languages may look very different syntactically to the English examples in (28), but they all have in common this property of restricting the set of possible items that the head noun refers to. Cross-linguistically, relative clauses often have other typical features too, as we will see.

Looking specifically at English relative clauses, there are two properties which should help you with their identification. First, we see from (28) that the relative clause in English may just follow straight after the head noun, as in (28d), or else it may begin with a word like *who*, *which* or *that*, as in (28a), (28b) and (28c) respectively. Although these words may help you to detect relative clauses, each of them also has other roles in English, so you need to be careful in using them to identify relative clauses. For example, *that* is of course a complementizer, and so can also introduce an ordinary embedded clause selected by a verb, as in *Lee believed [that they'd be back soon]*. (We can tell that this is not a relative clause because it doesn't modify a head noun, and doesn't have the property – outlined above – of referring to a subset.) And the words *who* and *which* can occur in *wh*-questions, as we saw in Section 8.1.

The second property of relative clauses in English is that, like *wh*-questions, they contain a gap, and that is why the embedded clauses could not be stand-alone clauses. More precisely, each relative clause in (28) has a 'missing' noun phrase, indicated with a dash in (29):

- (29) a. ___ hadn't read the book
 b. we discuss ___ next week
 c. we're going to ___ tonight
 d. they heard ___ in Newcastle

We understand the gap to refer back to the head noun that's modified by the whole relative clause. The relativized position is said to be coreferential with the head noun. So in (29a), the gap is understood to refer to the students, and in (29b), the gap refers to the paper that'll be under discussion. The gap within the relative clause is known as the RELATIVIZED position, and in English, any position that could contain a noun phrase can be relativized. In (29a), the relativized position is the subject position of the relative clause; in (29b) and (29d), it's the direct object position; in (29c), the object of the preposition *to*.

It is also possible in English (though not common cross-linguistically) to have the relativized position as a POSSESSOR noun phrase: an example is shown in (30):

- (30) This is the student [whose name I always forget ___].

There is a gap in direct object position in (30): the verb *forget* is transitive. However, the relativized position itself is actually a possessor NP: the phrase in bold in *I always forget **that student's name***. In standard English, however, relative clauses can't simply leave a gap in the possessor position: this would give something like **This is the student (who) I always forget ___'s name*. Although I've marked this as ungrammatical,

8.2.2 Cross-linguistic variation in relative clauses

Although relative clause constructions in other languages will contain a head noun and a ‘restricting’ relative clause that modifies it, they don’t necessarily share any of the other syntactic properties of English relative clauses. For instance, although European languages often have a counterpart to the so-called relative pronouns *who* or *which* introducing the relative clause, this is much less common in other parts of the world. Here, we’ll look at some of the cross-linguistic variation.

8.2.2.1 Order of the relative clause and the head noun

One major typological distinction (= a distinction in type) is in the order of the relative clause and the head noun. In English, the relative clause follows the head noun. For example, in *the students [who hadn’t read the book]*, the relative clause [*who hadn’t read the book*] follows *students*. This order is also found in a great many other languages. In (33) and (34), you’ll find two examples from languages unrelated to English. The relative clauses are bracketed, and the head nouns are shown in bold (SM in (33) stands for ‘subject marker’):

- (33) wa **mwiè** rra [nrâ sùveharru nrâ toni] nrâ truu numea (Tinrin)
 the woman there 3SG like SM Tony 3SG stay Noumea
 ‘The woman that Tony likes lives in Noumea.’

This example is from a Melanesian language, Tinrin. Just as in the English, the relativized position is the direct object position within the relative clause – the object of the verb *sùveharru*, ‘like’. The relative clause is not introduced by any relative pronouns or other special markers.

In the Yimas language of Papua New Guinea, the verbal prefix *m-*, glossed as RM for relative marker, ‘functions much like the *wh*-word or *that* in English – to mark the whole relative clause as a definite referring expression’ (Foley 1991: 413). Note, however, that this is not a relative pronoun, and does not come at the start of the relative clause – it’s simply an affix on the verb:

- (34) **ŋaykum** [irut m-naampa-nt-um] (Yimas)
 women mat RM-weave-PRES-3PL
 ‘the women who are weaving the mats’

Since it is so familiar to readers of English, it may seem natural that relative clauses follow the head noun. In verb-final languages, though, the relative clause often *precedes* the head noun. Consider the Japanese examples in (35) and (36): the head nouns are shown in bold type, and the relativized position is the gap within the relative clause:

- (35) [kimura-san-ga ___ katte-iru] **inu** (Japanese)
 Kimura-Mr.-NOM keeps-NONPAST dog
 ‘the dog that Mr. Kimura keeps’

- (36) [kimura-san-ga ___ inu-o ageta] **kodomo**
 Kimura-Mr.-NOM dog-ACC give.PAST child
 ‘the child to whom Mr. Kimura gave a dog’

In (35), the relativized position is the direct object of the verb *katte*, ‘keep’: recall that Japanese is an SOV language, so the ‘missing’ object NP immediately precedes the verb in the bracketed relative clause. And, in (36), the relativized position is the indirect object of the verb *ageta* ‘gave’; the basic position for an indirect object in Japanese is before the direct object, hence the position of the gap shown here. Note that there is no equivalent to the English relative pronouns *who* or *which* in Japanese, nor any other word introducing the relative clause, and that the relative clause simply comes right before the head noun.

This constituent order ‘relative clause – head noun’ is common in other HEAD-FINAL languages. For instance, the relative clause construction which is native to Turkish (a language with SOV constituent order) is also head-final in this way (Kornfilt 1997), and the same applies to Korean. Hungarian, however, has both types of relative clause – the English pattern ‘head noun – relative clause’ as well as the head-final pattern.

8.2.2.2 Relative clauses are complex NPs

The examples in (35) and (36) are not full sentences, of course, but noun phrases, consisting of a head noun modified by the relative clause; the same applies to their English translations, and indeed to all head noun + relative clause constructions. These are rather special NPs, though: a noun with a clausal modifier of any kind is known as a COMPLEX NP, so ‘head noun plus relative clause’ is one type of complex NP. As with the term ‘complex sentence’, this technical term doesn’t mean ‘complicated’, but simply indicates a construction with an embedded clause. If we put complex NPs into a sentence, we can see that – just like any other noun phrases – they can generally slot into whatever position an NP can fill. For instance, both of these complex NPs can be subjects, as in (37). The whole complex NP – head noun and the relative clause that modifies it – is bracketed:

- (37) [The dog that Mr. Kimura keeps] has a bad cough.
 [The child to whom Mr. Kimura gave a dog] has a bad cough.

Or alternatively, both complex NPs can be direct objects:

- (38) I’ve never liked [the dog that Mr. Kimura keeps].
 I’ve never liked [the child to whom Mr. Kimura gave a dog].

In the Japanese example in (39), and in its English translation, we see the whole complex NP used as the subject of a clause: the head noun *hon*, ‘book’, is again in bold:

- (39) [Kinoo Ziroo-ga ___ yondeita **hon**]-ga nakunatta. (Japanese)
 yesterday Ziro-NOM (ACC) was.reading book-NOM missing
 '[The **book** that Ziro was reading ___ yesterday] is missing.'

In Japanese, the whole complex NP (bracketed) is marked as the subject of the clause by the fact that it bears nominative case, the case for subjects – the *-ga* marker at the end of the complex NP signals this. Note, though, that the gap *within* the relative clause itself is a direct object gap in (39). In both languages, the relativized position is the object of the 'read' verb. For that reason, I have marked the gap in the gloss as 'accusative', the case of direct objects in Japanese.

8.2.2.3 Relative clauses that are not embedded

In the examples of relativization seen so far, the relative clause is embedded *within* the main clause: the relative clause plus the head noun that it modifies form an NP that occupies a standard NP position, such as subject or object. We saw this in data such as (37) to (39), where the whole bracketed complex NP acted as subject or object of the main clause. However, in some languages, the relative clause is not embedded within the main clause, as we'll now see. Our examples are from Bambara, a member of the African language family Mande (Niger-Congo), and come from Creissels (2000). Consider first a simple sentence, where the constituent order is SOV (subject-object-verb):

- (40) wùlú yé démíséŋ 'kíŋ (Bambara)
 dog PERF child bite
 'The dog bit the child.'

Next are two different relative clauses formed from (40). The way that Bambara shows that these are relative clauses is by using a relative marker (glossed RM), *míŋ*, which signals the relativized position. In (41a), this marker immediately follows *wùlú* 'dog', and the relativized position is the subject, while in (41b), *míŋ* immediately follows *démíséŋ* 'child', and the relativized position is the object:

- (41) a. wùlú 'míŋ yé démíséŋ 'kíŋ
 dog RM PERF child bite
 'the dog that ___ bit the child'
- b. wùlú yé démíséŋ 'míŋ 'kíŋ
 dog PERF child RM bite
 'the child that the dog bit ___'

Note that all that is changed between (40) and the two examples in (41) is the presence of the relative marker in (41). Now if we want to use one of these relative clauses in a sentence, we see that it is not embedded *within* the main clause, but is more like an adjunct which is tacked onto a following independent clause. The literal translation provided here gives a flavour of this. Compare the actual English translation, where the whole complex NP (bracketed) is the object of *saw* in the main clause:

- (42) wùlú yé démíséŋ míŋ 'kíŋ, n y' ó bòlìt'ó 'yé
 dog PERF child RM bite I PERF this.one running see
 'I saw [the child that the dog bit] running away.'
 (Literally, 'The child that the dog bit, I saw this one running away.')

The main clause *n y' ó bòlìt'ó 'yé* 'I saw this one running away' can indeed be a full independent clause in Bambara: nothing is missing from it. This strategy is the only relativization strategy found in most of the northern Mande languages (Creissels 2000: 255).

8.2.2.4 Relativization strategies

As noted in Section 8.2.1, in English more or less any position in a clause that can contain an NP can be relativized, including the subject, direct object, and object of a preposition. We also saw that the relativized position in such examples contains a gap: here, each gap is marked and its position (grammatical function) within the clause is shown too:

- (43) a. the forms [that __ arrived yesterday] • subject
 b. the paper [(which) we discuss __ next week] • direct object
 c. the film [(that) we're going to __ tonight] • prepositional object

So these relative clauses use what is termed the 'gap strategy': the relativized position is simply empty.

In the case of a possessor NP, as we saw in Section 8.2.1, standard English has a special strategy. This involves using the form *whose* to form the relative clause, and moving the whole of the possessive noun phrase from its basic position to the special position before the left edge of the clause that is reserved for *wh*-phrases: *this is the student whose name I always forget*. As we saw earlier, this strategy also leaves a gap, here in the direct object position after *forget*. However, in informal English, we often use an alternative construction, shown in (44). This has no gap following *forget*, as you can verify for yourself, but instead uses a RESUMPTIVE PRONOUN in the relativized position (shown in bold). This is called the 'resumptive' strategy:

- (44) This is the student [who I always forget **her** name] • possessor

There is one more position which may be relativized in English: the object of comparative *than*:

- (45) This is the guy who my cat is smarter than __ / **him** • object of comparison

As you can see, English doesn't much like these relative clauses: they somehow often don't sound quite right, either with a gap or with the resumptive pronoun (*him*) in the relativized position.

There are, then, around five possible NP positions which can be relativized: subject, direct object, object of preposition/postposition, possessor NP and object of comparison. Cross-linguistically, these NP positions each take a place in what is known as the Accessibility Hierarchy, as shown in (46), where the subject is the highest position on the hierarchy and the object of comparison the lowest. The '>' means 'is more accessible than' – that is, more accessible to relativization. This accessibility manifests itself in various ways cross-linguistically, as we'll see.

(46) NP Accessibility Hierarchy for relative clause formation

Su > Direct obj > Object of adposition > Possessor > Object of comparison

In some languages, we find a rather more fine-grained set of possible NP positions. For instance, Welsh treats the objects of finite verbs differently from the objects of non-finite verbs. Some languages also have a separate indirect object position, but in many languages, as in English, indirect objects (*Kim gave the book to her friend*) are syntactically the same as ordinary prepositional objects. So (46) shows a basic set of NP positions that are available, which may differ a bit from language to language.

In what sense, though, is (46) a hierarchy? First, every position on the hierarchy is a cut-off point for relative clause formation in some language or languages. Subjects are most accessible to relativization, and, indeed, virtually all known languages can relativize subjects. But some languages don't allow relative clauses formed on any position lower down the hierarchy. Tagalog is an example of a language that only relativizes subjects. Other languages only relativize subjects and direct objects (e.g. Tongan); others only relativize the highest three positions, and so on. The prediction is that there are no languages that relativize, say, the object of a preposition, but which would disallow a relative clause on some higher position, such as subject.

The hierarchy is also manifested within the grammars of individual languages. Although English has quite a lot of latitude in relative clause formation, we saw in (45) that the lowest position, object of comparison, is a bit marginal. What we expect, then, is that relative clauses formed on lower positions of the hierarchy may not sound as natural in a language. How far down the hierarchy these dispreferred 'lower' positions start will vary from language to language.

The other way in which the hierarchy is manifested is in the different strategies used for relativization. Basically, the gap strategy is expected in the highest positions, especially for relativized subjects, and very often for direct objects. And the resumptive strategy is often used for relative clauses formed in the less accessible positions lower down the hierarchy: in other words, having a pronoun rather than a gap in the relativized position seems to make the lower positions more accessible. Once the resumptive strategy 'kicks in' at some point on the hierarchy, it's normally expected that this strategy will also be used for all lower positions that the language can relativize. So if a language starts using resumptive pronouns, say, when a prepositional object is relativized, we'd predict that it would use the resumptive strategy for any lower positions too.

English is actually unusual, cross-linguistically, in using the gap strategy when the

relativized position is the object of a preposition (*the film we're going to* ___). Compare the Hausa in (47) with the English translation (the Hausa word *da* at the start of the relative clause is a relative marker):

- (47) wuqad [da ya kashe ta da ita] (Hausa)
 knife RM he killed her with it
 'the knife that he killed her with ___'

The relativized position in the Hausa and in the English translation is the object of the preposition *da*, 'with'. English uses the gap strategy, and allows prepositions to be 'stranded' at the end of the clause; i.e. left with no prepositional object. But most languages avoid this one way or another, for instance, by using the resumptive strategy as the Hausa does here.

The Accessibility Hierarchy is broadly supported by investigations of relative clause formation across many languages. If you know a language other than English, perhaps you can now test for yourself how relative clauses are formed (if at all) on each position on the hierarchy.

8.3 FOCUS MOVEMENTS AND SCRAMBLING

So far we have seen two kinds of *wh*-construction, questions and relative clauses, both of which often involve movement. Many languages use displacement of constituents in order to focus on a particular phrase, perhaps in order to emphasize it, or else to contrast it with other parts of the clause. Cross-linguistically, FOCUS constructions frequently move a particular constituent to a special position. These constructions typically have much in common with *wh*-fronting constructions. For instance, they often move a focalized constituent to a special position before the left edge of the clause, and also, this movement leaves a gap in the clause that corresponds to the moved XP (that is, a 'something' phrase). This occurs in English, as we first saw in Chapter 1, as in *Beans I like* __, *but spinach I can't stand* __. The gap shows the position of the focalized constituent: here, it's the direct object of the verb in both these co-ordinated clauses.

We have already seen a number of examples of this kind of fronting from other languages. In the discussion of Mam in Section 7.2, we saw that although the basic constituent order is verb-initial, an absolutive NP can be focalized through fronting. And in exercise 4 in Chapter 5, we saw that Welsh (also verb-initial in basic constituent order) uses fronting for the same purpose. Some similar Welsh examples are given here: (48) shows the normal constituent order, (49a) has a PP fronted for focus, and (49b), a fronted VP. The basic position of these fronted phrases is shown with a gap:

- (48) Mae Caryl yn palu yn yr ardd heddiw. (Welsh)
 be.PRES.3SG Caryl PROG dig.INFIN in the garden today
 'Caryl is digging in the garden today.'

- (49) a. [_{pp} Yn yr ardd] mae Caryl yn palu ___ heddiw.
 in the garden be.PRES.3SG Caryl PROG dig.INFIN today
 ‘It’s *in the garden* that Caryl is digging today.’
- b. [_{vp} Palu yn yr ardd] mae Caryl ___ heddiw.
 dig.INFIN in the garden be.PRES.3SG Caryl today
 ‘??It’s *digging in the garden* that Caryl is today.’

The English translation of (49b) sounds very odd (hence prefaced with two question marks) because in English a VP constituent can’t be focused in this way – it can’t simply be fronted, nor can it occur in the cleft construction. As I noted in Chapter 5, this doesn’t mean that there isn’t a VP constituent in English, just that not all syntactic processes necessarily apply to all constituents in a language.

In (50), (51) and (52), we see object-fronting for focus in three Oceanic languages which are normally subject-initial, i.e. SVO or SOV (data from Lynch 1998):

- (50) [La paia taume], eau kama hilo-a. (Nakanai)
 the dog your I not see-it
 ‘As for your dog, I haven’t seen it.’
- (51) [Boroma] Morea ese e-ala-ia. (Motu)
 pig Morea ERG he-kill-it
 ‘The pig, Morea killed it.’
- (52) [Nimwa aan n̄mataag-asuul] r-ɬm-atakɬn. (Lenakel)
 house that wind-big it-PAST-destroy
 ‘That house was destroyed by the cyclone.’

Although the pre-clause position is frequently used, cross-linguistically, for focusing a constituent, this is not the only option. For instance, in Hungarian, the position immediately preceding the verb is the position used for contrastive focus. The following illustrate: (53a) and (54a) both show a neutral sentence, i.e. one with no particular focus on any constituent, and the two (b) examples show a constituent moved to the pre-verbal focus position (shown in bold):

- (53) a. Péter olvasta a könyvet. (Hungarian)
 Peter read.DEF the book.ACC
 ‘Peter read/was reading the book.’
- b. Péter **a könyvet** olvasta.
 Peter the book.ACC read.DEF
 ‘It’s *the book* that Peter read.’
- (54) a. Tegnap vendégek érkeztek a szállodá-ba.
 yesterday guests arrived the hotel-in
 ‘Guests arrived at the hotel yesterday.’
- b. A vendégek **tegnap** érkeztek a szállodá-ba.
 the guests yesterday arrived the hotel-in
 ‘It’s *yesterday* that the guests arrived at the hotel.’

In Japanese and Korean, a leftward movement construction related to focus movement is known as SCRAMBLING. This construction results in a very free ordering of constituents, as we can see in (55), from Japanese. All of these sentences are grammatical, and the only restriction on order is that the verb must be in final position. The basic (neutral) constituent order in Japanese is SOV, as illustrated in (55a):

- (55) a. Kinoo Taroo-ga Ginza-de susi-o tabeta. (Japanese)
 yesterday Taro-NOM Ginza-in sushi-ACC eat.PAST
 ‘Taro ate sushi in Ginza yesterday.’
 b. Taroo-ga Ginza-de kinoo susi-o tabeta.
 c. Kinoo susi-o Taroo-ga Ginza-de tabeta.
 d. Susi-o kinoo Taroo-ga Ginza-de tabeta.
 e. Ginza-de Taroo-ga kinoo susi-o tabeta.
 f. Kinoo Ginza-de susi-o Taroo-ga tabeta.

In languages which have extensive case-marking, variations in phrase order resulting from scrambling are unlikely to cause any ambiguity, because each of the nominal constituents has a case-marker showing its grammatical relation (subject, object and so on). Japanese has nominative/accusative case-marking, and a fixed order is not required in order to show who is doing what. The variations are not glossed in (55), as the constituents are identical to those in (55a), but before finishing this chapter, please ensure that you can see what each phrase means.

8.4 SOME CONCLUSIONS

In this chapter, we have seen a variety of what are known as *wh*-constructions. Although these do not all contain an actual *wh*-word or phrase – or its equivalent in other languages – there are various properties which are common to these constructions, and this leads linguists generally to regard them as a related family of constructions. In English, two reliable signs of a *wh*-construction are the potential presence of a *wh*-expression (as in *The animals (which) I was filming yesterday*), plus the existence of a gap within the clause from which some phrase has moved. These same indications of *wh*-movement also occur in the constructions seen in (56) and (57):

- (56) What a strong swimmer Kim is ____ !
 How tired I feel __ these days!
 (57) Kim is stronger than Lee is __.
 Wrens are smaller than robins are __.

The examples in (56) are known as EXCLAMATIVES (something that you exclaim), and are reasonably transparently seen as *wh*-constructions with a fronted *wh*-phrase

and a gap. These are related to statements like *Kim is such a strong swimmer, or I feel so tired these days*.

The examples in (57), on the other hand, are less obviously *wh*-constructions, even though they do contain a gap. Note, however, that these COMPARATIVE constructions may contain an overt *wh*-word in non-standard English, as in *Kim is stronger than what Lee is*. Such evidence is regarded as a legitimate sign of a *wh*-construction.

We have seen that not all languages have what is known as *wh*-movement, either in interrogative clauses, or within a relative clause. However, despite the existence of superficial differences cross-linguistically, all these constructions are nonetheless regarded as closely related to the more familiar *wh*-constructions which do display movement, including the ones seen in this short section.

We have also looked briefly at focus constructions and scrambling, generally considered to be related to *wh*-constructions cross-linguistically. All of these movements differ from those discussed in Chapter 7 in that they specifically do *not* cause any change in the grammatical function of the moved phrase. Although most languages exploit the possibility of at least some movements of this type, there is a great deal of variation in terms of the freedom or the immobility of phrases.

FURTHER READING

On relative clauses, central readings are Keenan and Comrie (1977, 1979), Comrie and Keenan (1979), Comrie (1989: Chapter 7) and Keenan (1985b). On *wh*-questions and the idea that they leave a gap in the extraction site, see Radford (1988: Chapter 9). A seminal reading from the generative grammar tradition on *wh*-constructions and their general properties – though one which you will almost certainly find very challenging – is Chomsky (1977).

EXERCISES

1. Consider the Turkish *wh*-questions illustrated in (1) through (6), taken from Kornfilt (1997). What is the generalization concerning the position of the *wh*-phrase? (In other words, what single statement can you make about its position which will account for all the examples shown?) It may help you to recall that Turkish is an SOV language – its basic constituent order is subject-object-verb.

(1) bu kitab-1 kim oku-du?
 this book-ACC who read-PAST
 'Who read this book?'

(2) Hasan kitab-1 kim-e ver-di?
 Hasan book-ACC who-DATIVE give-PAST
 'To whom did Hasan give the book?'

- (3) Mehmet tarafından kim öl-dür-ül-dü?
 Mehmet by who die-CAUS-PASSIVE-PAST
 'Who was killed by Mehmet?'
- (4) Hasan ne-yi oku-du?
 Hasan what-ACC read-PAST
 'What did Hasan read?'
- (5) Hasan [sinema-ya kim git-ta] san-ıyor?
 Hasan cinema-DATIVE who go-PAST believe-PROG
 'Who does Hasan think went to the cinema?'
- (6) Hasan dün hangi kız-la dans-et-ti?
 Hasan yesterday which girl-with dance-do-PAST
 'Which girl did Hasan dance with yesterday?'

2. The data in this exercise are from Malayalam, a Dravidian language of India, and are taken from Asher and Kumari (1997). The examples show two kinds of data. There are seven examples that illustrate some basic, unmarked sentences, and the remaining eight are examples with various different constituents contrastively focused: the italics in the English translation enable you to work out which constituent in the Malayalam is being focalized. Your task is to indicate how this focus is achieved in Malayalam, and state which part of the clause is being focused in each example that has it.

Hints

- I have jumbled up the data illustrating neutral sentences and the sentences with focus, but you will probably find it helpful to sort the sentences out into an A set (neutral) and a B set (those with focus) before you start, and to group similar examples. You have enough data here to work out the essential facts concerning how focus is achieved in Malayalam.
- There are a few minor morphological (i.e. not syntactic) irregularities in the data; I have not ironed these out, but left them as examples of the natural, untidy nature of linguistic data. Comment on any that you find.

- (1) *naan* innale vannu
 I yesterday come.PAST
 'I came yesterday.'
- (2) *nii* pooyee tiiruu
 you go must
 'You really must go.'
- (3) *avan* atə ceytilla
 he it do.PAST.NEG
 'He didn't do it.'
- (4) *kutti* viittil illa
 child at.home NEG
 'The child is not at home.'

- (5) *naanee varaam*
I come.FUT
'I shall come.'
- (6) *avan varum*
he come.FUT
'He will come.'
- (7) *naan paṛayaan marannu*
I talk.INFIN forget.PAST
'I forgot to say.'
- (8) *nii pookanee paaṭilla*
you go.INFIN prohibition
'You should not go.'
- (9) *avan ezṭutanee paṛaṅṅuḷḷuu*
he write.INFIN tell.PAST
'He only told me to *write*.'
- (10) *avan pookaan paaṭilla*
he go.INFIN prohibition
'He mustn't go.'
- (11) *paṛayaanee paaṭilla*
talk.INFIN prohibition
'(You) should not *talk*.'
- (12) *poostṭ saadhaarṅa raṅṭə maṅikkə varunnu*
post usually two hour.DATIVE come.PRES
'The post usually comes at two o'clock.'
- (13) *avaneē varum*
he come.FUT
'*He alone* will come.'
- (14) *naaḷe pattə maṅikkeē varuu*
tomorrow ten hour.DATIVE come.IMPERATIVE
'Come at *ten o'clock* tomorrow.'
- (15) *avan atə ceyteeyilla*
he it do.PAST.NEG
'He *didn't* do it.'

3. In Kambera, certain verbs can take a nasal prefix (either *m-* or *ng-*), giving such pairs as these: *pata* 'to break something' and *mbata* 'to be broken'; *pana* 'to heat up something' and *mbana* 'to be warm/hot'; *kodang* 'to move something' and *nggodang* 'to be loose (e.g. a tooth)'.

Task: (i) Examine these pairs, along with the data in (1) and (2), which follow, (slightly adapted from Klamer 1994) and figure out what is the function of

the nasal prefix. (ii) In light of your answer, why is the relative clause in (3) ungrammatical?

Hints

- The example in (2a) is a relative clause with a gap in direct object position: see Section 8.2.
- It will help to consider what NP arguments the verbs have in (1) and (2). Recall that as a head-marking language, Kambera doesn't always have independent pronouns in argument positions – the pronominal markers on the verb can perform the same function. So in (1a), for instance, the subject – which I've translated as 'you' – is represented not by an independent pronoun in Kambera, but as a 2SG affix on the verb.

- (1) a. Ka u-**kunggul**-nya na ngohung!
so 2SG.SU-roll-3SG.OBJ the container
'You roll the container away!'
- b. **Nggungul**-nanya na ngohung
roll-3SG the container
'The container is rolling away.'
- (2) a. na kalembi na pa-**baha**-na ____
the clothes the REL-wash-3SG
'the clothes that she washed ____'
- b. Na-**mbaha** na kalembi-nggu nyungga
3SG.SU-be.wet the clothes-my I
'My clothes are wet.'
- (3) *na kalembi na pa-**mbaha**-na
the clothes the REL-be.wet-3SG
'(the clothes that she washed)'

4. The data in this exercise (taken from Hualde et al. 1994) are from the Lekeitio dialect of Basque. In each example, one constituent is focalized. The focalized constituents are indicated for you via the italics in the English translations; you will need to work out where they are in the Basque.

Task: How exactly is a constituent focalized in Basque? Give a generalization which covers all the data. (Allative in (5) is a case-marker, and gives the meaning expressed by the preposition *to* in English.)

- (1) lagunak txakurra ekarri-dau gaur goixian
friend.ERG dog.ABS bring-AUX today morning
'The friend brought *the dog* this morning.'
- (2) txakurra lagunak ekarri-dau gaur goixian
dog.ABS friend.ERG bring-AUX today morning
'*The friend* brought the dog this morning.'

- (3) txakurra gaur goixian ekarri-dau lagunak
dog.ABS today morning bring-AUX friend.ERG
'The friend brought the dog *this morning*.'
- (4) gaur goixian aitta etorri-da
today morning father come-AUX
'This morning, *father* arrived.'
- (5) estau nai Bilbora žun
NEG.AUX want Bilbao.ALLATIVE go
'S/he doesn't want to go *to Bilbao*.'
- (6) etxe barriža ikusi-dot, es subi barriža
house new.ABS see-AUX no bridge new.ABS
'I saw *the new house*, not the new bridge.'
- (7) Péruk esan dau bižar etorríko dala
Peru.ERG say AUX tomorrow come AUX
'Peru has said that he'll arrive *tomorrow*.'

5. This exercise is about RESPONSIVES in colloquial Welsh – answers to yes/no questions. These are questions which in many languages can simply be answered 'yes' or 'no'. As you will see, Welsh is more complex.

Task: Study the data that follow, and describe as accurately and concisely as possible the main principles that regulate the choice of the correct responsive. Your answers should take the form 'If the question ..., then the responsive ...'

Hints

- You will especially need to consider the (morpho)syntax of the finite verbs and auxiliaries in these data, which will mean examining the glosses very carefully. The finite element is clause-initial in Welsh. In some examples, you'll also need to consider the post-verbal syntax. Don't worry unduly about remaining syntactic features of the clause, which are generally not relevant to your answer.
- The form of *bod* 'be' in the third person singular present tense differs according to whether the subject is definite (*ydy*) or indefinite (*oes*).
- You don't need to attempt to account for the distinctions between the *affirmative* forms and the *negative* forms of the responsives in (8b), (9b) and (10b) (e.g. *gwnaf* vs. *Na wnaif*); this is not relevant to your answer. The distinction between the responsives themselves in (8), (9) and (10), however, *is* relevant.
- The example in (11) is distinct from all the data in (1) to (10). Your answer will need to reflect this.

- (1) a. Welaist ti 'r ffilm?
see.PAST.2SG you the film
'Did you see the film?'
- b. Do/Naddo
yes/no
'Yes/No.'

- (2) a. Ddaru Mair weld y ffilm?
 AUX.PAST Mair see.INFIN the film
 'Did Mair see the film?'
 b. Do/Naddo
 yes/no
 'Yes/No.'
- (3) a. Wnaeth Mair weld y ffilm?
 do.PAST.3SG Mair see.INFIN the film
 'Did Mair see the film?'
 b. Do/Naddo
 yes/no
 'Yes/No.'
- (4) a. Ydyn nhw 'n mynd?
 be.PRES.3PL they PROG go.INFIN
 'Are they going?'
 b. Ydyn / Nac ydyn
 be.PRES.3PL / NEG be.PRES.3PL
 'Yes/No.'
- (5) a. Ydyn nhw wedi mynd?
 be.PRES.3PL they PERF go.INFIN
 'Have they gone?' (*An appropriate answer is either (b) or (c).*)
 b. Ydyn / Nac ydyn
 be.PRES.3PL / NEG be.PRES.3PL
 'Yes/No.'
 c. Do/Naddo
 yes/no
 'Yes/No.'
- (6) a. Ydy 'r dŵr yn berwi?
 be.PRES.3SG the water PROG boil.INFIN
 'Is the water boiling?'
 b. Ydy / Nac ydy
 be.PRES.3SG / NEG be.PRES.3SG
 'Yes/No.'
- (7) a. Oes 'na goffi yn y gegin?
 be.PRES.3SG there coffee in the kitchen
 'Is there coffee in the kitchen?'
 b. Oes / Nac oes
 be.PRES.3SG / NEG be.PRES.3SG
 'Yes/No.'
- (8) a. Helpith y ferch heno?
 help.FUT.3SG the girl tonight
 'Will the girl help tonight?'

- b. Gwneith / Na wneith
do.FUT.3SG / NEG do.FUT.3SG
'Yes/No.'
- (9) a. Gwnei di agor y ffenest?
do.FUT.2SG you open.INFIN the window
'Will you open the window?'
- b. Gwnaf / Na wna
do.FUT.1SG / NEG do.FUT.1SG
'Yes/No.'
- (10) a. Fyddi di 'n dod i 'r ffilm heno?
be.FUT.2SG you PROG come.INFIN to the film tonight
'Will you be coming to the film tonight?'
- b. Byddaf / Na fyddaf
be.FUT.1SG / NEG be.FUT.1SG
'Yes/No.'
- (11) a. Cyngerdd welaist ti?
concert see.PAST.2SG you
'Was it a concert that you saw?'
- b. Ie / Nage
yes/no
'Yes/No.'

6. The data in this exercise are from Kurdish, specifically the northern variety known as Kurmanji (or Kurmanci), and are taken from Creissels (2008a, b), citing data from Blau and Barak (1999). You should re-read Section 6.3 before starting the exercise.

Hints and tasks

- Two distinct case-markings occur on NPs in these data, which are termed 'direct' case and 'oblique' case here (it's not unusual for language-specific terms to be used in the specialist literature on a language). In (1) and (2), for instance, the subject is in the 'direct' case. This case is formally and functionally unmarked; it has no inflection, and is the 'bare' citation form used for nouns. The nouns *mirov* in (7) and *Sînem* in (4) and (8) are also in the direct case. In fact, all NPs and pronouns that are not specifically marked in the gloss as oblique should be regarded as having direct case. All the oblique NPs and pronouns are specifically marked as such, as the glosses show (OBL): for NPs this involves a suffix *-ê* (feminine singular) or *-î* (masculine singular), and for pronouns, a distinct oblique form is used (compare English *she* vs. *her* etc.). Your task below, however, is to work out the alignment patterns in these data, using the standard notation set out in Chapter 6.
- In the data in (3) through (13), two different verb tenses occur. You will find it helpful to group examples together according to tense.

- First, consider the verb agreement in intransitive clauses, (1) and (2). (i) Which argument does the verb agree with, A, S or O?

(1) Ez dikev-im
1SG fall.PRES-1SG
'I'm falling.'

(2) Mirov dikev-e
man fall.PRES-3SG
'The man is falling.'

- Next, consider transitive clauses, here shown in two distinct tenses, present and perfective. (ii) Which argument(s) does the verb agree with in (3) through (13)? Answer in terms of the A, S and O relations. (iii) In what ways do the data in (3) through (13) pattern with the intransitive clauses in (1) and (2)? In what ways do they differ? (iv) Indicate the alignment pattern or patterns that you find, using the standard notation, and indicate why each pattern occurs. In other words, what conditioning factors do you find for each pattern? (v) What is the technical term for this pattern of data?
- If you would find it easier to answer these questions with a continuous narrative, rather than point by point, you may do so, but your answer must be clear and concise, and must cover all the issues raised.

(3) Ez Sînem-ê dibîn-im
1SG Sînem-OBL.F.SG see.PRES-1SG
'I see Sinem'

(4) Mîn Sînem dît-Ø
1SG.OBL Sînem see.PERFCTV-3SG
'I saw Sinem.'

(5) Tu Sînem-ê dibîn-î
2SG Sînem-OBL.F.SG see.PRES-2SG
'You see Sinem.'

(6) Ez mirov-î dibîn-im
1SG man-OBL.M.SG see.PRES-1SG
'I see the man.'

(7) Mîn mirov dît-Ø
1SG.OBL man see.PERFCTV-3SG
'I saw the man.'

(8) Te Sînem dît-Ø
2SG.OBL Sînem see.PERFCTV-3SG
'You saw Sinem.'

(9) Sînem min dibîn-e
Sînem 1SG.OBL see.PRES-3SG
'Sinem sees me.'

- (10) Sinem-ê ez dît-im
Sinem-OBL.F.SG 1SG see.PERFCTV-1SG
'Sinem saw me.'
- (11) Sinem te dibîn-e
Sinem 2SG.OBL see.PRES-3SG
'Sinem sees you.'
- (12) Mirov-î ez dît-im
man-OBL.M.SG 1SG see.PERFCTV-1SG
'The man saw me.'
- (13) Sinem-ê tu dît-î
Sinem-OBL.F.SG 2SG see.PERFCTV-2SG
'Sinem saw you.'

Asking questions about syntax

The title of this chapter is deliberately rather ambiguous. I am hopeful that by this point, you will be able to construct a basic syntactic description of a language, either a language that you speak well yourself, or one for which you can find a native speaker consultant. Section 9.1 outlines the kind of questions that you will need to investigate. Section 9.2 provides a short case study of Welsh, illustrating how these questions could be answered. These sections, then, refer to asking questions about syntax in the most literal way. However, I also hope that the discussion in the previous chapters has ignited some curiosity about the human language faculty more generally. In Section 9.3, I briefly outline some issues and questions surrounding our syntactic abilities that are currently widely debated within linguistics. Section 9.4 looks at possible further directions to pursue in your study of syntax.

9.1 SYNTACTIC DESCRIPTION: WHAT QUESTIONS TO INVESTIGATE

This section aims to give you a framework with which to write a basic syntactic description of a language that you know well, or for which you can access data readily. If appropriate, you can ask one or more native speakers to act as language consultant(s). Make sure you give the source(s) of your data, including attributions to the literature (i.e. cite your sources). Acknowledge any help given by language consultants.

- Give the name by which the language is known to its native speakers, plus its English name, if any. State its language family and the principal locations in which it is spoken.
- Your description should include *some or all* of the questions outlined in (1) to (13), depending on what features of the language you consider to be most interesting from a syntactic, morphosyntactic and typological point of view. Give enough information on (and illustration of) any feature to make it comprehensible to someone who has no prior knowledge of the language.
- *All* parts of the discussion must be illustrated with appropriate and sufficient data, glossed and translated. Number each example, following the conventions used in this textbook. If your language uses a writing system other than the Roman alphabet, cite data using whatever standard system of transliteration is used for this language.

- Give a list of abbreviations used in the gloss where these differ from those found in this textbook.
- You won't need to discuss syntactic properties that are not manifested in your language. For instance (question (8) below), not all languages mark morphologically the relationship between a head and its dependents; see Section 4.3.7. If you were discussing Chinese, question (8) wouldn't be relevant. In such cases, you can simply state that your language does not, for instance, display head- or dependent-marking. Similarly, you don't need to mention the antipassive construction unless your language has an ergative alignment (Chapter 7).
- You can collapse questions together where this makes sense for your language. For instance, questions (8) and (10) touch on the same kinds of data, and for some languages it would be appropriate to discuss them together.
- Make sure, when answering each question, that you provide adequate explanations: do not leave the reader to work out for themselves what your data show.

Some basic questions to consider:

- (1) What is the neutral, or unmarked, constituent order (sometimes termed 'word order') in the clause, if there is one? (Chapter 1, Chapter 6). If there is no neutral constituent order, describe the main principles of linearization. You should at least illustrate a transitive clause with two full NP arguments, and an intransitive clause. Are the orders the same in both these clause types?
- (2) What alternative neutral constituent orders are possible, if any? How marked are these?
- (3) What are the main word classes (or syntactic categories) in your language? Discuss any that have especially interesting properties. Focus on the LEXICAL classes N, V and A. You can expect any language to have a distinct class of nouns and verbs. Most languages will also have a distinct class of adjectives. Most will also have at least one or two (and maybe dozens) of adpositions. Justify all word classes that you posit: in other words, give evidence from its morphosyntactic properties and syntactic distribution to demonstrate that each proposed class should be regarded as distinct (Chapter 2, Chapter 4). Include some of the main FUNCTIONAL categories that your language distinguishes.
- (4) Is your language predominantly head-initial or predominantly head-final? Illustrate with data from more than one word class of heads. Are there any difficulties in establishing a predominant linearization? (Chapter 4). Remember that you are looking at the ordering of heads and their complements here, rather than the position of adjuncts with relation to heads.
- (5) How does your language express clausal negation? (Chapter 3).
- (6) Describe the main strategies for joining clauses together that are found in your language. What kinds of complementation occur? For instance, does the language have both finite and non-finite complement clauses? Does it have clausal

subjects? If so, can they be both finite and non-finite? Does your language rely largely on subordination, as is the case for typical European languages? Does it have nominalized embeddings? Or does it, for instance, use co-ordination or verb serialization? (Chapter 3)

- (7) Are the constituent orders occurring in subordinate clauses the same as those in root clauses, or different? If different, describe the differences carefully (Chapter 3). Are there any (other) noteworthy differences between root and embedded clauses?
- (8) How, *if at all*, does your language mark morphosyntactically the relationship between heads and dependents? (Chapter 4). In other words, is your language largely head-marking or largely dependent-marking? Does it display a mix of both strategies? Illustrate at least with reference to the verb and its arguments.
- (9) Does your language readily identify distinct constituents? (Chapter 5) If so, give at least two tests for constituency, illustrating with contrasting grammatical and ungrammatical data. Are there distinctive DISPLACEMENT processes for constituents in your language? Perhaps alternatively your language has free word order of the type found in Warlpiri (Chapter 6); if so, illustrate.
- (10) Describe the way(s) in which the grammatical functions A, O and S are identified in your language (Chapter 6). Does this rely predominantly on constituent order, on agreement or cross-referencing, or on case-marking? Does your language exemplify an accusative or an ergative alignment? Make sure you give enough data to illustrate this. If your language has ergative alignment, is this purely morphological, or is it also (a much rarer possibility) syntactic? In other words, is there a clear SUBJECT relation in your language? Is it possible to identify a syntactic pivot?
- (11) Your language almost certainly has some readily identifiable ways to change the grammatical functions or relations, either increasing or decreasing the valency of a verb (Chapter 7). Does it have a passive? If ergative, an antipassive? An impersonal construction? An applicative? A causative?
- (12) Describe how *wh*- (i.e. constituent) questions are formulated (Chapter 8). Does the language have *wh*-fronting or *wh*-in-situ, or perhaps both? Discuss the main strategies for forming relative clauses in your language (if appropriate, you could alternatively discuss these under question (6), as part of subordination). Are there other *wh*-constructions in your language; for instance, is focus movement found, and if so, is it similar to *wh*-question formation?
- (13) Are there any other interesting syntactic constructions that are not covered by these questions? If so, explain and illustrate them.

You may be wondering why it's worthwhile to investigate the grammars of languages. I hope that the preceding chapters have answered this question, but, in case not, you should consider the fact that every week, languages are becoming extinct. Today there are perhaps 6000 or so languages in the world; we don't know the

exact number, and to some extent the answer depends on what counts as a distinct 'language' rather than a 'dialect'. Most of these languages, perhaps as many as 90 per cent of the world's total, are endangered. A language that has only a handful of speakers, even a few hundred or a few thousand speakers, is unlikely to survive to the end of this century. Many languages will become extinct by the end of this decade (see Crystal 2000; Dixon 1997).

When a language dies out because its speakers have chosen to speak (or been browbeaten into speaking) one of the large 'global' languages, much of the culture of that society is likely to die out too (see Nettle and Romaine 2000). Just as biological diversity is endangered by the relentless march of westernized societies, so linguistic diversity is threatened by domination from the world's major languages, including English. Every time another language becomes extinct, we lose the opportunity to discover something more about the human language faculty; every language investigated to date has fascinating constructions and patterns that we may never know about unless linguists (including native speaker linguists) uncover them. If you decide that you want to undertake linguistic fieldwork, you will need professional training in all its aspects, which includes handling often complex socio-political situations.

(For more information about how to describe the syntax and morphosyntax of a language, see T. Payne (1997, 2006).)

9.2 A CASE STUDY: GRAMMATICAL SKETCH OF COLLOQUIAL WELSH

This section provides a necessarily brief grammatical sketch of Colloquial Welsh, illustrating the kinds of answer that could be given to questions in the previous section. The term 'Colloquial Welsh' is used by linguists to indicate, broadly speaking, the modern spoken language.

Colloquial Welsh is spoken in many, though not all, parts of Wales, where it has around half a million native speakers. Welsh speakers are also scattered throughout Britain, and there is a Welsh-speaking community in Argentina. The language is known as *Cymraeg* to its native speakers. Welsh is a member of the Celtic language family, a branch of Indo-European, and is thus related ultimately to English.

Let's now turn to the syntactic properties. In the examples that follow, I have deliberately left in place the 'messy' morphological details that characterize natural languages. See if you can spot some of these. If you'd like further information, please contact me.

In neutral constituent order, a finite element (either a main verb or an auxiliary) is in clause-initial position. The subject immediately follows. With an inflected main verb, this gives VS(O) order, as in (14) and (15); no other neutral word orders occur. Many VSO languages allow an alternative SVO order, but Welsh does not, as (16) shows:

- (14) Gwerthodd Elin y delyn.
 sell.PAST.3SG Elin the harp
 'Elin sold the harp.'

(15) Diflannodd y delyn.
 disappear.PAST.3SG the harp
 'The harp disappeared.'

(16) *Elin gwerthodd y delyn.
 Elin sell.PAST.3SG the harp
 ('Elin sold the harp.')

A fixed constituent order identifies the grammatical functions, A, S and O. Welsh has accusative alignment, and a clear subject relation. S and A are identical; both immediately follow the finite verb or auxiliary, and, in a VSO clause, O immediately follows S. This is seen in (14). Both the S and A relations trigger subject agreement on a finite verb, under restricted conditions, as is illustrated below.

A finite auxiliary (in bold) occurs in clause-initial position in both (17) and (18); again, the subject immediately follows, and there is also a non-finite lexical verb lower down in the clause:

(17) **Gwnaeth** Elin werthu 'r delyn.
 do.PAST.3SG Elin sell.INFIN the harp
 'Elin sold the harp.'

(18) **Mae** Elin wedi / yn gwerthu 'r delyn.
 be.PRES.3SG Elin PERF / PROG sell.INFIN the harp
 'Elin has sold/is selling the harp.'

The main difference between these two clauses, apart from the different auxiliaries, is that (18) is an overtly aspectual clause. Aspectual particles, including *wedi* (PERFECT) and *yn* (PROGRESSIVE) co-occur with the auxiliary *bod*, 'be', as shown in (18). *Bod* is the only aspectual auxiliary; there is no 'have' auxiliary in Welsh. (The auxiliary *bod* 'be' has suppletive (= unpredictable and irregular) morphology, as you can see from (18).)

Welsh is a strongly head-initial language. A verb precedes its direct object, as in *gwerthu'r delyn*, 'sell the harp', just seen in (18). All adpositions are prepositions, as in *wrth y drws* 'at the door'. Nouns precede their possessors, as in *ci Elin* (dog Elin), 'Elin's dog'. Across all categories, in fact, heads precede complements.

Negation is a particularly complex area of Welsh syntax. Here, I illustrate clausal negation in finite clauses:

(19) Ddiflannodd y delyn **ddim**.
 disappear.PAST.3SG the harp NEG
 'The harp didn't disappear.'

(20) Wnaeth Elin **ddim** gwerthu 'r delyn.
 do.PAST.3SG Elin NEG sell.INFIN the harp
 'Elin didn't sell the harp.'

(21) Dydy Elin **ddim** wedi / yn gwerthu 'r delyn.
 NEG.be.PRES.3SG Elin NEG PERF / PROG sell.INFIN the harp
 'Elin hasn't sold/isn't selling the harp.'

- (22) Werthodd Elin **mo** 'r delyn / *ddim y delyn.
 sell.PAST.3SG Elin NEG the harp NEG the harp
 'Elin didn't sell the harp.'

As the first three examples show, clausal negation involves a negative adverb, *ddim*, which occurs in post-subject position. However, the adverb *ddim* can't appear before a direct object, as (22) shows; instead, a form *mo* is used, which literally means 'nothing of'. There are also changes to the finite verbs and auxiliaries in initial position in (19) through (22), as you'll see if you compare them with the affirmative clauses seen earlier; these changes occur because the clause is negative.

The lexical categories N, V and A are clearly distinct in Welsh, as they are in other European languages. The morphosyntactic categories that these word classes inflect for are very familiar from European languages, so I will not illustrate these specifically; examples occur in the data that follow. The inflectional categories include: for nouns, number and gender (masculine/feminine); for verbs, tense and subject agreement; and for adjectives, comparison. Welsh has a large class of prepositions, and these are more interesting, since most of them inflect to agree with their pronominal objects. Table 9.1 illustrates a characteristic paradigm. The bare citation form of the preposition is *wrth*.

Table 9.1

Inflectional paradigm for the Welsh preposition *wrth*, 'at'

	Singular	Plural
First person	<i>wrth-a i</i> at-1SG me 'at me'	<i>wrth-on ni</i> at-1PL us 'at us'
Second person	<i>wrth-at ti</i> at-2SG you 'at you'	<i>wrth-och chi</i> at-2PL you 'at you (PL)'
Third person	<i>wrth-o fo</i> / <i>wrth-i hi</i> at-3SG.M him / at-3SG.F her 'at him' / 'at her'	<i>wrth-yn nhw</i> at-3PL them 'at them'

As the occurrence of an inflectional paradigm for prepositions indicates, Welsh is head-marking rather than dependent-marking. Unlike in English, there is no case-marking whatever on either pronouns or nouns. Welsh has extensive agreement morphology. Heads agree with a following pronominal argument for person, number and, in the third person singular forms only, gender (excluding finite verbs, which display no gender agreement). In all instances, agreement crucially co-occurs *only* with a following pronominal argument, and *never* with a lexical noun phrase. This is illustrated in (23) through (25). Agreement occurs on six distinct categories of head, of which three are illustrated here: finite verbs agree with a pronominal subject: (23); non-finite verbs agree with their pronominal object: (24); prepositions also agree with their pronominal object: (25). In each case, the agreeing head and the following pronominal that it agrees with are underlined:

- (23) a. Cerddon nhw i 'r dre.
walk.PAST.3PL they to the town
'They walked to town.'
- b. Cerddodd / *Cerddon y genod i 'r dre.
walk.PAST.3SG / walk.PAST.3PL the girls to the town
'The girls walked to town.'
- (24) a. Gwnaeth Meic eu gweld nhw.
do.PAST.3SG Meic 3PL see.INFIN them
'Mike saw them.'
- b. Gwnaeth Meic (*eu) weld y genod.
do.PAST.3SG Meic 3PL see.INFIN the girls
'Meic saw the girls.'
- (25) a. arni hi b. arnyn nhw
on.3SG.F her on.3PL them
'on her' 'on them'
- c. ar yr eneth / *arni 'r eneth
on the girl / on 3SG.F the girl
'on the girl'
- d. ar y genod / *arnyn y genod
on the girls / on.3PL the girls
'on the girls'

Example (23a) shows a verb agreeing with a plural pronominal subject in person and number. However, when the verb has a lexical noun phrase subject, as in (23b), the agreeing form is ungrammatical. For finite verbs, there is no 'bare' citation form; instead, the third person singular is the default form, as (23b) shows.

For non-finite verbs, such as *gweld* in (24a), the agreement element is not a verbal inflection, but rather a preverbal marker (here, *eu*) which agrees with the pronominal object – here, in person and number. As (24b) shows, the agreement marker cannot occur with a lexical noun phrase object.

Examples (25a) and (25b) show an inflecting preposition, *ar* 'on', agreeing with a pronominal object. In (25c) and (25d), we see once again that when the preposition has a lexical noun phrase object, there is no agreement, and instead, the preposition occurs in its 'bare' citation form, *ar*.

Unlike more canonical head-marking languages, Welsh does not have true pronominal affixes; in other words, person and number cannot be reliably identified solely from the verbal or prepositional inflection. Table 9.2 illustrates this with the past tense paradigm for a regular verb, *gweld*, 'see' (giving 'I saw', 'you (sg.) saw' etc.). Note that there are only three distinct forms of the verb, *gwelis*, *gwelodd* and *gwelso*: the pronouns that follow are therefore essential to identify the participant. Colloquial Welsh is, then, not what is termed a 'pro-drop' language: the subject pronouns cannot generally be omitted.

Table 9.2

Inflectional paradigm for the past tense of the Welsh verb *gweld*, 'see'

	Singular		Plural	
First person	<i>gwelis</i> see.PAST.SG	<i>i</i> I	<i>gwelso</i> see.PAST.PL	<i>ni</i> we
Second person	<i>gwelis</i> see.PAST.SG	<i>ti</i> you	<i>gwelso</i> see.PAST.PL	<i>chi</i> you(pl)
Third person	<i>gwelodd</i> see.PAST.3SG	<i>o / hi</i> he/she	<i>gwelso</i> see.PAST.PL	<i>nhw</i> them

There are two remaining inflectional paradigms for lexical verbs in Colloquial Welsh, namely the future tense and the conditional. Both of these have parallel properties to the past tense in terms of verbal agreement.

As is typical for a European language, Welsh makes extensive use of subordination. Constituent order is the same both in finite root and embedded clauses: (26) illustrates a finite embedded VSO clause, bracketed:

- (26) Dywedodd Aled [darllenith Elin y papur].
 say.PAST.3SG Aled read.FUT.3SG Elin the paper
 'Aled said that Elin will read the paper.'

As well as finite complement clauses, as in (26), Welsh has infinitival complement clauses. The syntax of the latter is actually rather complex, since some of these are interpreted as finite, others as non-finite. Examples (27) and (28) illustrate these two types: both are introduced by a small functional element, *i*, which I've glossed as 'to' since it looks identical to the preposition *i* 'to'. The complement clauses in these examples are superficially identical, but have very different meanings and properties:

- (27) Dywedodd Aled [i Elin ddarllen y papur].
 say.PAST.3SG Aled to Elin read.INFIN the paper
 'Aled said that Elin had read the paper.'
- (28) Disgwylodd Aled [i Elin ddarllen y papur].
 expect.PAST.3SG Aled to Elin read.INFIN the paper
 'Aled expected Elin to read the paper.'
 (i.e. *not* 'Aled expected that Elin had read the paper.')

The difference between the two clause types stems from the kind of verb that occurs in the matrix clause. When the 'upstairs' predicate is a verb like *dweud*, 'say', as in (27), or *meddwl*, 'think', the infinitival clause is interpreted as finite. In fact, there is good evidence that these clauses really are finite, including the fact that they have the same interpretation as ordinary tensed clauses. Conversely, when the 'upstairs' predicate is a verb like *disgwyl* 'expect' or *dymuno*, 'wish/want', the infinitival clause is not interpreted as finite; instead, very like its English translation, it tends to refer to

future events that have not yet happened. Syntactically, the embedded clause in (28) does not behave like a finite clause, either. For instance, it is negated with a distinct negator that occurs in non-finite clauses, here shown in bold:

- (29) Disgwyliodd Aled [i Elin **beidio** â darllen y papur].
 expect.PAST.3SG Aled to Elin NEG with read.INFIN the paper
 ‘Aled expected Elin not to read the paper.’

Clausal subjects are another form of subordination in Welsh, but these can only be non-finite, as in the bracketed clause in (30), and not finite, as in (31):

- (30) Mae [mynd i'r cyfandir] yn gyffrous.
 be.PRES.3SG go.INFIN to the continent PRED exciting
 ‘Going to the continent is exciting.’
- (31) *Mae [bydd Aled yn mynd] yn gyffrous.
 be.PRES.3SG be.FUT.3SG Aled PROG go.INFIN PRED exciting
 (‘That Aled will be going is exciting.’)

Various valency-changing operations occur in Welsh. The main valency-reducing process is the passive, illustrated in (33), which is formed from the active construction in (32):

- (32) Mae'r plismon wedi dal y lladron.
 be.PRES.3SG the policeman PERF catch.INFIN the thieves
 ‘The policeman has caught the thieves.’
- (33) Mae'r lladron wedi cael eu dal (gan y plismon).
 be.PRES.3SG the thieves PERF get.INFIN 3PL catch.INFIN by the policeman
 ‘The thieves have been caught (by the policeman).’

As in passives generally, the subject of the active sentence in (32) – *y plismon*, ‘the policeman’ – is demoted or deleted in the passive; in (33), there is an optional *gan* ‘by’ phrase, containing the agent. And the direct object of the active construction (*y lladron* ‘the thieves’) is promoted to subject position in the passive. These changes can be seen from the constituent order: as always, the subject immediately follows the finite verbal element in Welsh. Like many other languages, Welsh has an auxiliary-plus-main-verb passive construction: the added auxiliary is *cael* ‘get’ in (33). The lexical verb, *dal* ‘catch’, also has an agreement marker in the passive, which agrees with the promoted subject in (33): here, it is *eu*, third person plural, agreeing with the plural subject (*y lladron* ‘the thieves’). Literally, the passive reads ‘The thieves have got their catching by the policeman’. There is no morphological passive in Welsh. (Note also that in keeping with the usual restrictions on agreement in Welsh, the finite auxiliary does not agree with a lexical subject NP in (33), and is therefore singular rather than plural.)

Welsh also has a causative construction, which uses a causative verb such as

There are many more fascinating features of Welsh syntax (see Borsley *et al* 2007); the case study provides a brief sketch and illustrates the major typological properties of the language.

9.3 SOME QUESTIONS CONCERNING SYNTAX

In this section, I briefly address some questions and controversies that have been widely debated in recent linguistic research. Some central readings are provided to whet your appetite.

- Why is it so hard for adults to learn a new language?

Many of us have *attempted to learn* at least one language, in other words by making a conscious effort. And we mostly find it very difficult, even if we're immersed in the new culture. As adults, we can only envy the effortless, naturalistic way in which children learn the ambient language(s) of their culture without any instruction. Moreover, across the world, it's totally normal for children to learn – natively – more than one language; it's English-speaking cultures that are abnormal in being so overwhelmingly monolingual. So children have a head-start in language learning, and adult language learners are always at a grave disadvantage compared to children. As we approach puberty, our language-learning ability declines; beyond puberty, it largely atrophies. After that we may, with persistence, become fluent speakers of a new language, but we won't become native speakers – and we won't have the same intuitions about grammaticality as someone who learnt the language as a child. Language, in common with many other acquired skills, such as musical ability, has what is called a critical period (or a sensitive period) for learning. If learning takes place beyond that period, it is no longer effortless, and acquisition will probably be less than native-like.

In biological terms, none of this is terribly surprising. Many other animals exhibit sensitive periods for various systems, including motor systems (involving movement), sensory systems, and behaviour. An example often cited is birdsong. Some (though not all) species of songbirds have to learn their songs by hearing an adult model, inevitably inviting comparisons with language. If they don't have an appropriate adult model (for instance, if they are reared without a singing adult male bird), their song fails to develop properly. Some aspects of song are therefore genetically determined, but input from the environment is crucial. What is important in both birdsong and language, then (and indeed, in many other biological systems) is the interaction between genes and environment. Humans have a language faculty which is genetically specified. This does not mean that there is *in any sense* a single 'language gene', or, most likely, even a dedicated group of language genes. It means that in normal situations, we all acquire at least one language as children: that is hardwired in our species. The interaction of many genes is almost certainly involved. And input from the environment is needed before the child's brain can get to work building a language. No other

species has a language faculty, and no other species can acquire a human language, even under intensive instruction. But for the genetic predisposition to learn language to be triggered, cultural input is required. Language learning in children relies on normal human interaction, including exposure to language data: this is the environment.

How much of our linguistic ability is pre-specified by our genes, and how much is down to the influence of our environment, is currently a hot topic of debate within linguistics. If all languages have some feature in common, is it inbuilt, part of the genetic recipe for being human? Or can it be attributed to the fact that we are all exposed to human cultures, which present children with certain uniform experiences of the world? Frankly, we don't really know.

When adults do try to learn another language, we typically find it easier to learn languages that are closely related to our own, or that are similar typologically – in other words, languages which have similar characteristics, such as sharing the same word order. As in other spheres of life, the familiar is easier to grasp than the radically new. But this brings us to the next question.

- Are all languages equally complex? And are some languages harder to learn than others?

Impressionistically, few people (including professional linguists) are in any doubt that some languages are harder for each of us to learn as adults. But can languages be intrinsically hard, or intrinsically easy? It used to be generally considered that all languages were, essentially, equally complex, and that complexity in one area of the grammar would be balanced out by simplicity in another area. Recently, that view – which was more ideological than evidence-based – has been challenged from many quarters, and it now seems indefensible. The collection of papers in Sampson *et al.* (2009) provides much interesting discussion. One of the authors (Guy Deutscher) calls the claim that 'all languages are equally complex' nothing more than an urban legend!

It also now seems that the demographic properties of a language – including the number of speakers it has and the extent of its spread around the world – directly correlate with the linguistic complexity of the language (Lupyan and Dale 2010). At least in terms of their morphosyntax, large global languages which have many millions of speakers, such as English, have been found to be massively simpler than languages with small populations (fewer than 100,000 people) which are spoken only in one region of the world. One of the main factors seems to be that the large global languages are under pressure to become simpler over time because they are learned by adult learners – who, as we saw earlier, are not very good at learning the complexities of language when compared with child learners. Conversely, 'esoteric' languages – the small languages of remote communities – may maintain their linguistic complexity exactly because it facilitates learning by infants; complex morphosyntax seems to provide cues to language structure, and since children are so good at learning such complexities, there is no pressure for it to decrease within a small, closed community. For instance, Levinson (2006) discusses a language called Yélí Dnye, spoken by fewer than 4000 people on a remote island (Rossel

Island) several hundred kilometres off the coast of New Guinea. This language, Levinson suggests, is so complex that it lies at 'the boundaries of learnability': adult incomers to the community cannot learn it, and children seem to need an entire community of speakers to learn it successfully, so that if their parents migrate, the offspring may fail to acquire the language fully. It seems, then, that some languages genuinely are more complex than others.

Having reached the end of this book, you should be clear that complexity lies in differing areas of the grammar from language to language. But in standard circumstances (living within a normal linguistic community) children seem to learn each language *as a system* with equal ease, as far as is known. Certainly, there are no languages so hard that their speakers don't become fluent until they're 18 years old. So we can say that whatever complexities a language throws at children, they can cope. Does that mean that all languages are 'the same' in terms of their inherent difficulty for children? I leave this as an open question.

- Do all languages manifest broadly the same syntactic properties?

From reading this book, you will know that, on the face of things, there is a great deal of syntactic and morphosyntactic diversity between languages. Does that mean that languages can vary from each other at random, differing in essentially any way? Recently, some eminent linguists have suggested that this is the case (Evans and Levinson 2009). These authors reject the idea that languages are built to a universal pattern, citing many examples of 'esoteric' data that are not common to all languages. They claim that languages can differ in fundamental ways, resulting in a 'jungle' of linguistic complexity. Equally, there is, in their view, no language faculty – no innate template for language learning that is shared by all members of our species. But this seems to hugely overstate the case. Since all normal children are able to learn a language or several languages very quickly indeed, and without any instruction, and since no other species can achieve anything remotely similar, it seems wrong to deny that we are biologically pre-programmed for language learning.

A very different view to that of Evans and Levinson is outlined by linguists Ray Jackendoff and Peter Culicover (Jackendoff 2002; Culicover and Jackendoff 2005). These linguists suggest that what is often termed 'universal grammar' – the biological endowment for language learning in our species – provides a 'toolkit': a set of basic principles for building languages, which each language customizes in its own unique ways. There is no reason to expect that everything the toolkit can build will be found in all languages, and this is clearly correct. But the toolkit constrains what *can* be built; when properly investigated, languages do *not* vary from each other at random, but rather, look extremely similar. For instance, Morcom (2009) investigated whether or not languages all have a distinct lexical class of nouns and verbs, something which has been denied by certain linguists. She looked at the most controversial languages, and discovered that in each case, there was indeed a distinct noun word class and a distinct verb word class. Careful investigation of this nature by trained linguists often uncovers patterns that are not obvious on the surface.

Having finished revising this book for the third edition, I am left with the impression that despite the very evident cross-linguistic diversity in syntax and morphology, the languages of the world are similar in many crucial ways. All distinguish several word classes (Chapter 2) and, among these word classes, it seems that all languages have a class of nouns and a separate class of verbs. All languages have predicates and all have participants in the event denoted by the predication. All languages have ways of negating clauses, of asking questions, of giving commands. Most languages (perhaps all) distinguish between simple sentences and complex sentences, although not all languages make use of the kind of subordination that is familiar from European languages, as we saw in Chapter 3. All languages have heads which, together with their dependents, form phrases – the constituents of sentences; these were the topics of Chapters 4, 5 and 6. Although languages do not share the same set of constituents, the same sorts of test for constituent structure can be applied in all languages. These tests rely on harnessing the native-speaker intuitions which we all have about our native language(s), by calling on us to make grammaticality judgements. Hierarchical constituent structure is a universal linguistic feature, though it is clearly exploited more in some languages than others; free word order languages, such as those discussed in Chapter 6, make less use of constituency. All languages exhibit dependencies between elements in a sentence, such as those examined in Chapters 4, 5, and 8. All languages have at least one method of encoding grammatical relations – constituent order, morphological case or verbal agreement – as we saw in Chapter 6. All languages appear to exploit variations in constituent order or word order to foreground or background elements, to add focus and emphasis, or to show the topic of a sentence. The vast majority of languages have at least some valency-changing processes, and processes of promotion and demotion which change the basic grammatical relations borne by noun phrases (Chapter 7). The remarkable unity among totally unrelated languages is nicely illustrated in Chapter 7 by the applicative construction, which turns up again and again across the world, and which has much the same grammatical effects in each case. And though not all languages have *wh*-movement, all languages have a way of forming *wh*-questions, and probably relative clauses too (Chapter 8).

My overall impression, then, is that the syntactic diversity among languages from different families and different regions of the world is not trivial, but that the overwhelming homogeneity which exists between languages is far more impressive. In particular, when we look at language isolates (languages with no known relatives) and find that they too utilize the same ‘toolkit’, it seems safe to say that languages are, unmistakably, amazingly similar in design.

9.4 LAST WORDS: MORE SYNTAX AHEAD

My feeling when I started writing the first edition of this book (published by Arnold/OUP in 1998) was that there is an awful lot of syntax out there in the world, much of it rather daunting. This is a view that students of syntax often appear to share! I hope that by now you are familiar with many of the basic concepts needed in order to

understand the ways in which syntax operates in the natural languages of the world, and feel less daunted by its complexities. There is certainly much more syntax out there than a short book can cover, but my intention has been to introduce you to the major syntactic constructions found in the world's languages, and to the main ideas, terms and concepts used by linguists to discuss syntax.

Of course, the syntax part of the grammar of a language doesn't exist in isolation. We have seen in nearly every chapter how it interacts both with form (morphology) and meaning (semantics). Not all languages have much morphology, as I've often noted, in the sense of having variations in the form of words. However, many languages use morphology to signal the kinds of syntactic processes that I've talked about in this book, such as the formation of passives or of causatives. Many languages use morphology – case-marking, verbal agreement, or both – to distinguish between the core participants in a clause, although some languages rely almost entirely on word order to do this. Matters morphological have arisen over and over throughout this book, and if these have interested you, you may wish to move on to a specialized book on the topic, such as Lieber (2010).

Although the topics of semantics, pragmatics, and discourse are beyond the scope of this book, all of these areas also critically interact with syntax in all languages. For instance, there are important discourse factors involved in the selection of syntactic constructions; as an example, we saw in Chapter 7 how the need to allow noun phrases to be coreferential with previous noun phrases within the sentence can give rise to constructions such as the passive and the antipassive.

Having completed this introduction, you are now ready to further your study of syntax. There are various (overlapping) paths your study might take. One is to look at descriptions of languages, studying grammars written by linguists. In a good grammar, the chapters on syntax and morphology should be very prominent, and should ideally cover (at least) all the areas we've seen in this book: word classes, grammatical categories (Chapter 2); simple sentences and complex sentences (Chapter 3); heads and their dependents, head-initial or head-final syntax, head-marking or dependent-marking morphology (Chapter 4); constituent structure (Chapter 5); case, agreement, constituent order and grammatical relations (Chapter 6); syntactic processes which change grammatical relations, such as passives and/or antipassives, causatives and applicatives (Chapter 7); and *wh*-questions, relative clauses, and focus constructions (Chapter 8).

You might also take a course in linguistic fieldwork, which will build on the knowledge gained throughout this text, and might ultimately lead you to investigate the grammar of languages as yet undescribed (of which there are many).

A further way your study might proceed is by looking at syntactic theory. In order to explain the syntactic differences and similarities between languages, linguists need first to know how alike (and unlike) the world's languages are. This requires good descriptions of the sort mentioned earlier. Most linguists want not merely to describe languages in isolation, however, but to discover the ways in which their structures are related, even when there are no genetic relationships between the languages. For instance, the morphology and syntax in the majority of languages operates on the basis of either the nominative/accusative system or the ergative/

absolute system, with the former predominating cross-linguistically, as we saw in Chapter 6. Why do languages generally 'choose' one system or the other as their major system, in spite of the existence of several other logical possibilities? The likelihood is, as I suggested in Chapter 6, that the most economical way of, say, distinguishing between grammatical relations is to use one or other of these major systems.

Such economies in the grammar of a language are of interest to theoretical linguists, in part because we hope they will ultimately tell us something about how children can learn their native languages so quickly, regardless of all the complexities that exist. Linguists are also interested in language as a faculty unique to one species, *Homo sapiens*, and in addition, through the study of the human language faculty, we seek to discover more about the remarkable properties of the brain and of human cognition.

Sources of data used in examples

Data sources for the exercises are given in the exercise section of each chapter. Data not listed below are taken from personal knowledge or research or have been obtained from language consultants.

Chapter 1

Indonesian: Foley and Van Valin (1985).

Japanese: Kuno (1978); also in Chapters 2, 3 and 4.

Kwamera: Lindstrom and Lynch (1994); also in Chapter 2.

Rapa Nui: Chapin (1978).

Chapter 2

Akan: Schachter (1985).

Breton: Press (1986); also in Chapter 3.

Chadian Arabic: Abu-Absi (1995); also in Chapters 3 and 4.

ChiBemba: Chung and Timberlake (1985).

Chichewa: Baker (1988); also Chapter 7.

Chinese: Li and Thompson (1978); also in Chapters 4 and 6.

Chinook: Silverstein (1974).

Gunin: McGregor (1993).

Irish: Ó Siadhail (1989); also in Chapters 3, 5 and 6.

Jarawara: Dixon (2004b).

Mbalanhu: Fourie (1993).

Northern Sotho: Louwrens et al. (1995); also in Chapter 6.

Saliba: Mosel (1994); also in Chapter 4.

Yimas: Foley (1991); also in Chapter 8.

Chapter 3

Bare: Aikhenvald (1995); also in Chapters 4 and 7.

Chinese: Thompson (1973), cited in Foley and Van Valin (1984).

Comanche: Charney (1993), cited in Whaley (1997).
Evenki: J. Payne (1985a).
Greek: Horrocks (1983); Terzi (1992) cited in Baker (1996).
Irish: McCloskey (1979).
Japanese: Tsujimura (1996); also in Chapter 8.
Kambera: Klamer (1994); also in Chapters 4 and 6.
Ndyuka: Huttar and Huttar (1994); also in Chapter 4.
Nupe: Thompson and Longacre (1985).
Portuguese: Raposo (1987).
Swedish: Börjars (1991).
Ukrainian: Danylenko and Vakulenko (1995).
Vagala: Pike (1967), cited in Foley and Van Valin (1984).
Yoruba: Bamgbose (1974), cited in Foley and Van Valin (1984).
Wappo: T. Payne (1997).

Chapter 4

Ayacucho: Adelaar (2004).
Chechen: Nichols (1986).
Evenki: Nedjalkov (1997).
Mangga Buang: T. Payne (1997).
Marathi: Pandharipande (1997).
Southern Tiwa: Allen and Frantz (1978, 1983); Allen *et al.* (1984).
Tinrin: Osumi (1995); also in Chapter 8.
Turkish: Kornfilt (1997); also in Chapter 7.
Tzutujil: Dayley (1981) cited in Nichols (1986).

Chapter 5

Basque: Rebuschi (1989); Hualde *et al.* (1994); also in Chapter 6.
Malagasy: Keenan (1978).
Persian: J. Payne (1985b).

Chapter 6

Abaza: Allen (1956) cited in Dixon (1994).
Ancient Greek: Blake (2001a).
Apalai: Koehn and Koehn (1986).
Apurinã: Derbyshire and Pullum (1981).
Biri: Beale (1974) cited in Dixon (2002).
Dyirbal: Dixon (1994); also in Chapter 7.

Gilbertese: Keenan (1978).
Icelandic: Andrews (1985); Platzack (1987); Sigurðsson (1991); Zaenen *et al.* (1985).
Irish: Stenson (1981).
Kalkatungu: Blake (2001b).
Lezgian: Haspelmath (1993).
Malagasy: Keenan (1976).
Pitta-Pitta: Blake (1979).
Tagalog: Andrews (1985); Foley and Van Valin (1984); Schachter (1976).
Turkish: Comrie (1985a).
Warlpiri: Hale (1973, 1983); Austin and Bresnan (1996).

Chapter 7

Chamorro: Gibson (1980) cited in Baker (1988); Cooreman (1988) cited in Palmer (1994).
Chukchee: Kozinsky *et al.* (1988) cited in Palmer (1994).
Gilbertese: Keenan (1985a).
Greek: Joseph and Philippaki-Warbuton (1987) cited in Song (1996).
Indonesian: Chung (1976/1983).
Inuktitut (Greenlandic): Woodbury (1977).
Japanese: Iwasaki (2002); also in Chapter 8.
Korean: Sohn (1999).
Mam: England (1983a, b) cited in Manning (1996) and Palmer (1994).
Songhai: Comrie (1985a).
Tzotzil: Foley and Van Valin (1985).
Yidiny: Dixon (1977).

Chapter 8

Bambara: Creissels (2000).
Bulgarian: Boeckx and Grohmann (2003).
Chinese: Boeckx and Grohmann (2003).
Frisian: Radford (1988).
Hausa: Abraham (1959) cited in Keenan and Comrie (1977).
Hungarian: Kenesei *et al.* (1998).
Koromfe: Rennison (1997).
Lenakel: Lynch (1998).
Malagasy: Sabel (2003).
Moroccan Arabic: Radford (1988).
Motu: Lynch (1998).
Nakanai: Lynch (1998).
Serbo-Croatian: Stjepanović (2003), Boeckx and Grohmann (2003).

Chapter 9

Welsh: Additional data can be found in Borsley *et al.* (2007).

Glossary

This glossary contains brief definitions of some of the most important (and most difficult) terms and concepts used in the text. The definitions are intended to be a reminder rather than the last word on any given concept, and are thus largely based on examples from English where possible. More extensive cross-linguistic discussion, with data, can be found by looking up the terms in the subject index. For a more comprehensive listing, I recommend that the reader obtain a good dictionary of linguistics, such as Crystal (2008) 6th edition or Matthews (2007), or a more detailed text such as Hurford (1994).

absolutive case: The case of the two CORE ARGUMENTS S and O in an ERGATIVE/ABSOLUTIVE language. If overtly marked, may be indicated via case-marking or by cross-referencing (AGREEMENT) on the verb.

accusative case: The case of the CORE ARGUMENT O in NOMINATIVE/ACCUSATIVE languages. Centrally, the case assigned to DIRECT OBJECTS in languages such as Latin, Russian and German. May be used for other ARGUMENTS of verbs or ADPOSITIONS.

adjunct: A function represented by optional modifying phrases of various classes, e.g. *The post arrived promptly* (adverb phrase) / *in good time* (PP)/*this morning* (NP). Adjuncts are optional modifiers to a HEAD, i.e. not selected by a head, and typically have a rather loose relationship with the head that they modify.

adposition: A cover term for POSTPOSITION and PREPOSITION.

adverb: Typically, an optional modifier to a verb, an adjective, or another adverb. In English, a member of the word class whose central members are recognized by the *-ly* suffix (e.g. *gently, slowly, happily*). Note that not all adverbs have this suffix (e.g. *soon*), and that not all *-ly* words are adverbs (e.g. *friendly, ungodly*). Some adverbs are not optional, but are part of a verb's ARGUMENT structure, e.g. *badly* in *Kim treats Lee badly*.

adverbial: The traditional term for ADJUNCT, i.e. an optional modifying phrase. The term adverbial refers to a function which may be filled by phrases of various classes. Thus, not all adverbials are ADVERBS, and nor do all adverbs necessarily fulfil the adverbial function.

agreement: The marking of various morphosyntactic properties of a HEAD (such as person, number and gender) on the dependents of that head. For instance, within an NP, DETERMINERS and attributive adjectives often agree with the number and gender of the head noun. An example from French, illustrating agreement in gender, is *le livre vert* (the.M book(.M) green.M 'the green book') vs. *la porte verte* (the.F door(.F) green.F 'the green door'). In English, demonstratives such as *this* vs. *these* agree with the head noun in number: *this book, these books*.

agreement, verb (see also CROSS-REFERENCING): The occurrence of inflections or other morphological changes in the verb which reflect the morphosyntactic properties (such as person, number, gender) of some or all of the ARGUMENTS of the verb. For instance,

- many European languages have verbs which agree with their SUBJECTS, particularly in person and number. In English, a small amount of subject/verb agreement occurs in the third person singular, present tense only: *She sings*.
- antipassive: A syntactic process which occurs only in ERGATIVE languages, and which changes the basic GRAMMATICAL RELATIONS borne by CORE ARGUMENTS in the following way: demotes or deletes an O, and promotes an A to an S.
- applicative: A construction which creates two OBJECTS, a primary and a secondary object, from an NP-PP construction. In English, refers to the dative movement construction, which relates *Kim sent the parcel to Sue* with *Kim sent Sue the parcel*.
- argument: A phrase selected by a HEAD verb, adposition or other class of head. Typically refers to the set of obligatory dependents of a verb. INTRANSITIVE verbs have one argument; TRANSITIVE verbs have two, and so on.
- auxiliary: Sometimes termed a 'helping verb'. An element occurring in many, though not all, languages, and which represents the same type of grammatical information as is represented on verbs, e.g. tense, aspect, person/number etc. In English, two major classes: modals (*can, must, will* etc.) and aspectual auxiliaries (*have, be*).
- case-marking: The appearance of morphology on the NP ARGUMENTS of a HEAD (verb, adposition etc.) which marks the relationship each NP has with that head. Two major systems occur cross-linguistically: NOMINATIVE/ACCUSATIVE and ERGATIVE/ABSOLUTIVE.
- causative: A verb denoting a meaning such as 'cause' or 'make' someone do something. May be represented by a lexical verb or via verbal morphology. A causative agent is added to the verb's ARGUMENT structure.
- clause: The central unit of syntax: a 'sentence'. Contains a single PREDICATE.
- clause, embedded *or* subordinate: In a COMPLEX SENTENCE, any clause which is not the main clause. In English, recognized by characteristics which include the inability to take SUBJECT/AUXILIARY INVERSION, and the ability to take a COMPLEMENTIZER such as *that, whether*.
- clause, matrix *or* main: In a COMPLEX SENTENCE, the CLAUSE to which the clause below is subordinate.
- cleft: A type of focus construction which in English takes the form *It is/was [phrase] that Y*, e.g. *It was [last night] that we celebrated the happy event*. The portion bracketed is also shown to be a constituent.
- complement: A phrase which is selected by a HEAD, and is often obligatory. Complements typically have a close relationship with the head they modify. In English, the complements to heads generally immediately follow the head in neutral constituent order. However, the ARGUMENT of the verb which is the SUBJECT is also often considered a complement.
- complementizer: A word from a small class of grammatical items which introduce a CLAUSE, such as English *that* or *whether*. May indicate the tense or FINITENESS of the clause. In HEAD-FINAL languages, a complementizer often follows the clause.
- complex sentence: A sentence containing more than one CLAUSE. Includes co-ordinated clauses, but is more generally used to refer to clausal subordination.
- copula: A linking word with relatively little semantic content, such as the verb *be*, linking the SUBJECT and the PREDICATE. An English example, where the subject is *that huge dog*: *That huge dog is horribly wet and muddy*.
- core arguments: The SUBJECT of an INTRANSITIVE verb (S), and the subject (A) and DIRECT OBJECT (O) of a TRANSITIVE verb.
- cross-referencing (*see also* AGREEMENT): In many languages, verbs are morphologically marked with pronominal affixes which encode information about the verb's arguments.

- This phenomenon is known as cross-referencing. The term cross-referencing often indicates that the verb alone can constitute a sentence, and is thus sometimes distinguished from verbal agreement, a term indicating that the verb has independent arguments with which the head agrees. Extensive cross-referencing is found in HEAD-MARKING languages. Some linguists regard cross-referencing as part of agreement.
- dativ:** A CASE often used to mark the indirect object in a language with extensive case-marking; also often used for experiencer SUBJECTS.
- dependent-marking** (*see also* HEAD-MARKING): A language or construction which shows the relationship between a HEAD and its dependents by marking the dependents, rather than by marking the head. Classic dependent-marking is CASE-MARKING on the CORE ARGUMENTS of a verb or the COMPLEMENT of an ADPOSITION.
- determiner:** A member of a small, closed class of words which co-occur with a HEAD noun, and form a noun phrase. English examples include *the, a, this, that, these, those*. Considered by some linguists to constitute the head of the nominal phrase (hence, forming a 'determiner phrase') rather than a dependent to a head noun.
- direct object** *or* **object:** The O ARGUMENT of a TRANSITIVE verb is known as the (direct) object. The COMPLEMENT of an ADPOSITION is also known as its object.
- ditransitive verb:** A verb of the *give, send, take* type, which has three ARGUMENTS: [*Me*] *sent* [*the cake*] [*to her auntie*].
- echo question:** In English, a construction of the type: *She saw who at the ice-rink last night?*
- ellipsis:** A construction in which some portion is omitted when it can be understood from the context. Can often be used as a test for constituency.
- ergative/absolutive:** A language which indicates via CASE-MARKING and/or verb AGREEMENT the pairing of the S and O ARGUMENTS (known as ABSOLUTIVE) as opposed to the A argument (known as ERGATIVE) of the verb. Such languages are often termed simply 'ergative'.
- ergative case:** The case of the CORE ARGUMENT A in an ERGATIVE language. May be indicated via case-marking or by cross-referencing (AGREEMENT) on the verb.
- ergative language** *see* ERGATIVE/ABSOLUTIVE.
- finite verb** *and* **finite auxiliary:** One marked for such grammatical categories as tense, aspect, AGREEMENT with the verb's ARGUMENTS. May not necessarily bear overt inflections, e.g. in English *I/you/we/they sing*; *We must leave*. English requires a finite verb or auxiliary in root and independent clauses.
- grammatical relations** (*or* GRAMMATICAL FUNCTIONS): The functions fulfilled by the ARGUMENTS of a verb or ADPOSITION, or other HEAD. Examples are SUBJECT, OBJECT, indirect object.
- head:** The word which gives a phrase its word class; for instance, the verb in the VP, the PREPOSITION in the PP, the noun in the NP, and so on. Determines the meaning and grammatical properties of the phrase it heads. May require its DEPENDENTS to agree with it in terms of grammatical categories such as number and gender. It may sometimes be necessary to distinguish between the syntactic head, which determines the word class of the phrase, and the semantic head, which determines its central meaning.
- head-final:** A phrase in which the HEAD follows its COMPLEMENT(s). For instance, in a POSTPOSITION phrase such as Japanese *sanfuranshisuko made* (literally, *San Francisco to*), the P follows the postpositional object NP. Typically, languages are either predominantly head-final, meaning that the head follows the complement in all major phrase types, or else predominantly HEAD-INITIAL.
- head-initial:** A phrase in which the HEAD precedes its COMPLEMENT(s). For instance, the head verb precedes the direct object, and the head preposition precedes the prepositional

- object. Typically, languages are either predominantly head-initial, meaning that the head precedes the complement in all major phrase types, or else predominantly HEAD-FINAL.
- head-marking (see also DEPENDENT-MARKING): A language or construction which shows the relationship between a HEAD and its dependents by marking the head, rather than by marking the dependents. Classic head-marking gives rise to extensive verbal CROSS-REFERENCING: the verb is morphologically marked to reflect the grammatical categories of its dependents.
- infinitive: A NON-FINITE verb form. In English, the bare (uninflected) form of the verb which is used in the frames *Kim must* ___ (*that*) and *He needs to* ___ (*that*). May have a special marker in some languages, such as French *-er, -ir, -re*.
- intransitive verb: A verb taking just one ARGUMENT, namely its SUBJECT. Examples in English are *expire, disappear*. May have optional modifiers, e.g. *The permit expires in two days*.
- inversion, subject/auxiliary: The construction used in English to ask yes/no questions, e.g. *Will Kim be there later?*, in which an AUXILIARY moves to the left of the SUBJECT. Also occurs in *wh*-questions in English, e.g. *Where will Kim be then?*
- matrix: see CLAUSE, MATRIX OR MAIN.
- nominalization: A grammatical process which turns a word of a different word class into a noun. Typically, refers to the process by which verbs are turned into nouns, e.g. *approve/ approval; announce/announcement*.
- nominative/accusative language: A language which indicates via CASE-MARKING and/or verb AGREEMENT the pairing of the S and A ARGUMENTS (i.e. all SUBJECTS) as opposed to the O argument (the OBJECT) of the verb. Also refers to languages with little or no overt marking (e.g. English, Chinese) which organize aspects of their syntax along the same alignment, i.e. SA vs. O.
- nominative case: The case of the CORE ARGUMENTS S and A ('SUBJECTS') in NOMINATIVE/ACCUSATIVE languages. Is not necessarily shown overtly by any morphological CASE-MARKING.
- non-finite verb: Centrally, one which is not marked either for tense or person/number/gender contrasts. A non-finite clause contains only non-finite verbs or non-finite AUXILIARIES.
- object see DIRECT OBJECT.
- participle: A term used for certain NON-FINITE verb forms, but excluding the INFINITIVE. In English, refers to the *-ing* form of the verb in *Kim is sleeping* and the *-ed/-en* form of the verb (the past participle) in *Kim has taken/finished the cheese*.
- passive: A construction which changes the GRAMMATICAL RELATIONS of CORE ARGUMENTS in the following way: the original SUBJECT of an active sentence is demoted or deleted, and the object of the verb is promoted to the GRAMMATICAL RELATION of subject. The valency of a transitive verb is reduced, since it now contains a subject but no object. The change in grammatical relations is marked by changes to the verbal morphology. An example in English would be *Kim stole the cheese* (active) and *The cheese was stolen by Kim* (passive).
- postposition see PREPOSITION.
- predicate: Used in two different senses. May refer just to the verbal element in a CLAUSE, or to the verb and all its modifiers (a VP). There are also non-verbal predicates, for instance those headed by a noun or adjective.
- preposition and postposition: Typically, small words indicating location in time and space, such as in English *on, in, at, under, over, through, beside* etc. May be transitive or intransitive, i.e. may or may not have an OBJECT.

- proform: A word which stands for a full phrase. For instance, a pronoun replaces a full noun phrase. The existence of a proform for a given phrase can be used as a test for constituency.
- relative clause: An optional subordinate CLAUSE used to modify a HEAD noun. Restricts the possible referents of that noun to just the subset which the speaker wishes to refer to. For instance, in *Kim liked the dentist who she saw the other day*, there may be other dentists whom Kim does not like.
- semantic roles: Roles such as agent, theme or patient, goal and experiencer which are taken by NP ARGUMENTS of a HEAD, especially a head verb. Specific semantic roles are determined by lexical properties of the verb.
- subject: A GRAMMATICAL RELATION which refers to the grouping of the A and S ARGUMENTS of a verb. In English, subject pronouns have a special CASE-MARKING in the first and third person, i.e. *I, we, he/she, they*. Cross-linguistically, prototypical subjects are agents, but subjects may bear numerous other SEMANTIC ROLES.
- transitive verb: A verb taking two ARGUMENTS, typically referred to as the SUBJECT and (DIRECT) OBJECT. Examples of verbs in English which must be transitive are *assassinate, uncover*.

References

- Aarts, Bas (2008). *English syntax and argumentation*, 3rd edn. Basingstoke: Palgrave Macmillan.
- Abraham, Roy C. (1959). *The language of the Hausa people*. London: University of London Press.
- Abu-Absi, Samir (1995). *Chadian Arabic*. Languages of the world/Materials 21. Munich and Newcastle: Lincom Europa.
- Adelaar, Willem F. H. (2004). *The languages of the Andes*. Cambridge: Cambridge University Press.
- Aikhenvald, Alexandra (1995). *Bare*. Languages of the world/Materials 100. Munich and Newcastle: Lincom Europa.
- Aikhenvald, Alexandra, Dixon, R. M. W. and Onishi, Masayuki (eds) (2001). *Non-canonical marking of subjects and objects*. Amsterdam: John Benjamins.
- Allen, Barbara J. and Frantz, Donald G. (1978). Verb agreement in Southern Tiwa. *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society*: 11–17.
- Allen, Barbara J. and Frantz, Donald G. (1983). Advancements and verb agreement in Southern Tiwa. In David Perlmutter (ed.) *Studies in relational grammar* 1. Chicago, IL, and London: University of Chicago Press, 303–314.
- Allen, Barbara J., Gardiner, Donna and Frantz, Donald G. (1984). Noun incorporation in Southern Tiwa. *International Journal of American Linguistics* 50: 292–311.
- Allen, W. Sidney (1956). Structure and system in the Abaza verbal complex. *Transactions of the Philological Society* (1956): 127–176.
- Amberber, Mengistu (2000). Valency-changing and valency-encoding devices in Amharic. In Dixon and Aikhenvald (eds), 312–322.
- Anderson, Stephen (1976). On the notion of subject in ergative languages. In Li (ed.), 1–23.
- Anderson, Stephen (1985). Inflectional morphology. In Shopen (ed.), Volume III, 150–201.
- Andrews, Avery (1985). The major functions of the noun phrase. In Shopen (ed.), Volume I, 62–154.
- Asher, R. E. and T. C. Kumari. 1997. *Malayalam*. London and New York: Routledge.
- Austin, Peter and Bresnan, Joan (1996). Non-configurationality in Australian aboriginal languages. *Natural Language and Linguistic Theory* 14: 215–268.
- Baker, Mark (1988). *Incorporation: A theory of grammatical function changing*. Chicago, IL, and London: University of Chicago Press.
- Baker, Mark (1996). *The polysynthesis parameter*. Oxford and New York: Oxford University Press.
- Baker, Mark (2001). *The atoms of language: The mind's hidden rules of grammar*. New York: Basic Books.
- Baker, Mark (2003). *Lexical categories: Verbs, nouns, and adjectives*. Cambridge: Cambridge University Press.
- Bamgbose, Ayo (1974). On serial verbs and verbal status. *Journal of West African Languages* 9: 17–48.

- Beale, Anthony (1974). A grammar of the Biri language of North Queensland. Ms. Australian National University.
- Blake, Barry J. (1977). *Case marking in Australian languages*. Canberra: Australian Institute of Aboriginal Studies.
- Blake, Barry J. (1979). Pitta-Pitta. In Dixon and Blake (eds), 182–342.
- Blake, Barry J. (2001a). *Case*, 2nd edn. Cambridge: Cambridge University Press.
- Blake, Barry J. (2001b). The noun phrase in Australian languages. In Simpson et al. (eds), 415–425.
- Blau, Joyce and Barak, Veysi (1999). *Manuel de kurde (kurmanji)*. Paris: L'Harmattan.
- Blevins, James (2003). Passives and impersonals. *Journal of Linguistics* 39: 473–520.
- Boeckx, Cedric and Grohmann, Kleanthes K. (2003). Introduction. In Boeckx and Grohmann (eds), 1–15.
- Boeckx, Cedric and Grohmann, Kleanthes K. (eds) (2003). *Multiple wh-fronting*. Amsterdam and Philadelphia, PA: John Benjamins.
- Börjars, Kersti (1991). Complementation in the Scandinavian languages. In Nigel Vincent and Kersti Börjars (eds), *Complement structures in the languages of Europe – more preliminary surveys*. EURO TYP Working Papers (European Science Foundation, Programme in Language Typology) III: 2, 65–91.
- Börjars, Kersti and Burridge, Kate (2010). *Introducing English grammar*, 2nd edn. London: Hodder Education.
- Borsley, Robert D., Tallerman, Maggie and Willis, David (2007). *The syntax of Welsh*. Cambridge: Cambridge University Press.
- Burton-Roberts, Noël (2010). *Analysing sentences: An introduction to English syntax*, 3rd edn. Harlow: Longman.
- Campbell, Lyle (2000). Valency-changing derivations in K'iche'. In Dixon and Aikhenvald (eds), 223–281.
- Chapin, Paul G. (1978). Easter Island: A characteristic VSO language. In Lehmann (ed.), 139–168.
- Charney, Jean O. (1993). *A grammar of Comanche*. Lincoln, NE: University of Nebraska.
- Chomsky, Noam (1977). On *wh*-movement. In Peter Culicover, Thomas Wasow and Adrian Akamajian (eds), *Formal syntax*. New York: Academic Press, 77–132.
- Chomsky, Noam (1986). *Knowledge of language: Its nature, origin and use*. New York: Praeger.
- Chung, Sandra (1976). An object-creating rule in Bahasa Indonesia. *Linguistic Inquiry* 7: 1–37. Reprinted in David Perlmutter (ed.) (1983), *Studies in relational grammar* 1. Chicago and London: University of Chicago Press, 219–271.
- Chung, Sandra and Timberlake, Alan. (1985). Tense, aspect and mood. In Shopen (ed.), Volume III, 202–258.
- Clamons, Robbin, Mulkern, Ann E., Sanders, Gerald and Stenson, Nancy (1999). The limits of formal analysis. Pragmatic motivation in Oromo grammar. In Michael Darnell, Edith Moravcsik, Frederick J. Newmeyer, Michael Noonan and Kathleen Wheatley (eds), *Functionalism and formalism in linguistics. Volume II: Case studies*. Amsterdam: John Benjamins, 59–76.
- Comrie, Bernard (1976). *Aspect*. Cambridge: Cambridge University Press.
- Comrie, Bernard (1985a). *Tense*. Cambridge: Cambridge University Press.
- Comrie, Bernard (1985b). Causative verb formation and other verb-deriving morphology. In Shopen (ed.), Volume III, 309–348.
- Comrie, Bernard (1989). *Language universals and linguistic typology*, 2nd edn. Oxford: Basil Blackwell.
- Comrie, Bernard and Keenan, Edward L. (1979). Noun phrase accessibility revisited. *Language* 55: 649–664.

- Cook, Anthony (1988). Participle sentences in Wakiman. In Peter Austin (ed.), *Complex sentence constructions in Australian languages*. Amsterdam and Philadelphia, PA: John Benjamins, 69–95.
- Cooreman, Ann (1988). The antipassive in Chamorro: Variations on a theme of transitivity. In Shibatani (ed.), 561–593.
- Corbett, Greville (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville (2000). *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville (2006). *Agreement*. Cambridge: Cambridge University Press.
- Creissels, Denis (2000). Typology. In Heine and Nurse (eds), 231–258.
- Creissels, Denis (2008a). Remarks on split intransitivity and fluid intransitivity. In Olivier Bonami and P. Cabredo Hofherr (eds), *Empirical issues in syntax and semantics 7*. Paris: Colloque de Syntaxe et Sémantique à Paris, 139–168.
- Creissels, Denis (2008b). Direct and indirect explanations of typological regularities: The case of alignment variations. *Folia Linguistica* 42: 1–38.
- Croft, William A. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. New York and Oxford: Oxford University Press.
- Culicover, Peter W. and Jackendoff, Ray (2005). *Simpler syntax*. Oxford: Oxford University Press.
- Crystal, David (2000). *Language death*. Cambridge: Cambridge University Press.
- Crystal, David (2008). *A dictionary of linguistics and phonetics*, 6th edn. Oxford: Wiley Blackwell.
- Danylenko, Andrii and Serhii Vakulenko (1995). *Ukrainian*. Languages of the world/Materials 5. Munich and Newcastle: Lincom Europa.
- Davenport, Mike and Hannahs, S. J. (2010). *Introducing phonetics and phonology*, 3rd edn. London: Hodder.
- Dayley, Jon P. (1981). *A Tzutujil grammar*. Unpublished PhD dissertation, University of California, Berkeley.
- Deering, Nora and Delisle, Helga (1976). *Mohawk: A teaching grammar*. Kahnawake, Quebec: Thunderbird Press.
- Derbyshire, Desmond C. and Pullum, Geoffrey K. (1981). Object initial languages. *International Journal of American Linguistics* 47: 192–214.
- Dixon, R. M. W. (1972). *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (1977). *A grammar of Yidiny*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (1979). Ergativity. *Language* 55: 59–138.
- Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (1995). Complement clauses and complementation strategies. In F. R. Palmer (ed.), *Grammar and meaning: essays in honour of Sir John Lyons*. Cambridge: Cambridge University Press, 175–220.
- Dixon, R. M. W. (1997). *The rise and fall of languages*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (2000). A typology of causatives: form, syntax and meaning. In Dixon and Aikhenvald (eds), 30–83.
- Dixon, R. M. W. (2002). *Australian languages: Their nature and development*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (2004a). Adjective classes in typological perspective. In R. M. W. Dixon and Alexandra Y. Aikhenvald (eds), *Adjective classes: A crosslinguistic typology*. Oxford: Oxford University Press, 1–49.
- Dixon, R. M. W. (2004b). *The Jarawara language of Southern Amazonia*. Oxford: Oxford University Press.

- Dixon, R. M. W. and Aikhenvald, Alexandra (eds) (2000). *Changing valency: Case studies in transitivity*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. and Blake, Barry J. (eds) (1979). *Handbook of Australian languages 1*. Canberra: Australian National University and Amsterdam: John Benjamins.
- Donohue, Mark and Brown, Lea (1999). Ergativity: Some additions from Indonesia. *Australian Journal of Linguistics* 19: 57–76.
- Dryer, Matthew S. (1991). SVO languages and the OV:VO typology. *Journal of Linguistics* 27: 443–482.
- Dryer, Matthew S. (1997). Are grammatical relations universal? In Joan Bybee, John Haiman and Sandra A. Thompson (eds), *Essays on language function and language type. Dedicated to T. Givón*. Amsterdam: John Benjamins, 115–143.
- Emonds, Joseph E. (1986). Parts of speech in generative grammar. *Linguistic Analysis* 16: 247–285.
- England, Nora (1983a). Ergativity in Mamean (Mayan) languages. *International Journal of American Linguistics* 49: 1–19.
- England, Nora (1983b). *A grammar of Mam, a Mayan language*. Austin, TX: University of Texas Press.
- Evans, Nicholas and Levinson, Stephen C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32: 429–492.
- Everett, Dan and Kern, Barbara (1997). *Wari'*. London and New York: Routledge.
- Foley, William A. (1991). *The Yimas language of New Guinea*. Stanford, CA: Stanford University Press.
- Foley, William A. and Robert D. Van Valin Jr. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Foley, William A. and Robert D. Van Valin Jr. (1985). Information packaging in the clause. In Shopen (ed.), Volume I, 282–364.
- Fourie, David J. (1993). *Mbalanhu*. Languages of the world/Materials 3. Munich and Newcastle: Lincom Europa.
- Frajzyngier, Zygmunt (2002). *A grammar of Hdi*. Berlin and New York: Mouton de Gruyter.
- Fromkin, Victoria, Rodman, Robert and Hyams, Nina (2007). *An introduction to language*, 8th edn. Boston, MA: Thomson Wadsworth.
- Gazdar, Gerald, Klein, Ewan and Pullum, Geoffrey K. (eds) (1983). *Order, concord and constituency*. Dordrecht: Foris.
- Gibson, Jeanne D. (1980). *Clause union in Chamorro and in universal grammar*. Unpublished PhD dissertation, University of California, San Diego.
- Greenberg, Joseph (1966). Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg (ed.), *Universals of language*, 2nd edn. Cambridge, MA: MIT Press, 73–113.
- Hale, Kenneth L. (1973). Person marking in Walbiri. In Stephen R. Anderson and Paul Kiparsky (eds), *A festschrift for Morris Halle*. New York: Holt, Rinehart, & Winston, 308–344.
- Hale, Kenneth L. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory* 1: 5–47.
- Haspelmath, Martin (1993). *A grammar of Lezgian*. Berlin and New York: Mouton de Gruyter.
- Haspelmath, Martin (2007). Pre-established categories don't exist: consequences for language description and typology. *Linguistic Typology* 11: 119–132.
- Haviland, John (1979). Guugu Yimidhirr. In Dixon and Blake (eds), 26–180.
- Hawkins, John A. (1983). *Word order universals*. New York: Academic Press.

- Heine, Bernd and Nurse, Derek (eds) (2000). *African languages: An introduction*. Cambridge: Cambridge University Press.
- Horrocks, Geoffrey (1983). The order of constituents in Modern Greek. In Gazdar et al. (eds), 95–111.
- Hualde, José Ignacio, Elordieta, Gorka and Elordieta, Arantza (1994). *The Basque dialect of Lekeitio*. Bilbao: University of the Basque Country.
- Huddleston, Rodney and Geoffrey K. Pullum (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Huddleston, Rodney and Geoffrey K. Pullum (2005). *A student's introduction to English grammar*. Cambridge: Cambridge University Press.
- Hudson, Richard A. (1984). *Word grammar*. Oxford: Basil Blackwell.
- Hudson, Richard A. (1987). Zwicky on heads. *Journal of Linguistics* 23: 109–132.
- Hudson, Richard A. (1990). *English word grammar*. Oxford: Basil Blackwell.
- Hudson, Richard A. (2007). *Language networks: The new word grammar*. Oxford: Oxford University Press.
- Hurford, James R. (1994). *Grammar: A student's guide*. Cambridge: Cambridge University Press.
- Huttar, George L. and Mary L. Huttar (1994). *Ndyuka*. London and New York: Routledge.
- Jackendoff, Ray (1993). *Patterns in the mind: Language and human nature*. New York and London: Harvester Wheatsheaf.
- Jackendoff, Ray (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Iwasaki, Shoichi I. (2002). *Japanese*. Amsterdam and Philadelphia, PA: John Benjamins.
- Jones, B. M. and Thomas, Alan R. (1977). *The Welsh language: Studies in its syntax and semantics*. Cardiff: University of Wales Press.
- Joseph, Brian D. and Philippaki-Warbuton, Irene (1987). *Modern Greek*. London: Croom Helm.
- Keenan, Edward L. (1976). Towards a universal definition of 'subject'. In Li (ed.), 303–333.
- Keenan, Edward L. (1978). The syntax of subject-final languages. In Lehmann (ed.), 267–327.
- Keenan, Edward L. (1985a). Passive in the world's languages. In Shopen (ed.), Volume I, 243–281.
- Keenan, Edward L. (1985b). Relative clauses. In Shopen (ed.), Volume II, 141–170.
- Keenan, Edward L. and Comrie, Bernard (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8, 63–99.
- Keenan, Edward L. and Comrie, Bernard (1979). Data on the noun phrase accessibility hierarchy. *Language* 55: 333–352.
- Kenesei, István, Vago, Robert M. and Fenyvesi, Anna (1998). *Hungarian*. London and New York: Routledge.
- Klamer, Marian (1994). *Kamera: A language of Eastern Indonesia*. Amsterdam: Holland Institute of Generative Linguistics.
- Koehn, Edward and Koehn, Sally (1986). Apalai. In Desmond C. Derbyshire and Geoffrey K. Pullum (eds), *Handbook of the Amazonian languages* 1. Berlin: Mouton de Gruyter, 32–127.
- Kornfilt, Jaklin (1997). *Turkish*. London and New York: Routledge.
- Kozinsky, I. Sv., Nedjalkov, V. P. and Polinskaya, M. S. (1988). Antipassive in Chukchee: Oblique object, object incorporation, zero object. In Shibatani (ed.), 651–706.
- Kuno, Susumo (1978). Japanese: A characteristic OV language. In Lehmann (ed.), 57–138.
- Lehmann, Winfred P. (ed.) (1978). *Syntactic typology: Studies in the phenomenology of language*. London: Harvester Wheatsheaf.

- Levinson, Stephen C. (2006). Introduction: The evolution of culture in a microcosm. In Stephen C. Levinson and Pierre Jaisson (eds), *Evolution and culture*. Cambridge, MA: MIT Press, 1–41.
- Li, Charles N. (ed.) (1976). *Subject and topic*. New York: Academic Press.
- Li, Charles N. and Thompson, Sandra A. (1978). An exploration of Mandarin Chinese. In Lehmann (ed.), 223–266.
- Lieber, Rochelle (2010). *Introducing morphology*. Cambridge: Cambridge University Press.
- Lindstrom, Lamont and Lynch, John (1994). *Kwamera*. Languages of the world/Materials 2. Munich and Newcastle: Lincom Europa.
- Lobeck, Anne (2000). *Discovering grammar: An introduction to English sentence structure*. New York and Oxford: Oxford University Press.
- Louwrens, Louis J., Kosch, Ingeborg M. and Kotzé, Albert E. (1995). *Northern Sotho*. Languages of the world/Materials 30. Munich and Newcastle: Lincom Europa.
- Lupyan, Gary and Dale, Rick (2010). Language structure is partly determined by social structure. *PLoS ONE* 5(1): e8559. doi:10.1371/journal.pone.0008559.
- Lynch, John (1998). *Pacific languages: An introduction*. Honolulu, HI: University of Hawai'i Press.
- Lyons, John (1966). Towards a 'notional' theory of the 'parts of speech'. *Journal of Linguistics* 2: 209–236.
- McCloskey, James (1979). *Transformational syntax and model-theoretic semantics: A case study in Modern Irish*. Dordrecht and Boston, MA: D. Reidel Publishing Company.
- McCloskey, James (2000). Quantifier float and *wh*-movement in an Irish English. *Linguistic Inquiry* 31: 57–84.
- McGregor, William B. (1993). *Gunin/Kwini*. Languages of the world/Materials 11. Munich and Newcastle: Lincom Europa.
- McMahon, April M. S. (1994). *Understanding language change*. Cambridge: Cambridge University Press.
- Manning, Christopher D. (1996). *Ergativity: Argument structure and grammatical relations*. Stanford, CA: CSLI Publications.
- Matras, Yaron (2002). *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Matthews, Peter (2007). *The concise Oxford dictionary of linguistics*, 2nd edn. Oxford and New York: Oxford University Press.
- Millar, Robert McColl (2007). *Trask's historical linguistics*. London: Hodder and New York: Oxford University Press.
- Morcom, Lindsay (2009). *The universality and demarcation of lexical categories cross-linguistically*. Unpublished DPhil. thesis, University of Oxford.
- Mosel, Ulrike (1994). *Saliba*. Languages of the world/Materials 31. Munich and Newcastle: Lincom Europa.
- Nedjalkov, Igor (1997). *Evenki*. London and New York: Routledge.
- Nettle, Daniel and Romaine, Suzanne (2000). *Vanishing voices: The extinction of the world's languages*. Oxford and New York: Oxford University Press.
- Nichols, Johanna (1986). Head-marking and dependent-marking grammar. *Language* 62: 56–119.
- Noonan, Michael (1985). Complementation. In Shopen (ed.), Volume III, 42–140.
- Ó Siadhail, Mícheál (1989). *Modern Irish: Grammatical structure and dialectal variation*. Cambridge: Cambridge University Press.
- Osumi, Midori (1995). *Tinrin grammar*. Honolulu, HI: University of Hawai'i Press.
- Otsuka, Yuko (2005). Two derivations of VSO: A comparative study of Niuean and Tongan.

- In Andrew Carnie, Heidi Harley and Sheila Ann Dooley (eds), *Verb first: On the syntax of verb-initial languages*. Amsterdam/Philadelphia, PA: John Benjamins, 65–90.
- Palmer, F. R. (1994). *Grammatical roles and relations*. Cambridge: Cambridge University Press.
- Pandharipande, Rajeswari V. (1997). *Marathi*. London and New York: Routledge.
- Payne, John (1985a). Negation. In Shopen (ed.), Volume I, 197–242.
- Payne, John (1985b). Complex phrases and complex sentences. In Shopen (ed.), Volume II, 3–41.
- Payne, Thomas E. (1997). *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.
- Payne, Thomas E. (2006). *Exploring language structure: A student's guide*. Cambridge: Cambridge University Press.
- Pike, Kenneth (1967). *Grammar as wave*. Georgetown University Monographs on Language and Linguistics 20. Washington, DC: Georgetown University Press, 1–14.
- Pinker, Steven (1994). *The language instinct*. London: Penguin Books.
- Platzack, Christer (1987). The Scandinavian languages and the null-subject parameter. *Natural Language and Linguistic Theory* 5: 377–401.
- Press, Ian (1986). *A grammar of modern Breton*. Berlin: Mouton de Gruyter.
- Pullum, Geoffrey K. and Ladusaw, William A. (1996). *Phonetic symbol guide*, 2nd edition. Chicago, IL, and London: University of Chicago Press.
- Radford, Andrew (1988). *Transformational grammar: A first course*. Cambridge: Cambridge University Press.
- Raposo, Eduardo (1987). Case theory and Infl-to-Comp: The inflected infinitive in European Portuguese. *Linguistic Inquiry* 18: 85–109.
- Rebuschi, Georges (1989). Is there a VP in Basque? In László Marácz and Pieter Muysken (eds), *Configurationality: The typology of asymmetries*. Dordrecht: Foris, 85–116.
- Rennison, John R. (1997). *Koromfe*. London and New York: Routledge.
- Sabel, Joachim (2003). Malagasy as an optional *wh*-fronting language. In Boeckx and Grohmann (eds), 229–254.
- Sampson, Geoffrey, Gil, David and Trudgill, Peter (eds) (2009). *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Schachter, Paul (1976). The subject in Philippine languages: Topic, actor, actor-topic or none of the above. In Li (ed.), 491–518.
- Schachter, Paul (1985). Parts-of-speech systems. In Shopen (ed.), Volume I, 3–61.
- Shibatani, Masayoshi (ed.) (1988). *Passive and voice*. Amsterdam and Philadelphia, PA: John Benjamins.
- Shopen, Timothy (ed.) (1985). *Language typology and syntactic description*. Volume I: *Clause structure*; Volume II: *Complex constructions*; Volume III: *Grammatical categories and the lexicon*. Cambridge: Cambridge University Press.
- Shopen, Timothy (2001). Explaining typological differences between languages. In Simpson et al. (eds), 187–197.
- Siewierska, Anna (2004). *Person*. Cambridge: Cambridge University Press.
- Sigurðsson, Halldór Ármann (1991). Icelandic case-marked PRO and the licensing of lexical arguments. *Natural Language and Linguistic Theory* 9: 327–363.
- Sigurðsson, Halldór Ármann (2006). The Icelandic noun phrase: Central traits. *Arkiv för nordisk filologi* 121: 193–236.
- Silverstein, Michael (1974). Dialectal developments in Chinookan tense–aspect systems: An areal-historical analysis. *International Journal of American Linguistics*, Memoir 29: 44–99.
- Simpson, Jane, Nash, David, Laughren, Mary, Austin, Peter and Alpher, Barry (eds) (2001). *Forty years on: Ken Hale and Australian languages*. Canberra: Pacific Linguistics.

- Sohn, Ho-Min (1999). *The Korean language*. Cambridge: Cambridge University Press.
- Song, Jae Jung (1996). *Causatives and causation: A universal-typological perspective*. London and New York: Longman.
- Song, Jae Jung (2001). *Linguistic typology: Morphology and syntax*. London: Longman.
- Stenson, Nancy (1981). *Studies in Irish syntax*. Tübingen: Gunter Narr Verlag.
- Stjepanović, Sandra (2003). Multiple *wh*-fronting in Serbo-Croatian matrix questions and the matrix sluicing construction. In Boeckx and Grohmann (eds), 255–284.
- Stucky, Susan (1983). Verb phrase constituency and linear order in Makua. In Gazdar et al. (eds), 75–94.
- Terzi, Arhonto (1992). *PRO in finite clauses: A study of the inflectional heads of the Balkan languages*. Unpublished PhD dissertation, City University of New York, NY.
- Thompson, Sandra A. (1973). Resultative verb compounds in Mandarin: A case for lexical rules. *Language* 49, 361–379.
- Thompson, Sandra A. and Longacre, Robert E. (1985). Adverbial clauses. In Shopen (ed.), Volume II, 171–234.
- Tomlin, Russell (1986). *Basic word order: Functional principles*. London: Croom Helm.
- Tsujimura, Natsuko (1996). *An introduction to Japanese linguistics*. Oxford: Blackwell.
- Van Valin, Robert D., Jr. (1985). Case marking and the structure of the Lakhota clause. In Johanna Nichols and Anthony C. Woodbury (eds), *Grammar inside and outside the clause: Some views of theory from the field*. Cambridge: Cambridge University Press, 363–413.
- Watters, John R. (2000). Syntax. In Heine and Nurse (eds), 194–230.
- Whaley, Lindsay J. (1997). *Introduction to typology: The unity and diversity of language*. Thousand Oaks, CA, and London: Sage Publications.
- Woodbury, Anthony (1977). Greenlandic Eskimo: Ergativity and relational grammar. In Peter Cole and Jerry Sadock (eds), *Grammatical relations* (Syntax and Semantics 8), 307–336.
- Zaenen, Annie, Maling, Joan and Thráinsson, Höskuldur (1985). Case and grammatical functions: The Icelandic passive. *Natural Language and Linguistic Theory* 3: 441–483.
- Zwicky, Arnold M. (1985). Heads. *Journal of Linguistics* 21: 1–29.

Language index

The purpose of this list is to give genetic information about each of the languages used in the examples, and give you an idea of where they are spoken. Each entry is structured as follows:

Language name (Branch of family: Family name – Location)

Where necessary (because the language family has a number of different branches) the branch of the family is listed first, and the language family itself follows the colon. The main country or area(s) in which the language is (or was formerly) spoken follows the dash. For instance, Breton is a member of the Celtic branch of the Indo-European language family, and is spoken in France. The genetic affiliation of languages is often controversial, and not all linguists will agree with all the classifications listed. The classifications used here generally follow the practice of the Ethnologue, which can be found at the following URL: <http://www.ethnologue.com>

- Abaza (North Caucasian – Russia, Turkey) 188–9, 208
Akan (Kwa: Niger-Congo – Ghana) 55
Amharic (Semitic: Afro-Asiatic – Ethiopia) 241
Ancient Greek (Greek: Indo-European – Greece) 200
Apalai (Carib – Brazil) 174
Apurinā (Arawakan – Brazil) 174
Ayacucho (Quechuan – Peru) 129
Bambara (Mande: Niger-Congo – Mali) 256–7
Bare (Arawakan – Brazil) 79, 97–9, 122, 238–9
Basque (Language isolate – Basque Country (i.e. parts of Spain, France)) 78, 82, 133, 135, 147, 166, 178, 181, 184, 216, 265–6
Biri (Pama-Nyungan: Australian – Australia) 188, 200
Breton (Celtic: Indo-European – France) 3, 56, 57, 89–90, 104
Bulgarian (Slavonic: Indo-European – Bulgaria) 132, 249, 250
Chadian Arabic (Semitic: Afro-Asiatic – Chad) 40, 75, 122, 131, 132, 133
Chamorro (Malayo-Polynesian: Austronesian – Guam) 214, 219
Chechen (North Caucasian – Chechnya) 125, 132, 136
ChiBemba (Bantoid: Niger-Congo – Tanzania, Zaire, Zambia) 43
Chichewa/Nyanja (Bantoid: Niger-Congo – Malawi, Mozambique, Zambia) 44–5, 225–6, 227–8, 231, 232–3
Chinese Mandarin (Sino-Tibetan – China, Malaysia, Singapore, Taiwan etc.) 12, 18, 37, 55–6, 66, 96, 97, 98, 132, 134, 177, 185, 204, 247–8, 272
Chinook (Penutian – United States) 42
Chukchee (Chukotko-Kamchatkan – Russia) 218
Comanche (Uto-Aztecan – United States) 96
Dyirbal (Pama-Nyungan: Australian – Australia) 132, 182–3, 184, 198, 199, 204, 215, 222, 223–4, 226–7, 228
English (Germanic: Indo-European – global spread) *examples throughout*
Evenki (Tungusic: Altaic – China, Russia) 79, 82, 103, 122, 136, 237–8
Fijian (Malayo-Polynesian: Austronesian – Fiji Islands, New Zealand etc.) 71–2, 168, 174

- French (Romance: Indo-European – Canada, France etc.) 3, 9, 13, 16, 17, 51, 56, 58, 61, 80, 83, 110, 130, 133, 185–6, 207, 228, 230, 248, 250, 291
- Frisian (Germanic: Indo-European – Netherlands) 246
- German (Germanic: Indo-European – Germany) 11, 44, 51, 57, 62, 82, 83, 89, 90, 124–5, 126–7, 132, 133–4, 174, 179, 180, 185, 200, 201, 207, 213, 215, 216–7, 291
- Greek (Greek: Indo-European – Greece) 56, 66, 93, 132, 181, 200, 231–2,
- Gunin/Kwini (Wororan: Australian – Australia) 45
- Guugu Yimidhirr (Pama-Nyungan: Australian – Australia) 238–9
- Hausa (Chadic: Afro-Asiatic – Cameroon, Niger, Nigeria etc.) 259
- Hawrami (Indo-Iranian: Indo-European – Iran) 130
- Hdi (Chadic: Afro-Asiatic – Cameroon, Nigeria) 137
- Hebrew (Semitic: Afro-Asiatic – Israel) 126, 175
- Hungarian (Finno-Ugric: Uralic – Hungary) 56, 136, 166, 255, 260
- Icelandic (Germanic: Indo-European – Iceland) 28, 192–4, 199, 210
- Indonesian (Malayo-Polynesian: Austronesian – Indonesia) 6–7, 12, 66, 225–6, 227, 239–40
- Inuktitut, Greenlandic (Eskimo-Aleut – Greenland) 215, 218
- Irish (Celtic: Indo-European – Ireland) 39, 40, 48, 66, 102–3, 147, 160–1, 174, 176
- Italian (Romance: Indo-European – Italy) 51, 66, 185
- Japanese (Japanese – Japan) 12, 13, 14, 18, 19, 23–4, 55, 65, 70, 87, 102, 111, 121, 124, 126, 176, 179, 185, 201, 214, 215, 242, 247, 248, 249, 250, 254–5, 256, 261
- Jarawara (Arauan – Brazil) 41, 52, 58–9
- Kalkatungu (Pama-Nyungan: Australian – Australia) 203, 235–6
- Kambera (Malayo-Polynesian: Austronesian – Sumba Island) 46, 75, 94–5, 96, 110, 124, 127, 128, 130, 173, 186, 199, 201, 234–5, 264–5
- K'iche'/Quiché (Mayan – Guatemala) 241
- Korean (Language isolate/or possibly Altaic – North Korea, South Korea) 104, 229–30, 255, 261
- Koromfe (Gur: Niger-Congo – Burkina Faso) 245
- Kru (Kru: Niger-Congo – Liberia, Sierra Leone) 30, 104
- Kurmanji (or Kurmanci) Kurdish (Indo-Iranian: Indo-European – Afghanistan, Iran, Iraq etc.) 268–70
- Kwamera (Malayo-Polynesian: Austronesian – Vanuatu) 16, 17–18, 28–9, 59
- Lakhota/Lakota (Siouan – United States) 132, 206–7
- Latin (Italic: Indo-European – widely in medieval Europe) 53, 172–3, 179–80, 184, 199, 200, 201, 217, 291
- Lenakel (Malayo-Polynesian: Austronesian – Vanuatu) 28–9, 260
- Lezgian (North Caucasian – Azerbaijan, Daghestan) 82, 121, 180–1, 184, 194–6, 198, 200, 210
- Lobala (Bantoid: Niger-Congo – Democratic Republic of the Congo) 30
- Makua (Bantoid: Niger-Congo – Mozambique, Tanzania) 205–6
- Malagasy (Malayo-Polynesian: Austronesian – Madagascar) 19, 165, 185, 191, 250–1
- Malay (Malayo-Polynesian: Austronesian – Malaysia) 50, 53, 56–7, 66, 69
- Malayalam (Dravidian – India) 132, 169–70, 263–4
- Mam (Mayan – Guatemala, Mexico) 221, 259
- Mangga Buang (Malayo-Polynesian: Austronesian – Papua New Guinea) 128
- Marathi (Indo-Iranian: Indo-European – India) 130, 131
- Mbalanhu (Bantoid: Niger-Congo – Namibia) 39
- Mende (Mande: Niger-Congo – Sierra Leone) 30
- Middle English (Germanic: Indo-European – British Isles) 3, 8–9, 10
- Mohawk (Iroquoian – Canada, United States) 75, 132, 136, 174
- Moroccan Arabic (Semitic: Afro-Asiatic – Morocco) 246
- Motu (Pidgin, Motu based – Papua New Guinea) 260
- Nakanai (Malayo-Polynesian: Austronesian – Papua New Guinea) 260
- Nakanamanga/Nguna (Malayo-Polynesian: Austronesian – Vanuatu) 71
- Nanai (Tungusic: Altaic – China, Russia) 104
- Northern Sotho (Bantoid: Niger-Congo – South Africa) 51–2, 173, 174, 186–7
- Ndyuka (Creole, English based – Surinam) 75, 122, 123
- Nupe (Benue-Congo: Niger-Congo – Nigeria) 96–7
- Orok (Tungusic: Altaic – Russia) 103–4
- Oromo (Cushitic: Afro-Asiatic – Ethiopia, Kenya, Tanzania) 170–1
- Persian (Indo-Iranian: Indo-European – Afghanistan, Iran) 130, 165
- Pitta-Pitta (Pama-Nyungan: Australian – Australia) 183–4
- Portuguese (Romance: Indo-European – Brazil, Portugal) 81
- Rapa Nui (Malayo-Polynesian: Austronesian – Easter Island) 14, 15

- Romani (Indo-Iranian: Indo-European – numerous European countries, especially in the Balkans) 135, 136
- Russian (Slavonic: Indo-European – Russia) 18, 19, 50, 53, 55, 66, 132, 174, 291
- Saliba/Suau (Malayo-Polynesian: Austronesian – Papua New Guinea) 42, 50–1, 129, 131, 135
- Serbo-Croatian (Slavonic: Indo-European – Bosnia and Herzegovina, Croatia, Serbia) 249, 250
- Songhai/Sonrai (Nilo-Saharan – Burkino Faso, Mali, Niger) 232
- Southwest Tanna (Malayo-Polynesian: Austronesian – Vanuatu) 28–30
- Spanish (Romance: Indo-European – widely in Latin America, Spain, United States) 11–12, 51, 130, 133, 185
- Swahili (Bantoid: Niger-Congo – Kenya, Tanzania) 30, 208–9
- Swedish (Germanic: Indo-European – Sweden) 53–4, 90, 185
- Tagalog (Malayo-Polynesian: Austronesian – Philippines) 196–8, 258
- Tinrin (Malayo-Polynesian: Austronesian – New Caledonia) 30–1, 106–7, 120–1, 254
- Tongan (Malayo-Polynesian: Austronesian – Tonga) 174, 209, 258
- Turkish (Turkic: Altaic – Turkey) 53, 66, 121, 135, 136, 174, 179, 185, 199, 200, 217, 231, 255, 262–3
- Tzotzil (Mayan – Mexico) 132, 215, 216
- Tzutujil (Mayan – Guatemala) 125, 131
- Ukrainian (Slavonic: Indo-European – Ukraine) 75
- Vagala (North Volta-Congo: Niger-Congo – Ghana) 97
- Wari' (Chapakuran – Brazil) 136
- Wappo (Yuki – United States) 92–3
- Warlpiri (Pama-Nyungan: Australian – Australia) 79, 189–90, 199, 200–4, 205
- Welsh (Celtic: Indo-European – Wales) 5, 13, 14, 19, 43, 51, 55, 58, 81, 89, 104, 105, 120, 126, 131, 135, 168–9, 207–8, 213, 245, 258, 259–60, 266–8, 274–81
- Yaqui (Uto-Aztecan – Mexico) 102
- Yalarnnga (Pama-Nyungan: Australian – Australia) 208
- Yidiny (Pama-Nyungan: Australian – Australia) 132, 204, 219–20
- Yimas (Sepik-Ramu – Papua New Guinea) 59–61, 104, 137–9, 254
- Yoruba (Benue-Congo: Niger-Congo – Benin) 97
- Zina Kotoko (Chadic: Afro-Asiatic – Cameroon) 69–70

Subject index

- absolutive case (*see case*)
- Accessibility Hierarchy 258–9, 280
- accusative case (*see case*)
- accusative pattern (*see nominative/accusative pattern*)
- active 22–3, 44–5, 211–24, 229, 279, 294
- adjective 2, 27, 33–5, 37, 38, 51, 55, 56–62, 64, 65, 66, 67, 68, 75, 82, 83, 108, 109, 110, 112, 113, 116, 117, 118, 120, 121, 123, 129–30, 132, 133, 180, 272, 276
- attributive 56–7, 58, 59, 123, 129–30, 133, 291
- predicative 56–7, 59, 62
- tests for adjective status 34, 57–8
- adjunct function 63–4, 87–8, 89, 92, 93, 94, 98, 105, 111–14, 115, 117–18, 119, 134, 145, 148, 149, 158, 244, 247, 256, 272, 291
- adverb 35, 61, 62, 63, 109, 110, 112, 114, 118, 276, 291
- adverbial function (*see adjunct function*)
- adposition (*see also preposition; postposition*) 65, 258, 291, 292, 293
- affix (*see also infix; prefix; suffix*) 15, 23, 45, 61, 66, 79, 81, 123, 124, 142, 196, 204, 214, 221, 231, 232, 234, 254, 265
- agent (*see also semantic role*) 46–7, 48, 138, 177, 190, 196, 197, 198, 216, 217, 229, 279, 292, 295
- agreement (*see also cross-referencing*) 26, 44, 48, 49, 58, 60, 61, 67, 74, 75, 79, 80, 81, 82, 110, 111, 124, 128, 129, 131, 132, 133–4, 173, 177, 184–90, 193, 194, 196, 197, 199, 200–1, 204, 205, 206, 207, 208, 209, 212, 215, 216, 218, 219, 221–2, 231, 233, 269, 273, 275, 276–7, 279, 284, 285, 291, 292, 293, 294.
- accusative (*see also nominative/accusative pattern*) 185–188, 189–90, 201
- ergative (*see also ergative/absolutive pattern*) 188–90, 216
- object/verb 187, 218, 231, 233
- of adjectives with N 58, 110, 129–30, 133
- of adpositions with complement 125–6, 277, 280
- of determiners with N 110, 54–5, 58, 119, 125, 130, 133, 180, 291
- subject/verb 45, 48, 49, 67, 133–4, 185–6, 193, 194, 212, 292
- with more than one argument 184, 186, 187–8, 200, 216, 218
- alignment (of arguments) 179, 180, 182–3, 186, 188–9, 194, 198, 208, 209, 215, 218, 220, 268, 269, 272, 273, 275, 294
- antipassive (*see also passive*) 217–22, 224–5, 227, 229, 233, 272, 273,
- summary of properties of 224–5
- applicative (*see also dative movement*) 205, 225–8, 232, 233, 234, 284, 292
- summary of properties of 226
- argument 39–41, 44–5, 46, 47, 49, 50, 84, 87, 93, 111, 123, 124, 125, 126–7, 128, 148, 149, 172, 177–9, 180, 182, 183, 184–5, 186, 187–8, 189, 190, 191, 194, 195, 196, 198, 199, 200, 201, 202, 203, 206, 207, 212–3, 215, 216, 217, 218, 219, 220, 221, 224, 225, 227, 229, 230, 232, 233, 245, 265, 269, 272, 276, 280, 291, 292, 293, 294, 295
- core 23, 41, 45, 172, 173, 177–9, 183, 184, 185, 190–201, 204, 212, 213, 216, 217, 218, 219, 220, 224, 227, 229, 233, 285, 292
- oblique 213, 214, 218, 219, 222, 225, 226, 227, 228, 232, 239, 268,
- article 15, 53, 54, 180
- definite 15, 53, 54, 180
- indefinite 15, 54
- aspect 15, 41–3, 45, 55–6, 67, 77–8, 79, 80, 82, 83, 84, 98, 183, 275, 292
- habitual 43, 136
- progressive 14, 15, 42–3, 59, 78, 79, 82, 275
- perfect 43, 78, 79, 83, 88–9, 275
- perfective 60, 183, 269
- auxiliary 8–9, 42, 43, 45, 48, 49, 74, 76–8, 79, 80, 81, 82, 83, 84, 91–2, 94, 100–1, 103, 105, 168, 176, 186, 189, 201, 202, 203, 204, 212, 213, 245, 274, 275, 279, 292, 293, 294
- aspectual 77, 84, 275, 292
- modal 43, 77, 80, 100–1, 292
- negative 79, 103
- NICE properties of 100

- birdsong 295
- brackets 21, 24, 25, 26, 61, 85, 86, 87, 102, 108, 112, 114, 119, 135, 141–2, 144, 145, 146, 147, 148, 150, 151, 153, 154, 163, 164, 165, 167, 292
- branch 148–65
- case (*see also* case marking) 48–9, 52–3, 111, 121, 124–5, 126–7, 133, 136, 172–3, 177–84, 200, 201, 203, 204, 208, 209, 215–6, 223, 227, 231, 238, 249, 250, 256, 261, 265, 268
- ablative 136
- absolutive 179, 180–2, 184, 188–90, 194, 195, 196, 199, 201, 202, 208, 209, 215, 216, 218, 219, 220, 221–2, 222–3, 224, 227, 238, 259, 286
- accusative 53, 111, 125, 126, 169, 172, 178, 179–80, 182–3, 184, 191, 192, 199, 200, 214, 231, 261, 291
- and pronouns 48–9, 50, 53, 180–1, 182–3, 212, 295
- dative 121, 125, 136, 147, 188, 192, 193, 194, 195, 196, 199–200, 210, 218, 223, 224, 225, 226, 227, 231, 242, 293
- ergative 179, 181–4, 189, 190, 194, 195, 196, 200, 201–3, 215, 216, 218, 219, 220, 223, 224, 226–7, 238–9, 293
- genitive 129
- instrumental 218, 227
- nominative 49, 53, 67, 75, 102, 111, 124, 126–7, 134, 172–3, 178, 179–80, 183–4, 192, 193, 194, 199, 212, 214, 220, 229, 256, 294
- case marking (*see also* case) 48, 53, 125, 127, 129, 133, 147, 172–3, 177–84, 189–90, 194–5, 199–200, 201–204, 210, 214, 215, 216, 220, 221, 227, 249, 250, 261, 268, 273, 291, 292, 293, 294, 295
- causative 229–33, 234, 240, 279, 292
- morphological 230–1
- summary of properties of 232
- clause 67, 73–6, 78, 81, 83–94, 97, 99, 101–2, 105, 109, 111, 115–6, 117, 119, 123, 126–8, 147, 148–9, 172, 173–7, 183, 191, 194, 195–6, 198, 205–6, 213, 214, 218, 220, 221, 222–4, 229, 230, 238–9, 244, 245, 256, 247, 251–7, 259, 261, 272, 278, 279, 280, 292, 294
- adjunct 88, 89, 93, 94, 98, 105
- adverbial (*see* clause, adjunct)
- conditional 88
- embedded (*see also* clause, subordinate) 24, 73, 81, 84, 85–9, 92, 93, 96, 97, 99, 102, 105, 175, 183, 195, 198, 205–6, 230, 246–7, 251, 252, 255, 256, 278, 279, 292
- gerund-participial 116
- matrix 84–7, 89, 92, 93, 94, 96, 97, 98, 99, 105, 195, 198, 230, 251, 253, 278, 292
- relative (*see* relative clause)
- root 89–92, 94, 95, 99, 101, 105, 174–5, 246, 247, 273, 278, 293
- subordinate (*see also* clause, embedded) 73, 84–94, 96, 98, 99, 101, 105, 174–5, 176, 251, 273, 278, 279, 292
- subordinate in English 84–9, 91, 92, 93, 94, 96, 99, 105–6, 176, 251
- cleft 146–7, 151, 152, 160–1, 162, 163, 260, 292
- closed class (*see also* open class) 38, 58, 59, 66, 82, 117, 118, 119, 293
- comparison (of adjectives) 27–8, 58
- equative 58
- comparative 27–8, 58, 62, 257, 262
- superlative 58
- complement 21, 84–7, 88, 89, 92, 93–4, 99, 105, 109, 111–8, 119, 120–3, 134, 149, 157, 158, 162, 163, 176, 191, 195, 198, 199, 200, 225, 230, 234, 272, 275, 278, 292
- clausal 84–7, 88, 93–4, 105, 115, 116, 117, 121, 176, 191, 198, 200, 272, 278
- structure 111–7
- to adjectives 113, 116, 121
- to nouns 116–7, 120–1
- to prepositions 47, 50, 64–6, 109, 111, 112, 116, 118, 120–1, 123–6, 162, 258–9, 293
- to verbs 84–7, 88, 89, 92–3, 105, 109, 112, 113–6, 123, 126–7, 191, 195, 199, 230
- vs. adjunct 117–8
- complementary distribution 62
- complementation strategies 84–99, 272,
- complementizer 85, 86, 88, 90, 91, 94, 101–3, 117, 119, 176, 230, 246, 252, 292
- complex NP (*see* noun phrase, complex)
- complex sentence (*see* sentence, complex)
- complexity, linguistic 282–3
- concord 184
- conjoined phrases (*see* co-ordination)
- conjunction, co-ordinating 83–4, 94, 165, 169–71
- constituent 19, 141–67, 173–7, 178, 180, 185, 186–7, 191, 192, 195, 201, 202, 203–4, 205, 239, 243, 247, 249, 259–61, 263, 265, 273, 280, 284
- definition of 142, 143, 155–6
- discontinuous 202–4
- order (*see also* word order) 19, 148, 168, 172, 173–7, 178, 186–7, 191, 195, 201, 204, 205, 221, 239, 245, 250, 255, 256, 259, 261–2, 272, 273, 274, 275, 278, 284, 292
- tests for 141–66
- conventions used in linguistics 3, 15, 16, 40, 113, 154, 271
- co-ordination 83–4, 94–5, 165–6, 170, 273, 292
- copula 50, 56, 61, 75, 292
- core (*see* argument, core)
- co-reference 222, 223, 224, 252, 285

- critical period 281
- cross-reference (*see also* agreement) 45, 58, 124, 128, 173, 184–6, 188, 189, 273, 292
- dative case (*see* case)
- dative movement (*see also* applicative) 225, 226, 227, 228, 229, 234, 292
- daughter (*see also* mother) 155, 156
- definiteness 15, 38, 53–4, 55, 180
- degree modifier (*see* intensifier)
- demonstrative 15, 38, 54, 291
- demotion (*see also* promotion) 22–4, 45, 211, 212–4, 216, 217, 218, 219, 220, 221, 224, 225, 226, 227, 229, 231, 232, 233, 240, 279, 284, 292, 294
- dependency 41, 111, 147–8, 244
- dependent 108–34, 135, 136, 150, 177, 184, 272, 293
- dependent-marking (*see also* head-marking; split-marking) 123–34, 272, 273, 293
- dependent sentence (*see* clause, embedded)
- determiner 53–5, 58, 96, 110, 118–9, 125, 128, 130, 133, 134, 144, 154, 165, 180, 291, 293
as head of nominal phrase 55, 118–9, 134
possessive 54, 128
wh- 54
- detransitivization 219, 224
- direct object (*see* object)
- displacement 41, 147–8, 244, 259, 273
- distribution 33–7, 48, 55, 57, 62, 66, 68, 74, 76, 80, 82–3, 109, 119, 272
- distributional tests for word class 32–9, 74, 76, 80, 82–3
- diversity 12, 274, 283, 284
- dominate (*see also* immediately dominate) 155–6, 164
- dual (*see also* number) 17, 29–30, 31, 51, 60, 188, 201–2
- echo question 145–6, 152, 153, 161, 247, 248, 293
- ellipsis 76, 100, 161, 163, 166, 194, 198, 222–4, 238, 293
- embedding (*see* clause, embedded and subordination)
- ergative/absolutive pattern (or ergative pattern) 179, 180–4, 185, 188–90, 199, 201, 208, 209, 215, 291, 292, 293
- ergativity (*see also* case, ergative) [please delete 'ergative']
181, 189, 205, 221
and agreement (*see* agreement)
split ergative pattern 182–4, 189–90, 201
- examples, making use of 3, 4, 6, 11–19, 26, 103, 145, 153, 163
- exclamative 261
- experiencer (*see also* semantic role) 46–8, 195, 293, 295
- finiteness 48, 49, 53, 73–8, 79, 80–1, 84, 85, 86, 88–99, 105, 115, 116, 117, 168, 193, 195, 196, 245, 258, 266, 272–3, 274, 275, 276, 278, 279, 293
- focus 20, 25, 127, 146–7, 151, 154, 160, 168, 174, 175, 202, 216, 217, 221, 259–61, 263, 273, 280, 284
- fronting 20, 147, 221–2, 239–40, 243–7, 248, 249, 250, 259–60, 261, 280
- function 23, 30, 33–6, 39, 46, 50, 52, 53, 56, 62, 63, 64, 66, 70, 77, 91, 95, 101, 102, 111, 137, 220, 244, 246, 251, 257, 262, 291
- functional categories 38, 43, 54, 55, 57, 58, 66, 85, 272, 278
- gender 26, 45, 51–2, 53, 58, 66, 82, 98, 110, 111, 124, 125, 126, 129, 132, 185, 187, 189, 210, 231, 233, 276, 291
- generalization 76, 184, 206, 262, 265
- genitive case (*see* case)
- gerund 83, 96, 116
- gloss 13–19, 29, 30, 75, 95, 122, 138, 272
and grammatical information 14–6
and lexical information 14–6
- goal (*see also* semantic role) 47, 200, 226, 295
- grammar 1–4, 7, 11, 12, 61, 63, 80, 82, 83, 87, 116, 158, 167, 200, 262, 273, 282–6
descriptive 1
prescriptive 2, 3, 4, 7
- government 111, 125
- grammatical categories 15, 16, 31, 32, 41–6, 50–6, 58, 67, 78–80, 98, 103, 104, 110, 293, 294
for adjectives 58
for nouns 50–6
for prepositions 65–6
for verbs 41–6, 78–80, 104, 293
inherent 58, 110
- grammatical information (*see* gloss)
- grammatical relations 23, 47, 48–9, 52–3, 64, 111, 124, 127, 133, 172–3, 177, 190–201, 204, 205, 207, 211–33, 239–40, 244, 245, 249, 253, 257, 261, 273, 275, 284–6, 293
nonprimary 220
primary 220, 221, 222, 223, 224, 292
- grammaticality judgement 26, 100, 148, 234, 284
- head 73, 93, 108–36, 144, 145, 146, 158, 162, 163, 165, 177, 180, 184, 190, 200, 243, 251, 252, 253, 254–5, 272, 275, 276, 284, 293
- head-final (*see also* head-initial) 121, 122, 125, 134, 135, 255, 272, 292, 293

- head-initial (*see also* head-final) 120–1, 122–3, 134, 135, 272, 275, 293
- head-marking (*see also* dependent-marking; split-marking) 123–34, 136, 201, 273, 276, 277, 293, 294
- hierarchical structure 24, 87, 89, 119
- homogeneity 12, 284
- immediately dominate (*see also* dominate) 115–6, 165
- imperative 100, 191
- impersonal 216–7, 280
- independent sentence (*see also* dependent sentence) 24, 73–4, 75, 83, 84, 85, 89, 93, 96, 147, 251, 256, 257
- indicative (*see also* mood; subjunctive) 43–4, 229
- infinitival clause 81, 93, 116, 191, 195, 278
inflected 81
- infinitival marker 80
- infinitive 2, 77, 80–1, 84, 93, 100, 116, 294
- infix 214
- inflections 16, 42, 65, 68, 74, 77, 79, 80, 94, 103, 126, 127, 182, 183, 184, 186, 219, 223, 276, 277, 278, 280
- innate language faculty 5, 11, 204, 271, 274, 281–2, 283, 286
- instrument (*see also* semantic role) 46, 47, 226, 227
- intensifier 57–8, 62, 63
- inversion 91–2, 94, 100, 105, 193, 247, 292
subject/auxiliary 91–2, 94, 100, 105, 247, 292
subject/verb 193
- labelled brackets and tree diagrams 61–2, 114, 148, 149–65, 167
- language change 7–11, 26
- language isolate 181, 216, 284
- language phylum 51
- lexical classes 32, 62, 65, 66, 67, 76, 93, 105, 272, 283
- lexicon 13, 15
- locative 64, 196, 226, 244
- marked (*see also* unmarked) 20, 168, 173–5, 184, 196, 187, 191, 272
- modal (*see* auxiliary)
- modification 21, 33, 34, 35, 37, 57, 60, 110, 129, 139, 145, 251, 254, 255, 256
- mood (*see also* indicative; subjunctive) 43–4, 78, 80, 98, 229
- morphological tests for word class 35–7
- morphology 15–16, 35, 37, 42, 43, 52, 53, 54, 59, 61, 67, 68, 74, 75, 79, 82, 95, 124, 132, 134, 180, 182, 189, 190, 196, 198, 199, 213, 225, 226, 231, 232, 240, 275, 276, 285, 292, 294
- morphosyntax 15–16, 30, 35, 36, 39, 59, 66, 67, 274, 282, 283
- mother (*see also* daughter) 155
- negation 3, 8, 9, 13, 38, 78, 79, 92, 98, 99, 100, 103–4, 266–8, 275–6, 279, 284
double 3
- node 154, 155–6, 164
- nominalization 95–6, 99, 273, 294
- nominalizer 96, 104
- nominative (*see* case)
- nominative/accusative pattern (or accusative pattern) 178, 179–80, 184, 185–8, 189–90, 198, 199, 208, 209, 214, 215, 261, 285–6, 294
- non-finite verb (*see also* finiteness; participle; infinitive) 76, 77, 78, 79, 80–3, 84, 85, 86, 88, 89, 91, 93, 94, 99, 105, 115, 116, 117, 258, 275, 276, 277, 278, 279, 294
- non-singular (*see also* number) 17
- non-standard usage (*see also* standard usage) 3, 4, 9, 10, 11, 28, 139–40, 252–3, 262
- noun 46–56, 110, 117, 128–30
as head of NP 108, 117, 120, 122, 123, 128–30, 251, 252, 253, 254–5
count 10, 51
mass 10, 51
proper 54, 214
- noun class (*see* gender)
- noun phrase (NP) 10, 36, 37, 45, 46–56, 63, 86, 95, 111, 118–9, 123, 128–9, 132, 133, 134, 135, 172–204, 211–33, 252, 255–9, 277, 285
as core argument (*see* argument)
as oblique argument (*see* argument)
as object (*see* object)
as subject (*see* subject)
complex 117, 255–6
semantic roles of (*see* semantic roles)
- number (*see also* dual; non-singular; plural; singular; trial) 16–18, 26, 42, 44, 45, 48, 49, 50–1, 58, 59, 66, 67, 74, 75, 77, 79, 80, 81, 96, 98, 103, 111, 124, 125, 126, 127, 128, 129, 130, 133, 185, 186, 188, 189, 192, 201, 210, 276, 277, 291, 292, 293, 294
- object (includes direct object) 13, 20–1, 22, 23, 29, 38, 41, 45, 46, 47, 49–50, 51, 53, 63, 64, 65, 66, 67, 75, 84, 90, 95, 97, 102, 109, 111, 112, 113, 114, 115, 118, 120, 123, 124, 125, 126, 127, 128, 131, 132, 133, 134, 135, 148, 149, 150, 157, 158, 162, 164, 166, 169, 172–90, 198, 199–201, 203, 205, 212–15, 217, 218, 219, 220, 222, 224, 225, 226, 227, 228, 231, 232, 233, 236, 239, 244, 247, 248, 249, 250, 252, 253, 254, 255, 256, 257, 258, 259, 260, 275, 276, 277, 279, 280, 291, 292, 293, 294, 295

- double 200
- indirect 115, 195, 200, 205, 226, 231, 249, 255, 258, 293
- prepositional (or postpositional) 47, 50, 54, 63, 66, 125, 131, 228, 248, 252, 253, 257, 258, 259, 276, 277, 280, 293
- oblique (*see* argument, oblique)
- open class (*see also* closed class) 38, 58, 59, 66, 82
- paradigm, inflectional 16, 126, 276, 278
- participial (*see* participle)
- participle 45, 78, 80, 82–3, 116, 203, 212, 213, 217, 228, 230, 294
- partitive 218
- passive (*see also* antipassive) 22–4, 44–5, 83, 113, 192, 211–6, 217, 218, 219, 220, 225, 227–8, 229, 231, 232, 233, 236–7, 279, 285, 294
- in ergative languages 215–6
- morphological 214, 231, 279
- summary of properties of 22–3, 213
- past participle (*see* participle)
- patient (*see also* semantic role) 46, 47, 48, 175, 212, 309
- person 16–18, 26, 29, 42, 45, 48, 49, 59, 66, 74, 75, 79, 80, 81, 95, 96, 98, 103, 110, 124, 125, 126, 127–8, 129, 130, 131, 133, 136, 178, 183, 185, 186, 188–9, 190, 191, 192, 199, 201, 208, 219, 236, 266, 276–8, 279, 291, 292
- exclusive 17–18, 29
- inclusive 17–18, 29
- pivot 222–4, 238–9, 273
- plural (*see also* number) 10–11, 16, 17, 18, 29, 33, 34, 37, 48, 50–1, 52, 81, 119, 134, 185, 186, 187, 189, 193, 210, 212, 276, 277, 278, 279
- possessed N (*see also* possessor NP) 52, 66, 123, 128–9, 131, 136, 168, 253
- possession 52, 70, 128–9, 136, 168
- alienable 52, 168
- inalienable 52, 168
- possessive marker 25, 52, 63, 95, 96, 121, 124, 128–9, 131, 133, 136, 142, 143, 168, 253
- possessor NP (*see also* possessed N) 123–4, 128–9, 131, 136, 252–3, 257, 258, 275
- postposition 65, 121, 122, 125, 131, 176, 248, 291, 293, 294
- predicate 39, 50, 56, 73–4, 75, 88, 97, 99, 109, 149, 160–1, 173, 175, 176, 229, 230, 278, 284, 292, 294
- prefix (*see also* affix; infix; suffix) 15, 17, 18, 29, 42, 46, 51–2, 59, 75, 95, 98, 125, 127, 131, 138, 168, 186–7, 189, 205, 216, 236, 254, 264–5
- preposition 2, 38, 47, 50, 63–6, 80, 108, 109, 111, 112, 113, 115, 116, 118, 120, 124–6, 147, 151, 154, 157–64, 176, 196, 205, 226, 257, 258, 259, 265, 275, 276, 277, 280, 291, 293, 294
- and inflection 66, 126, 276, 277, 280
- modifiers for 64–5, 158
- pro-drop 12, 277
- proform 160, 162, 166, 167, 294
- promotion (*see also* demotion) 6–7, 22–4, 45, 192, 199, 212–4, 216, 217, 218, 220, 221–2, 223, 224, 225, 226–8, 229, 232, 233, 240, 279, 284, 292, 294
- pronominal 49, 66, 126, 128, 183, 212, 276, 277
- affix 17–18, 46, 59, 66, 75, 79, 95, 98, 99, 126, 127, 128, 130, 131, 173, 186, 188, 201, 204, 236, 265, 277, 292
- bound 17, 75, 98, 99, 127, 173, 186, 188, 204, 236
- free 17, 59, 66, 127, 173, 186
- pronoun 9, 10, 11, 12, 17, 18, 26, 27, 30, 38, 46, 48, 49, 50, 53, 54, 59, 69, 75, 79, 91, 95, 98, 106, 124, 126, 127, 132, 133, 134, 143, 148, 160, 162, 173, 178, 180, 181, 182–3, 184, 185, 186, 188, 189–90, 191, 192, 193, 194, 199, 201, 204, 212, 224, 236, 257, 258, 276, 277, 280, 294, 295
- as determiner 54–5
- resumptive 257, 258, 280
- quantifier 54, 139–40, 210
- floating 139–40, 210
- reading 142, 144, 164, 165, 218, 223, 224
- recipient (*see also* semantic role) 40, 41, 46–7, 121, 188, 200, 201, 205, 249
- reciprocal 191
- recursion 24, 86, 150
- reflexive 26–7, 191,
- relative clause 93, 251–9, 262, 265, 273, 280, 284, 285, 294
- non-restrictive 251
- restrictive 251
- relativized position 252–9
- relativization strategies 256, 257–8
- responsive 266–8
- scrambling 261
- semantic roles (*see also* agent; experiencer; goal; instrument; patient; recipient; stimulus; theme) 46–7, 48, 177, 195, 196–8, 200, 212, 295
- sentence, complex 73, 83, 84, 86, 88, 89, 91, 94, 99, 205, 229, 230, 251, 255, 284, 292
- sentence, simple 73–4, 75, 76, 78, 83, 84, 91, 205, 229, 256, 284
- sentence fragment 143–5, 148, 152, 153, 160, 162, 163
- sentential subject (*see* subject, clausal)
- serialization (*see* verb, serial)
- singular (*see also* number) 10–11, 16–18, 27, 37, 42, 44, 45, 48, 50–1, 77, 79, 80, 81, 94, 100, 119, 134, 136, 180–1, 185, 236, 266, 276–8, 279, 292

- specifier 118
- standard usage (*see also* non-standard usage) 3, 4, 9–10, 26–8, 44, 48, 62, 100, 139, 252–3, 257
- statistical patterns in constituent order 175–6
- stimulus (*see also* semantic role) 46–7
- structure 14, 24–6, 44, 87, 89, 109, 116, 119, 141–153, 157, 165, 166–7, 203–4, 244
- complement 116
- constituent 141–67, 203–4, 243, 284
- structure dependency 26
- subject 12, 13, 22–3, 45, 46, 47–9, 53, 59, 63, 67, 69, 73, 75, 81, 85, 86, 87, 88, 91, 93, 95, 97, 100, 109, 111, 114, 116, 124, 126, 128, 132, 133, 134, 148, 160, 170, 172–99, 211–17, 220–2, 228, 229–32, 236, 238, 247, 250, 252, 257, 258, 260, 265, 273, 274, 275, 276, 279, 292, 293, 294, 295
- clausal 86, 87, 89, 279
- dummy 217
- cross-linguistic properties of 190–9
- properties of in English 48–9, 212
- quirky 193–4
- sentential (*see* clausal)
- subjunctive (*see also* mood; indicative) 43–4, 93
- subordination 73, 84–94, 105, 174–5, 176, 251, 273, 278, 292, 294
- suffix (*see also* affix; infix; prefix) 15–16, 23–4, 27, 29, 35, 37, 42, 43, 45, 46, 50, 54, 58, 60, 62, 74, 79, 80, 81, 82, 95, 96, 100, 127, 129, 130, 131, 133, 142, 173, 181, 183, 186, 188, 201, 203, 205, 216, 218, 223, 224, 226, 227, 231, 237, 291
- syntactic tests for word class 32–9, 55, 57–8, 64–5, 74, 81, 83
- tag question 9–10, 91–2, 94, 105, 142
- temporal (property of PPs and NPs) 64, 244
- tense 15–16, 41–3, 44, 45, 48, 49, 60, 67, 74–5, 77, 78–9, 80, 81, 82, 83, 96, 98, 100, 103, 133, 154–5, 173, 183, 201, 266, 276, 277–8, 280, 292, 293
- thematic role (*see* semantic role)
- theme (*see also* semantic role) 46–7, 50, 176, 196–8, 200, 201, 212, 249, 295
- topic 170–1, 174, 175, 187, 196–8, 202
- topicalization 187
- tree diagram 118, 119, 148–60, 163–5, 167
- trial (*see also* number) 17, 29, 51
- typology 61, 123, 124, 132–4, 230, 254, 272, 281
- ungrammaticality 4–5, 7, 19, 24, 25, 28, 37, 40, 41, 60, 76, 83, 89, 91, 106, 113, 116, 138, 140, 145, 146, 148, 153, 157, 161, 162, 163, 165, 166, 167, 196, 203, 206, 207, 221, 222, 227, 228, 236, 244, 249, 273, 277
- universal 5, 12, 38, 58, 67, 90, 94, 198–9, 228, 283, 284
- Universal Grammar 283
- unmarked (*see also* marked) 20, 125, 126, 168, 174–5, 184, 186, 220, 223, 263, 268, 272
- valency 44–5, 213, 219, 225, 229, 233, 273, 294
- valency-changing processes 44–5, 211–20, 229–33
- verb
- and agreement (*see* agreement)
- ditransitive 40, 41, 114, 187, 200, 293
- finite (*see* finiteness)
- intransitive 39–40, 41, 44, 112, 113, 114, 157, 158, 172, 177, 178, 179, 180, 181, 186, 188, 190, 191, 194, 212, 213, 216, 218, 219, 220, 221, 222, 224, 229, 231, 232, 233, 292, 294
- lexical (*see also* main) 76, 105, 186, 230, 275, 278, 279
- main (*see also* lexical) 9, 42, 43, 45, 76, 78, 79, 80, 84, 88–9, 92, 97, 100, 105, 126, 168, 212, 214, 232, 251, 274
- phrasal 65, 157–63, 164
- prepositional 115, 157–63
- serial 96–9, 106, 137–9, 230
- transitive 37, 41, 44, 49, 97, 109, 111, 112, 113, 114, 118, 126, 138, 149, 157, 164, 176, 177, 203, 212, 216, 222, 231, 244, 295
- verb-initial order 89, 90, 120–1, 175, 207, 221, 266, 274–5
- voice (*see also* active; passive) 78
- wh*-construction 243–62, 280
- wh*-in-situ 247–8, 249, 250, 273
- wh*-movement 243–7, 248, 250, 261, 262
- wh*-question 243–51, 252, 253, 262–3, 290
- multiple 249–51
- word class
- morphological criteria for identification of 35–6
- syntactic criteria for identification of 35–6
- word order (*see also* constituent order) 4, 6, 12, 13, 14, 19–21, 23, 53, 55–6, 65, 89–90, 91, 92, 94, 126–7, 129, 132, 139, 47, 173–7, 201–4, 205, 207, 274, 284
- writing systems 18–19, 271
- yes/no question 8, 91, 193, 266–8, 294