

Tetun Dili
A grammar of an
East Timorese language

Also in Pacific Linguistics

Catharina van Klinken, 1999. *A grammar of the Fehan dialect of Tetun: An Austronesian language of West Timor.*

Charles E. Grimes (ed.), 2000. *Spices from the East: Papers in languages of Eastern Indonesia.*

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

Catriona Hyslop, 2001. *The Lolovoli dialect of the North-East Ambae language, Vanuatu.*

Fay Wouk and Malcolm Ross (eds), 2001. *The history and typology of western Austronesian voice systems.*

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, the Philippines, Indonesia, East Timor, southeast and south Asia, and Australia.

Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at the Australian National University. The Editorial Board of Pacific Linguistics is made up of the academic staff of the school's Department of Linguistics. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise, who are usually not members of the editorial board.

The 470 publications before Publication 501 were numbered in four series, A, B, C and D, but the series were merged from the beginning of the year 2000 because the rationale for them had gradually disappeared.

FOUNDING EDITOR: Stephen A. Wurm

EDITORIAL BOARD: Malcolm D. Ross, John Bowden and Darrell T. Tryon
(Managing Editors),
I Wayan Arka, Andrew Pawley, Paul Sidwell

**Shorter Grammar
Pacific Linguistics 528**

Tetun Dili
A grammar of an
East Timorese language

Catharina Williams-van Klinken,
John Hajek
and
Rachel Nordlinger



Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University

Published by Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University
Canberra ACT 0200
Australia

Copyright © The authors
First published 2002

National Library of Australia Cataloguing-in-Publication entry:

Williams-van Klinken, Catharina Lumien.
Tetun Dili: A grammar of an East Timorese language

Bibliography
Includes index
ISBN 0 85883 509 6

1. Tetun language - Indonesia - Timor Timur - Grammar.
2. Tetun language - Morphology. 3. Tetun language - Phonology.
I. Nordlinger, Rachel. II. Hajek, John. III. Australian
National University. Research School of Pacific and Asian
Studies. Pacific Linguistics. IV. Title. (Pacific Linguistics 520).

499.225

Cover design by Emily Brissenden
Printed and bound by Union Offset Printers, Fyshwick, Canberra

Contents

1.	Introduction	1
1.1	Language overview.....	1
1.2	Variation and register	3
1.3	Previous research on Tetun.....	4
1.4	Data and approach for this study	5
1.5	Language structure.....	5
1.6	Standardisation	6
1.7	Orthography.....	6
2.	Phonology.....	8
2.1	Consonants.....	8
2.1.1	Introduction	8
2.1.2	Stops	9
2.1.3	Fricatives	9
2.1.4	Other consonants	10
2.2	Vowels	11
2.2.1	The basic vowels	11
2.2.2	Diphthongs	12
2.2.3	Nasalisation	12
2.2.4	Reduction of compounds and reduplications	12
2.3	Stress and vowel length	12
2.4	Phonotactics	13
3.	Word classes	14
3.1	Introduction.....	14
3.2	Nominal and verbal classes	14
3.3	Other word classes.....	15

4.	Morphology.....	17
4.1	Overview.....	17
4.2	Affixation and zero derivation.....	17
4.2.1	Derivation from intransitive to causative verb	17
4.2.2	Derivation from noun to verb	18
4.2.3	Derivation from verb or adjective to abstract noun	19
4.2.4	Detransitivising derivation	19
4.2.5	Derivation of actor nouns and adjectives	20
	4.2.5.1 <i>-dór</i>	20
	4.2.5.2 <i>-tén</i>	20
	4.2.5.3 <i>nain</i>	21
4.3	Reduplication.....	21
4.3.1	Iteration	21
4.3.2	Collectivity	21
4.3.3	Deriving adverbs from adjectives	22
4.4	Compounds.....	22
4.4.1	Introduction	22
4.4.2	Nominal compounds	23
4.4.3	Other	24
5.	Noun phrases	25
5.1	Basic noun phrase structure	25
5.2	Pronouns	26
5.3	Determiners.....	27
5.3.1	Introduction	27
5.3.2	Demonstratives	27
5.3.3	Interrogatives	29
5.3.4	<i>ida</i> 'one, a'	30
5.3.5	<i>sira</i> 'PLURAL' and Portuguese plurals	30
5.3.6	Quantifiers	31
5.3.7	Other	32
5.4	Possessive and associative constructions.....	33
5.4.1	Overview	33
5.4.2	Order of possessor and possessum	34
5.4.3	Combination of two 'possessors'	35
5.4.4	Genitive head-marking	36
5.5	Noun phrase as modifier.....	36
5.6	Adjective phrase	36

5.7	Numeral phrase.....	37
5.7.1	Numerals	37
5.7.2	Numeral classifiers	39
5.8	Complements of nouns	40
6.	Prepositions.....	41
6.1	Introduction.....	41
6.2	<i>iha</i> ‘LOC’	43
6.3	<i>ba, mai</i> ‘to (location, person), for (person)’	44
6.4	<i>hosi</i> ‘from’, <i>tó</i> ‘until’	45
6.5	<i>ho, hodi, lori</i> ‘with’	45
6.6	Temporal prepositions	45
6.6.1	Introduction	45
6.6.2	‘before’	46
6.6.3	‘during, for’	46
6.6.4	‘after’	46
6.6.5	‘from, since’	47
6.6.6	‘until’	47
6.7	Other prepositions.....	48
7.	Basic clauses.....	51
7.1	Verbal clauses.....	51
7.1.1	Basic clause order	51
7.1.2	Subject and object	52
7.1.3	Object fronting	53
7.1.4	Subject–object–verb (object incorporation)	55
7.1.5	Existential and possessive	55
7.1.6	Other verb–subject clauses	56
7.1.7	Verbs taking adjectival complements	57
7.2	‘Body-good’ expressions	57
7.3	Other non-verbal clauses	60
7.4	Reciprocal	61
7.5	Reflexives	63
7.6	Interrogatives	64
7.6.1	Information interrogatives and their answers	64
7.6.2	Polar and alternative questions and their answers	67
7.7	Commands and invitations	69
7.8	Focus (<i>mak</i>)	69
7.9	Peripheral constituents.....	71

8.	Clausal modification	73
8.1	Introduction.....	73
8.2	Tense–aspect.....	73
8.2.1	Introduction	74
8.2.2	Focus on the completed event	74
8.2.2.1	<i>tiha</i> ‘PERFECTIVE’	75
8.2.2.2	<i>ona</i> ‘ANTERIOR’	75
8.2.2.3	<i>tiha ona</i> ‘PERFECT’	76
8.2.2.4	<i>hotu</i> ‘be finished’	77
8.2.2.5	<i>foin</i> ‘only just, very recently’	77
8.2.2.6	<i>lai</i> ‘FIRST, before doing something else’	78
8.2.2.7	<i>kedas</i> ‘IMMED, in advance, immediately’	79
8.2.3	Focus on the middle: continuity	79
8.2.3.1	<i>hela</i> ‘CONTINUOUS’	79
8.2.3.2	<i>daudauk</i> ‘PROGRESSIVE’	80
8.2.3.3	<i>sei</i> ‘still’, <i>seidauk</i> ‘not yet’	81
8.2.3.4	<i>nafatin</i> ‘continue’	82
8.2.4	Future	82
8.2.4.1	<i>sei</i> ‘FUTURE’	82
8.2.4.2	<i>atu</i> ‘IRREALIS’	83
8.3	Deontic modality	84
8.3.1	Introduction	84
8.3.2	<i>bele</i> ‘can, may’	84
8.3.3	<i>lalika(n)</i> ‘need not’	86
8.3.4	<i>keta</i> ‘PROHIBITIVE’	86
8.3.5	<i>tenki</i> ‘must’	87
8.4	Negation.....	87
8.5	Intensifiers, comparative, superlative	89
8.6	Manner modifiers.....	91
9.	Serial verbs and causative constructions	93
9.1	Introduction.....	93
9.2	Motion–action serialisation	94
9.3	Motion–direction serialisation	94
9.4	Instrumental serialisation.....	95
9.5	Other shared-subject	97

9.6	Causation	97
9.6.1	Introduction	97
9.6.2	<i>ha-</i> prefixation	97
9.6.3	Constructions with <i>halo</i> ‘make, do’	98
9.6.4	Causation using <i>fó</i> ‘give’	99
9.6.5	Resultative serial verbs	100
10.	Complex sentences	102
10.1	Introduction.....	102
10.2	Topics	102
10.3	Parataxis.....	103
10.4	Phrasal and clausal coordination	103
10.4.1	Introduction	103
10.4.2	Conjunction	104
10.4.3	Disjunction	105
10.5	Complement clauses	106
10.5.1	Introduction	106
10.5.2	Sentential complements	106
10.5.3	Reduced complements	108
10.6	Adverbial clauses.....	110
10.6.1	Introduction	110
10.6.2	Temporal	110
10.6.3	Reason	111
10.6.4	Purpose	112
10.6.5	Condition	113
10.6.6	Concession	113
10.7	Other means of connecting clauses.....	114
10.8	Relative clauses	115
Texts.....	118
Text 1:	History behind setting up an NGO	118
Text 2:	Why people create disturbances	120
Text 3:	Our escape from the militia	122
Text 4:	A press report.....	125
Text 5:	A gospel portion	126
References.....	127
Index	130

List of tables

Table 2.1: Consonants	8
Table 5.1: Personal pronouns	26
Table 5.2 Numerals.....	38
Table 5.3: Numeral classifiers	40
Table 6.1: Location nouns	42
Table 8.1: Intensifiers	89

Acknowledgments

Catharina would like to thank the many people who welcomed me and assisted me in Dili. Special thanks goes to the family of Senhor Cornelio Luz Ferreira and Senhora Maria Paixão de Jesus Costa Ferreira, along with Senhora Angelina dos Reis Magno, who graciously accepted me into their home and their lives.

Those who taught me, encouraged me, fed me, or collected or transcribed texts are too numerous to mention, but include Senhora Ana Maria de Fatima Costa Soares and Senhor Alfonso de Jesus, Vic. Samuel Marçal, Senhora Natercia dos Reis Ferreira, Senhor Natalino da Conceição da Silva and Senhora Carla, Senhor Serafino Dias da Costa, Senhora Ofelia Jacinta Ricardinha Salsinha and Senhor Ismael da Costa Babo, Senhor Bernardino Menezes, Menina Rita Angelina Mota, Pastor Carlos Marçal and Senhora Rita da Costa Marçal, Menina Maria Sabina Tilman Carceres, Mestri João Nunes, Mestri José Maria and his fellow teachers at Dare, Senhor Edio da Costa and Senhora Candida Lopez Ribeiro, Senhor Fausto de Sousa, Menina Domingas de Araújo, Senhor Firmino Ferreira and Senhora Mónica de Araújo Fernandes Mendonça, Senhora Julieta dos Reis, Senhor João Gonçalves Silveira and Senhora Palmira Mendonça, Senhor Patrocínio dos Reis Magno and Senhora Rosel de Jesus do Rego Magno, and Senhor Hernânio Vicente. Thank you all!

John and Rachel would like to thank Alexandre Vital Tilman, Gil Santos and Emanuel Braz for their kind help in Australia over the years. Special acknowledgment must go to A.V. Tilman ('Alex' to us) for his patient assistance in the early days as language helper in our Field Methods course where we first grappled with the grammatical description of Tetun Dili. John would also like to thank the many people in East Timor who generously shared their time and language—sometimes unwittingly—with me.

The authors also appreciate financial assistance from an ARC small grant made available at the University of Melbourne

Abbreviations

Abbr.	Meaning	Abbr.	Meaning
1S	1 st person singular (<i>hau</i>)	INTR	intransitiviser (e.g. <i>hak-</i>)
1PE	1 st person plural exclusive (<i>ami</i>)	IRR	irrealis (<i>atu</i>)
1PI	1 st person plural inclusive (<i>ita</i>)	JUST	only just, very recently (<i>foin</i>)
2P, 3P	2 nd and 3 rd person plural (<i>imi, sira</i>)	lit.	literally
2S, 3S	2 nd and 3 rd person singular (<i>o, nia</i>)	LOC	general locative (<i>iha</i>)
2S.HON	2 nd person singular respectful (<i>ita</i>)	MASC	masculine
Adj	adjective	N	noun
AGENT.N	agentive nominaliser (<i>-dór</i>)	pers.	personal communication
AGENT.A	agentive adjectivaliser (<i>-tén</i>)	comm.	
ANT	anterior (<i>ona</i>)	PL	plural (<i>sira, -s</i>)
CLS:	classifier for the following category	POS	possessive (<i>nia</i>)
CONT	continuous (<i>hela</i>)	PRF	perfective (<i>tiha</i>)
EXIST	exist, there is/are, be present (<i>iha</i>)	PROH	prohibitive (<i>keta</i>)
FEM	feminine	PROG	progressive (<i>daudauk</i>)
FIRST	first, before doing other things (<i>lai</i>)	RDP	reduplication
FOC	focus (<i>mak</i>)	RECIP	reciprocal (<i>malu</i>)
FUT	future (<i>sei</i>)	REL	relative clause marker (e.g. <i>nebé</i>)
HES	hesitation	Vi	intransitive verb
IMMED	immediately, in advance (<i>kedas</i>)	Vt	transitive verb

Conventions for examples

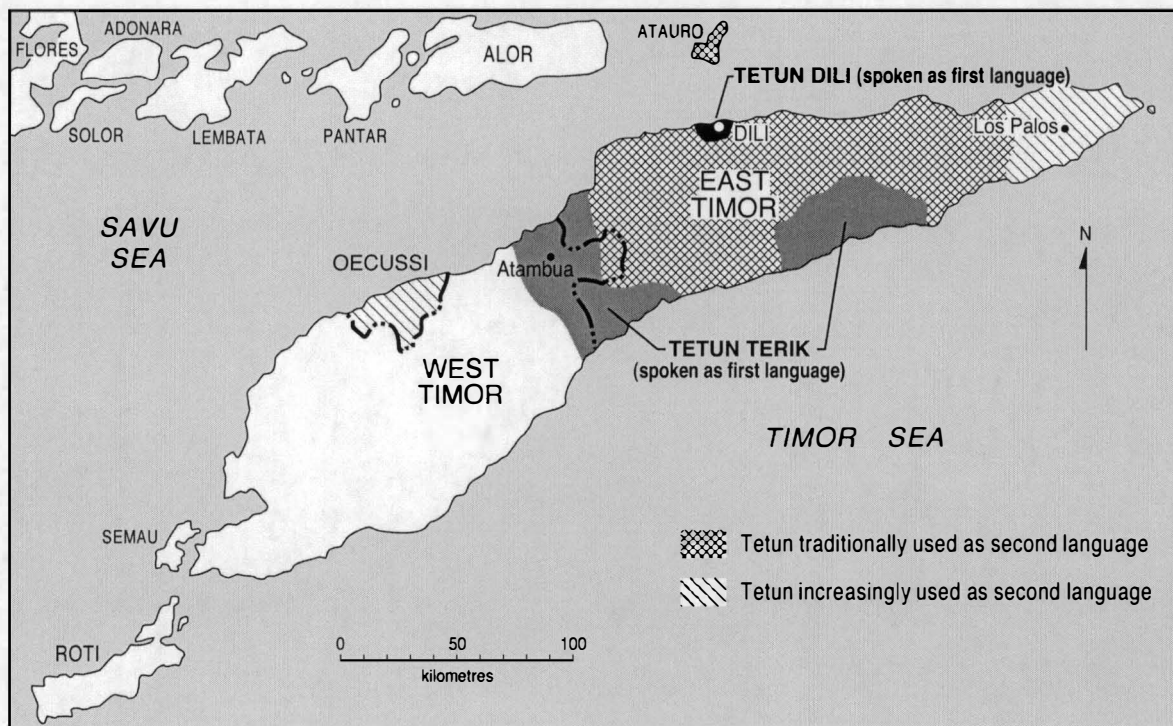
In Tetun examples:

- * marks the following example as ungrammatical
- ... indicates ellipsis
- á the acute accent marks stress, where it is not penultimate
- N—— substitutes for names to protect anonymity

The spelling system used is explained in section 1.7.

In English glosses, full stops separate words which gloss a single Tetun word (e.g. *sés* ‘move.aside’).

In translations, gender, tense and aspect (which are often not inferable from the Tetun) are given as per the original context. Brackets surround comments, as well as the translation or paraphrase of those parts of the example that are omitted by ellipsis.



Map 1: The distribution of Tetun in Timor (based on Thomaz 1981:56)

1. Introduction

1.1 Language overview

Tetun (alternatively spelled Tetum) is the best known indigenous language of East Timor, located between Indonesia and Australia. It is one of at least twelve Austronesian languages spoken in this newly independent nation, in addition to a smaller number of Papuan languages.

Tetun is a member of the Central Malayo-Polynesian grouping within Austronesian. The most closely related languages include Idate, Kemak, Lakalei, Mambae, and Tukudede.

There are at least two major varieties of Tetun: (1) Tetun Dili, spoken as a first language in the capital Dili and the focus specifically of this grammar; and (2) Tetun Terik, a much more conservative rural variety spoken as a first language in two distinct areas some distance from Dili. The first is the Soibada–Viqueque district on the south coast and the second is a much larger area that extends from Suai on the south coast across and along the border with the Indonesian province of West Timor. Available estimates (Grimes 2000) suggest that Tetun Terik has more than 300,000 native speakers, with the large majority of these being resident in West Timor. The map on the preceding page (based on Thomaz (1981)) shows the distribution of Tetun in East Timor.

Although Tetun Dili and Tetun Terik are often assumed to be variants of the one language, this view is not uncontroversial. It has been noted (Grimes et al. 1997:52) that the two varieties are not easily mutually comprehensible, leading to the suggestion that they should be treated as two separate languages. This high degree of separation results primarily from two different factors: (1) a significant number of grammatical differences; but especially (2) substantial variation in the number and source of loanwords.

The dialects of Tetun Terik spoken in West Timor have long shown greater Malay/Indonesian influence with relatively little historical contact with Portuguese, the colonial language of East Timor from the 1500s until 1975. At the other extreme, Tetun Dili has had long and intensive contact with Portuguese since the establishment of Dili in the late 1700s. In the period before the departure of the Portuguese in 1975, the local population of Dili was reported to have been largely bilingual in Tetun and Portuguese for some time. Such close contact has led to a massive influx of Portuguese loan words, especially of a technical or abstract nature, into Tetun Dili. Effects on other aspects of Tetun Dili are also evident. This process of so-called Lusification was partly broken during the Indonesian administration of East Timor from 1975 to 1999, but is now rapidly being restored as a result of local policy

initiatives in favour of Portuguese in East Timor, including its progressive reintroduction into schools as a medium of instruction, and its increasing use in the media.

Portuguese influence notwithstanding, Tetun Dili does show some Malay influence, which is divisible into two historical layers. The first results from centuries of contact through the longstanding use of Malay as a trade language in the Timor area; this was broken in the mid-nineteenth century by Portuguese policy (Fox 1997; Thomaz 1981). Intensive contact with Malay (in the form of Indonesian) was re-established from 1975 until 1999 during the period of Indonesian control of East Timor. Effects of both periods are most evident on the lexicon. Old borrowings are considered to be fully nativised (e.g. *buá* 'betel' from Malay *buah* 'fruit' and *durbasa* 'translator'). More recent borrowing is especially evident in higher-level vocabulary (e.g. *otonomi* 'autonomy', *listrik* 'electricity', *kepala desa* 'village head'). There is now an increasing tendency, especially in writing, to avoid more recent Malay/Indonesian borrowings, by replacing them with Portuguese loans (e.g. *autonomia* 'autonomy', *eletrisidadi* 'electricity' and *xefi suku* 'village head').

Still to be determined is the extent to which Mambae and Galolen, two Austronesian languages spoken in the vicinity of Dili, have influenced Tetun Dili. The town of Dili is situated in a very small coastal pocket implanted within an otherwise totally Mambae speaking area. Galolen, spoken a relatively short distance to the east of Dili, was used by Catholic missionaries in the nineteenth century to spread the faith along East Timor's northern coast.

The relatively heavy contact phenomena exhibited by Tetun Dili, and its divergence from vernacular Tetun Terik, lead Grimes et al. (1997:52) to label Tetun Dili a 'creole', using the term in a loose sense that indicates a language that has diverged significantly from its source (Mühlhäusler 1986:6; Barbara Dix-Grimes, pers. comm.). Tetun Dili does not, however, exhibit all the features of classic creoles as strictly defined (for instance by Thomason (1997)). In particular, the vocabulary is predominantly indigenous, whereas in classic creoles much of the vocabulary derives from the superstratum language (in this case Portuguese). It is also not known whether, like classic creoles, Tetun Dili passed through a pidgin stage in its early history. In fact much detail regarding the early history and genesis of Tetun Dili is not available, including the sociolinguistic situation in Dili following its establishment in the late 1700s.

Although there is great local awareness of the difference between Tetun Dili and Tetun Terik, including problems of intercomprehension already referred to, local naming practice tends to be inclusive. Whilst the terms 'Tetun Dili' and 'Tetun Terik' are widely known and used, it is also common practice to refer to both or either as 'Tetun'. An older term for Tetun Dili, but no longer much used, is Tetun Prasa (lit. 'marketplace Tetun'). The alternative spelling of Tetun as 'Tetum' reflects Portuguese convention with respect to marking final nasalisation, and the word is never in fact pronounced with final [m] in Tetun itself.

Tetun Dili is widely used as a lingua franca throughout most of East Timor, and has now been designated as the 'national language' of East Timor. Current estimates suggest that some 60–75% of the population is able to speak at least one form of Tetun. For the overwhelming majority of these speakers, this is likely to be Tetun Dili. Only some 50,000, all native to Dili, are first language speakers of Tetun Dili, but the number of second-language speakers would be in the hundreds of thousands. Thomaz (1981) reported that only in two

areas of East Timor was Tetun not known: (1) the geographically isolated Oe-Cussé enclave and (2) the Fataluku-speaking extreme eastern tip of the island. However, visits to both areas since 1999—by one of the authors and others—show a strong and increasing penetration of Tetun Dili in both regions, especially among young adults, mainly as a result of periods of residence in Dili for study or work.

Available census figures confirm that 50% of the local East Timorese population is today fluent in Indonesian (Hajek in press); this is much higher than the 5–20% reported for Portuguese but less than recent estimates of 60–75% for Tetun. Indonesian is however still widely used in the written press, and to a lesser extent on radio and television. It is currently being rapidly phased out of the education system as a medium of instruction, and its future influence on Tetun can be expected to decline over time, especially as Tetun and Portuguese increasingly take over domains in which Indonesian has dominated since 1975.

In addition to the renewed use of Portuguese in schools and media, local policymakers in 2000 reiterated an earlier policy decision to designate Portuguese as East Timor's new 'official language'. The reintroduction of Portuguese in East Timor has however been matched by the long-term rise in status of Tetun since the 1970s. In 1981 Tetun replaced Portuguese as the liturgical language of the Catholic Church. Later it was reportedly introduced as the medium of instruction in the first two years of schooling in some Catholic primary schools in the Dili diocese. It is increasingly used in the media and schools, and is also used as the default language of verbal communication across a wide range of domains amongst the East Timorese. English is widely spoken by foreign aid workers, and civilian and military personnel working under United Nations supervision since the end of 1999. It is also spoken by a small number of local East Timorese.

More detailed accounts of the sociolinguistic situation in East Timor—before 1975, during the Indonesian period and after 1999—are provided by Thomaz (1981), Hull (1994), and Hajek (2000; in press).

1.2 Variation and register

A significant issue in any description of Tetun Dili is the unusually high levels of interspeaker variation—seemingly across all areas of the grammar, including the lexicon and phonology. It is possible as a result for linguists to report quite different findings, a fact which contributes to differences between some of the conclusions we draw in this study and those presented by other researchers, such as Hull (1996b).

Variation occurs along a very large continuum and also interacts with social and register differences. There are two distinct registers used in formal interactions, although some mixing between the two does occur. One register predominates in the media and in formal writing; it is characterised by particularly high levels of Portuguese loans. The other register predominates in religious writing and sermons, and is characterised by significant Tetun Terik influence. This 'church Tetun' is discussed below. These two formal registers are further discussed in Williams-van Klinken (in preparation). They are illustrated by texts 4 and 5 in the appendix, while texts 1–3 illustrate spoken Tetun.

Tetun Dili speakers do not all share the same degree of access to or knowledge of Portuguese from which to draw. Portuguese influence is greatest amongst speakers who have

received Portuguese education (in East Timor or in Portugal) or who are from families where knowledge of Portuguese was maintained at home after 1975. Knowledge of Portuguese and evidence of its influence are therefore greatest amongst members of East Timor's traditional educated elite and political leadership. Apart from the social aspect, Portuguese influence also increases with the greater formal and/or technical nature of a particular linguistic context.

The precise extent of Portuguese influence is therefore difficult to establish, especially since even the same speaker can show substantial differences according to context, leading to a high number of Portuguese–Tetun doublets. However, Portuguese loans fall into four general categories:

- (1) loans which everyone uses and accepts as Tetun, e.g. *kama* 'bed', *kafé* 'coffee';
- (2) loans which tend to be used by more Portuguese-influenced speakers but which all speakers understand, e.g. *sín* 'yes', *komu* 'as, because';
- (3) loans used mainly by educated Portuguese–Tetun bilinguals typically in more formal and intellectual contexts, but which are not necessarily widely understood, e.g. *sobre* 'about';
- (4) loans used almost exclusively by educated Portuguese–Tetun bilinguals, but which are either not understood or not accepted as Tetun by others, e.g. *alein de* 'besides'.

The particularly high use of Portuguese loanwords in Tetun television broadcasting (up to 50% in some short televised reports) is beginning to affect this scale by spreading knowledge of loanwords in categories (3) and (4) amongst the wider community.

The formal register which we term 'church Tetun' is used primarily in religious writings, sermons, and liturgy. Some Tetun speakers also use it as a model, especially in writing and formal speeches. This register appears to use a Tetun Dili base with strong mixing of Tetun Terik. Portuguese influence seems less obvious in this register than that seen generally in Tetun Dili, through a tendency to adopt Tetun Terik terms instead of Portuguese loanwords. However, more subtle effects of contact with Portuguese (the traditional language of the Catholic Church hierarchy) are evident in other areas such as syntax and semantics.

1.3 Previous research on Tetun

Published research on East Timorese languages has always tended to focus on Tetun Dili and Tetun Terik. Hull (1998) and Van Klinken (1999a) provide useful literature reviews of all previous linguistic and anthropological work, including dictionaries and pedagogical materials in English, Portuguese, Dutch and Indonesian.

The only full descriptive grammar is Van Klinken's (1999a) detailed grammar of the Fehan variety of Tetun Terik, as spoken in West Timor.¹ A condensed version of this research is presented in Van Engelenhoven and Williams-van Klinken (in preparation). Hull's (1996b) pedagogical manual of Tetun for English-speakers focuses on Tetun Dili but with substantial additional information on Tetun Terik. His dictionary too (Hull 1999) incorporates terms from both varieties.

¹ Hull and Eccles' (2001) reference grammar of Tetun became available after the final version of this manuscript was submitted.

1.4 Data and approach for this study

This grammatical description of Tetun Dili is based on the analysis of a large corpus of language data collected between 1999 and 2001, primarily in Dili in East Timor. The primary source consists of over 50 texts recorded from a wide variety of speakers, including men and women from a wide age range, and from a diversity of ethnic and educational backgrounds. Most were recorded privately by the authors or East Timorese colleagues, but a minority were formal interviews by journalists. While we have focussed mainly on oral texts and oral interaction, we have also made use of a large corpus of written materials to enable comparison with formal written registers of Tetun. This corpus includes gospel translations and liturgical materials, press and other official reports, and correspondence over the Internet. Since Tetun Dili is a language of inter-ethnic communication, and is spoken fluently by many people for whom it is not their native language (particularly in the ethnically mixed city of Dili), we have included data from non-native speakers in this grammar. In the phonology chapter, we have also made reference to pronunciations common amongst less fluent second-language speakers, who are found particularly in the rural areas.

The approach taken in this grammar is descriptive rather than prescriptive. As such we explicitly acknowledge inter-speaker variation, and attempt to note when constructions which are commonly used by one section of the population are considered unacceptable by others. As a reference grammar, this work follows the general principles of modern linguistics, without making a commitment to any particular formal model.

In addition to trained linguists, this grammar is intended to provide useful information to non-linguists both from within East Timor and from elsewhere who may be interested in learning more about Tetun Dili. We have therefore attempted to make it accessible to as wide a readership as possible by limiting the use of linguistic terminology and simplifying the technical presentation wherever it would not result in the loss of important linguistic and grammatical information. The grammar, however, remains primarily a linguistic description of Tetun Dili. Those interested in a much briefer, less technical, introduction to the language are referred to the Lonely Planet phrasebook (Hajek and Tilman 2001). Of course, since this present grammar is itself a rather brief preliminary one, not all aspects of the language are covered in as much depth as we would like.

1.5 Language structure

The linguistic structure of Tetun Dili may be summarised as follows.

Phonology: The language has a simple five vowel system. A large number of consonant phonemes and almost all examples of clusters have been borrowed from Portuguese. Stress is contrastive, but usually penultimate.

Lexicon: Tetun Dili shows a high degree of borrowing and calquing on Portuguese. Although especially evident in areas of technical and abstract vocabulary, the phenomenon also covers almost all greetings, all common vocatives, many common words, numerals (used in addition to Tetun and Indonesian ones), prepositions and conjunctions.

Morphology: There is limited productive morphology. In contrast to the more conservative Tetun Terik, there is no person–number marking on verbs. There is evidence of some

borrowing of Portuguese morphology, in particular the productive agentive suffix *-dór*. Portuguese verbs are normally borrowed in their 3S form (e.g. Tetun *kanta* from Portuguese *canta* ‘sing (3S)’). Portuguese nouns and adjectives are usually borrowed in the masculine form (e.g. Tetun *falsu* from Portuguese *falso* ‘false.MASC’), unless the feminine form specifically refers to women, such as masculine *sobrinhu* ‘nephew’ and feminine *sobrinha* ‘niece’.

Syntax: Clause order is essentially subject–verb–object, with some use of serial verbs. There is no passive voice. Modifiers follow the noun within the noun phrase. Numeral classifiers are mainly restricted to humans.

1.6 Standardisation

Although the Catholic Church has spent considerable time preparing religious and primary school materials in Tetun, there is no agreed standard as yet. As a result, a number of competing spelling systems are in use (see section 1.5 for a description of the orthography adopted in this grammar), and examination of written press in Tetun confirms our report above of substantial grammatical and lexical variation. Hull (1996a) and Hajek (2000) discuss language policy in East Timor, including the problem of standardisation.

1.7 Orthography

There is at present no generally accepted orthography for Tetun Dili. Areas of particular variation include:

- (1) the spelling of Portuguese loans: Some spell them as they are pronounced, others spell them as per Portuguese, and others take compromise solutions such as the omission of diacritics from Portuguese spelling (e.g. ‘komparasaun’, ‘comparaçãõ’ and ‘comparacao’ are alternative spellings of the Portuguese loan *comparaçãõ* ‘comparison’).
- (2) whether to represent the glottal stop, which is a phoneme in Tetun Terik but not in Dili (e.g. ‘ha’u’ versus ‘hau’ for *hau* ‘1S’).
- (3) whether and how stress or vowel length are marked (e.g. ‘moos’, ‘mos’ and ‘mós’ are common alternative spellings of /mos/ ‘also’).

Since the target readership of this book includes both native speakers of Tetun Dili and readers who have never heard the language, we are adopting a largely phonemic representation, while keeping in mind those orthographic traditions which are well established.² Taking into account the phonetic description of Tetun phonemes in chapter 2, all letters used have the expected values except that ‘nh’ represents a palatal nasal /ɲ/ and ‘lh’ a palatal lateral approximant /ʎ/,³ ‘j’ marks a voiced alveo-palatal fricative /ʒ/, and ‘x’ marks

² We break from tradition mainly in consistently marking stress, even where native speakers feel this to not be required. We follow tradition at the expense of pronunciation by spelling the perfective marker as *tihã* even though it is pronounced /tia/.

³ We here follow a relatively strong tradition of representing these Portuguese loan phonemes just as they are spelled in Portuguese. A reasonable alternative is to use ‘ny’ and ‘ly’, which, being more intuitive and more like Indonesian orthography, have been found to be easier to read for those Tetun speakers who have not been

its voiceless equivalent /ʃ/. For Indonesian loans, we follow Indonesian orthography in using ‘c’ to indicate the voiceless palatal stop /c/.

Most Tetun words have penultimate stress. Where stress is not penultimate, or where a word with a single vowel carries stress, we mark it by an acute accent over the vowel (e.g. *Timór* ‘Timor’, *hát* ‘four’).⁴ Note that rising diphthongs (/ei/, /ai/, /oi/, /ui/, /au/, and /eu/) are treated from this perspective as single vowels, and so are not marked for stress if they are penultimate (e.g. *kadeira* ‘chair’ and *ideia* ‘idea’ have stress on the /ei/). Where such rising vowel sequences are in fact not diphthongal, however, penultimate stress is explicitly marked (e.g. *saúdi* ‘health’).

exposed to Portuguese spelling conventions. Hull (1996b:xxvi) recommends ‘ñ’ and ‘ll’, as per medieval Portuguese.

There are three bimorphemic lexemes in Tetun Dili in which /nh/ occurs as a sequence of two phonemes, namely *bain-hira* ‘when’ (lit. ‘day-how.many’), *bain-hát* ‘in four days’ time’ (lit. ‘day-four’), and *bain-hitu* ‘in seven days’ time’ (lit. ‘day-seven’). These are written in this grammar with a hyphen, to prevent the ‘n-h’ being interpreted as a palatal nasal. There appear to be no words in which a sequence of /l/ and /h/ occurs.

⁴ The representation of stressed vowels is a difficult issue. We have chosen to mark it by an acute accent in this grammar since this method allows non-penultimate stresses to be marked consistently.

An alternative is to use double vowels rather than acute accents for stressed vowels in native Tetun words, as per Hull (1999). Many East Timorese authors use this system for some words (e.g. spelling *bót* ‘big’ as ‘boot’), while leaving stress unmarked in many others (e.g. spelling *nebé* ‘REL’ as ‘nebe’). In Portuguese loans, however, there is widespread consensus that vowel doubling is not acceptable orthographically. So, for instance, it is possible for people to accept spelling native Tetun *nanál* ‘tongue’ as ‘nanaal’ (although ‘nanal’ is more common), but people cannot accept spelling the Portuguese loan *animál* ‘animal’ as ‘animaal’. There are in addition some native Tetun words which are stressable, but for which orthographic vowel-doubling is not readily accepted by most writers; these include various words for which the vowel can be long or short according to context (e.g. *bá* ‘go’, *ba* ‘to’), and *sá* ‘what’.

2. Phonology

2.1 Consonants

2.1.1 Introduction

The maximal consonant phoneme inventory in Tetun is presented in Table 2.1,¹ where the headings give approximate place of articulation only, with more exact ones being given in the descriptions below. Unless otherwise indicated below, symbols have their usual IPA values in terms of voicing, place and manner of articulation.

Table 2.1: Consonants

	Bilabial	Labio-dental	Coronal	Alveo-palatal	Palatal	Velar	Glottal
Voiceless stop	p		t			k	
Voiced stop	b		d			g	
Voiceless fricative		f	s	ʃ			h
Voiced fricative		v	z	ʒ			
Nasal	m		n		ɲ		
Liquid			l		ʎ		
Trill			r				
Tap			ɾ				
Approximant	w				j		

The consonantal system in Tetun is extremely unstable and shows great variation within and across speakers, reflecting such things as educational background, social class and differences between urban and rural speakers. The maximal system given above is most commonly found in the speech of the traditional educated elite, who speak both Tetun and Portuguese. The influence of Portuguese and to a lesser degree Malay/Indonesian on the phonology of Tetun has been extensive; this is evident in the presence of the following

¹ Additional phonemes (e.g. a voiceless palatal stop /c/) are found in recent Indonesian loans. These loans are however not felt by East Timorese to be integrated into Tetun, so the loan phonemes are not discussed here.

phonemes: /p g v z ʒ ʒ̃ ɲ ʎ r j/ and partly also /w/. Portuguese influence is also the source of numerous allophonic processes found in the speech of many speakers. On the other hand, many speakers do not have the full set of consonant phonemes listed above, because the variable pronunciation of some of the phonemes results in phonemes overlapping or merging. In particular, /r/ and /r̃/ merge for most speakers. In addition, for many, especially those who are not native speakers of Tetun Dili, there is the possibility of merging for: /v/–/b/, /ʃ/–/s/, /ʒ/–/z/, /ɲ/–/n/, /ʎ/–/l/ and /d/–/r/.

Specific details about a number of phonemes and some phonological processes are given here:

2.1.2 Stops

All voiceless stops are unaspirated, and are generally unreleased in final position.

/b/ has a fricative allophone [β] (and less commonly [v]), which is not uncommon in intervocalic position, both within words and across word boundaries; e.g. *baba* ['baba] ~ ['baβa] 'drum'.

/t/ is clearly lamino-dental, but its voiced counterpart /d/ is a post-alveolar stop. The latter also has, for some speakers, a strongly retracted variant with retroflex quality, while for still others the voiced stop is optionally alveolar. In intervocalic position /d/ is often tapped; this results in a possible merger with /r/ (and/or /r̃/) in some words; e.g. *odamatan* [oda'matan] ~ [ora'matan] ~ [or'matan] 'door'.

/k/ is normally a voiceless velar stop. When it occurs in word-initial consonant clusters of the form /kC/, it has a range of pronunciations depending on the speaker. These include a weakly articulated [k̠] with the following consonant in a cluster, a clearly released [k] with vowel offglide, and a glottal variant [ʔ]; e.g. *kbú* ['k̠bi:t̠], [ki'bi:t̠], and [ʔbi:t̠] 'power'.

2.1.3 Fricatives

/v/ is a voiced labio-dental fricative which tends to be unstable and optionally alternates with or is replaced by /b/; e.g. *avó* [a'vo:] ~ [a'bo:] 'grandparent'.

/s/ is a voiceless fronted alveolar sibilant fricative. For many speakers it has a retracted post-alveolar variant [s̠] word-finally and before voiceless stops; e.g. *hás* ['ha:s] ~ ['ha:s̠] 'mango', *feſta* ['feſta] ~ ['feſta] 'party'. For other speakers, especially those who are fluent in Portuguese, this retracted allophone is in fact regularly backed even further to alveo-palatal [ʃ], as one finds in European Portuguese today; e.g. ['ha:ʃ], ['feʃta]. The whole phenomenon of /s/-retraction in Tetun is conditioned by a very similar and somewhat earlier phonological process of sibilant-retraction in European Portuguese. As our examples show, it can be found even in native Tetun words such as /has/.

/z/ is a voiced alveolar sibilant fricative. It appears initially (e.g. /zero/ 'zero'), medially (e.g. /gazo'lina/ 'petrol'), post-consonantly (e.g. /onzi/ 'eleven') and before voiced consonants (e.g. /ez'mola/ 'begging'). Before sonorants, some speakers optionally retract /z/ to [ʒ] or even [ʒ̃], once again due to European Portuguese influence; hence [ez'mola] ~ [eʒ̃'mola]. In

fixed expressions—borrowed from Portuguese—/z/ also surfaces at internal word boundaries; e.g. /es'taduz u'nidus/ 'United States', /'duaz 'oras/ 'two o'clock'. This reflects an earlier process of regular sibilant voicing between vowels in Portuguese. Some speakers may sometimes transfer this voicing to native Tetun compounds; e.g. *lís asu* ['li:s 'asu] ~ ['li:z 'asu] 'garlic' (lit. 'onion dog').

/ʃ/ (spelled 'x') is a voiceless alveo-palatal fricative; e.g. /kaiʃa/ 'box'. As noted in the discussion of /s/ above, many speakers reflect European Portuguese patterns by pronouncing /s/ as [ʃ] before voiceless stops and word-finally; e.g. *feira* ['feira] ~ ['feʃta] 'party'. On the other hand, many less Portuguese-influenced people pronounce expected /ʃ/ as [s], resulting in a merger between the /ʃ/ and /s/ phonemes; e.g. *xá* ['ʃa:] ~ ['sa:] 'tea'.

/ʒ/ (spelled 'j') is a voiced alveo-palatal fricative; e.g. /ʒa'nela/ 'window'. Like its voiceless counterpart, it is highly unstable across speakers and is often replaced by [z]; e.g. [ʒa'nela] ~ [za'nela]. Hypercorrect [dʒ] can now also be found in the speech of some younger speakers, due to the influence of Indonesian orthography; e.g. [a'dʒuda] instead of the expected [a'ʒuda] for *ajuda* 'help'.

/h/ is a voiceless glottal fricative; e.g. /'hare/ 'uncooked rice'. In intervocalic position a breathy voiced variant is common. It may be reduced further and optionally deleted; e.g. *ahi* ['ahi] ~ ['afi] ~ ['ai] 'fire'.

2.1.4 *Other consonants*

/n/ is a voiced apico-alveolar nasal. There is substantial inter-speaker variation in word-final position: while some speakers always maintain final [n], for others it is regularly velar [ŋ]. The velar nasal is often so reduced as to be absent, leaving only nasalisation on the preceding vowel (see section 2.2.3); e.g. *jardín* [ʒar'dĩ:n] ~ [ʒar'dĩ:ŋ] ~ [ʒar'dĩ:] 'garden'.

/ɲ/ (spelled 'nh') is a voiced palatal nasal; e.g. /'ʒuɲu/ 'June'. Speakers not native to Dili often reduce it to [jɲ] or [ɲ].

/l/ is a voiced lamino-alveolar lateral approximant; e.g. /'lalar/ 'fly'. It can be markedly velarised in syllable-final position, especially when word-final; e.g. *Portugal* [portu'gɔt] 'Portugal'. Less marked velarisation is also common in intervocalic position; e.g. *janela* [ʒa'neɫa] 'window'.

/ʎ/ (spelled 'lh') is a voiced palatal lateral approximant; e.g. /ba'ruʎu/ 'noise'. It is often reduced to [jɫ], [l] and more rarely [j] amongst speakers not native to Dili.

Some speakers contrast trilled /r/ and tapped /r/, although the distinction is not very stable and can only be found, where it exists, in intervocalic position in Portuguese loan words; e.g. /ka'reta/ 'car' and /fev'reiru/ 'February'. It appears that for most speakers the two rhotics are in free variation or in complementary distribution. For this reason, we do not distinguish the trill and the tap in our orthography.

/r/ is a voiced alveolar trill. A duration of 3–4 taps is common, often reducing in faster speech. Although alternation with /r/ is common, the trill seems the preferred rhotic in

syllable-initial position at least in careful speech; e.g. *riku* [ˈriku] ‘rich’, *hamrík* [hamˈri:k] ‘stand’.

/r/ is a voiced alveolar tap, with an optional post-alveolar variant. In non-final position, the single tap commonly alternates with a very short trill with 2 taps. In word-final position, the tap is usual; e.g. *fiar* [ˈfiar] ‘believe’.

/w/ is a voiced labio-velar approximant found in a small number of loanwords; e.g. /waiˈmoa/ ‘Waimoa’ (an ethnic group in East Timor), /ˈwarun/ ‘stall’ (from Indonesian *warung*). Although /w/ is common in Tetun Terik, the corresponding Tetun Dili words contain /b/; e.g. Tetun Terik /wainˈhira/ and Tetun Dili /bainˈhira/ ‘when’.

/j/ (spelled ‘i’ in Portuguese loans and ‘y’ in Indonesian loans) is a voiced palatal approximant found only in loans; e.g. *yayasan* /jajasan/ ‘foundation’ (Indonesian), *iodu* /ˈjodu/ ‘iodine’ (Portuguese).

2.2 Vowels

2.2.1 The basic vowels

Tetun has a five vowel system: /i, u, a, e, o/.

/i/ is a high front unrounded vowel.

/u/ is a high back rounded vowel. A slight fronting to [ɯ] is not uncommon in unstressed final open syllables; e.g. *etu* [ˈetu] ~ [ˈetɯ] ‘cooked rice’. Laxing to [ʊ] is possible in unstressed closed syllables; e.g. *rihun* [ˈrihʊn] ‘thousand’.

/a/ is a low unrounded central vowel. It can appear raised to [ɐ] in unstressed syllables, and can then be reduced further to [ə]; e.g. *oras* [ˈoras] ~ [ˈorɐs] ~ [ˈorəs] ‘time’, *sasík* [saˈsi:k] ~ [saˈsi:kʰ] ‘riddle’. This raising and reduction is well-known in European Portuguese, and is often most evident, albeit not exclusively, in the speech of Tetun–Portuguese bilinguals.

/e/ is a front unrounded mid vowel. The precise height of this vowel appears to vary greatly according to speaker and to context. For many speakers, the basic quality may be true-mid (or even slightly lower), with some evidence of limited vowel harmony seen in slight raising to [e̝] before following high vowels (especially /u/); e.g. *hemu* [ˈhemu] ~ [ˈhɛ̝mu] ‘drink’. For other speakers the front mid vowel is generally high-mid regardless of the following vowel. All speakers tend to open the vowel when it is next to /h/; e.g. *makehe* [kaˈkɛfiɛ] ‘fan’.

/o/ is a back rounded mid vowel. Once again speakers show considerable variation. For many speakers, the basic allophone is open [ɔ]; e.g. *tohar* [ˈtɔhar] ‘break’. For others, the basic allophone tends to true-mid. When the following syllable contains a high vowel (/i, u/), /o/ is usually raised to close [o] by vowel harmony; e.g. *tohu* [ˈtoɦu] ‘sugar-cane’. The vowel is also usually close before another vowel; e.g. *koi* [ˈkoi] ‘scratch’, *foun* [ˈfou̯n] ‘new’ and *moe* [ˈmoe̯] ‘shame’.

/e/ and /o/ tend to be expressed as a higher [i] and [u] in unstressed position in Portuguese loanwords; this same phenomenon is found in many varieties of Portuguese spoken around the world. Raising is particularly common in word-final position; e.g. [pre'zente] ~ [pre'zenti] from Portuguese *presente* 'present'. However it also occurs in unstressed initial syllables; e.g. [portu'gaŋ] ~ [purtu'gaŋ] from Portuguese *Portugal* 'Portugal'. When the unstressed vowel occurs word-initially before /sC/ clusters, it is often absent altogether, as in European Portuguese; e.g. [es'trada] ~ [is'trada] ~ [i'trada] ~ [ʃ'trada] from Portuguese *estrada* 'street'.

2.2.2 *Diphthongs*

In native Tetun vocabulary, sequences of vowels are easily distinguished as separate vowels in careful speech; however in normal speech they tend to coalesce into diphthongs; e.g. *sae* ['sa.e] ~ ['sae] 'ascend'. Diphthongs are also frequent in Portuguese loanwords but in normal speech speakers tend to reduce many of these to single vowels; e.g. *kadeira* [ka'deira] ~ [ka'dera] 'chair', *senoura* [se'noura] ~ [se'nora] 'carrot'.

2.2.3 *Nasalisation*

All vowels adjacent to nasals are usually nasalised; e.g. *masin* ['māsĩn] 'salt', *mama* ['māmā] 'chew betelnut', *kanta* ['kānta] 'sing'.

2.2.4 *Reduction of compounds and reduplications*

In compounds and reduplicated forms, the final coda of the first word is usually deleted. In addition, the vowel of the final syllable is in normal speech typically weakly articulated, and indeed often deleted altogether, leaving the onset of the final syllable to function as the coda to the preceding one; e.g. *masin-midar* ['māsĩn 'mĩdar] ~ ['māsĩ 'mĩdar] ~ [mās 'mĩdar] 'sugar' (lit. 'salt sweet'), *manu-tolun* ['mānʷ 'tolũn] ~ [mān 'tolũn] 'egg' (lit. 'egg bird'), *loro-loron* ['loro 'lorõn] ~ [lor 'lorõn] 'daily'.

2.3 *Stress and vowel length*

Stress is fully contrastive in Tetun; e.g. /'kabas/ 'cotton, thread, yarn' versus /ka'bas/ 'shoulder', /'kaben/ 'marry' versus /ka'ben/ 'saliva'. Nevertheless it falls most commonly on the penultimate syllable; e.g. /ha'karak/ 'want', /ka'miza/ 'shirt'. Final stress occurs in both native and loaned items; e.g. /ha'ris/ 'bathe', /si'nal/ 'sign'. Antepenultimate stress is much less common and is found only in a small number of Portuguese loanwords; e.g. /'mediku/ 'doctor', /a'merika/ 'America'.

Unstressed vowels are always short, as are non-final stressed vowels; e.g. [a'merika] 'America', [ha'karak] 'want'. However, all vowels, regardless of height or quality, are regularly lengthened in stressed monosyllables and stressed final position; e.g. /'hat/ ['ha:t] 'four', /ka'bas/ [ka'ba:s] 'shoulder', /ha're/ [ha're:] 'see'. The possibility that vowel length may be contrastive in Tetun cannot, however, be totally excluded; some speakers for instance appear not to always lengthen the vowel as expected. As a result we can find /ani'mal/ [ãni'mã:] instead of expected [ãni'mã:t] 'animal' alongside regular /na'nal/ [nã'nã:t] 'tongue'. Further investigation is required on this issue.

2.4 Phonotactics

The syllable template in Tetun is (C)(C)V(V)(C). There are only two single-vowel words, namely /o/ '2s, you' and /i/ 'and'.

The range of initial clusters in native vocabulary is restricted to a small set of /kC/ combinations which—as we have seen in section 2.1.2—are in any case prone to elimination through optional vowel insertion; e.g. /'kmanek/ ['kmanek'] ~ [ka'manek'] 'wonderful'. Intensive borrowing from Portuguese and to a lesser degree from Malay/Indonesian has led to the introduction of a large number of previously inadmissible clusters, including (s)C+liquid sequences; e.g. /planta'saun/ 'plantation', /'primu/ 'cousin', /es'traga/ 'damage'. As in Portuguese, /sC/ clusters never appear word-initially, at least phonemically.

3. Word classes

3.1 Introduction

This chapter presents an overview of word classes. Many of the classes will be discussed in more detail in subsequent chapters.

Word classes are determined by a number of criteria. These are the range of syntactic functions they can fulfil, their distribution within syntactic units, the categories which are specifiable for members of that class (e.g. nouns can be marked as plural), derivational possibilities, and semantic criteria. As is to be expected, the various criteria sometimes conflict. In this case priority is given to the syntactic criteria, and the discrepancies are noted.

Some words are members of more than one word class, with a difference in meaning between the two classes. The significant overlap between nouns and verbs is discussed in sections 4.2.2 and 4.2.3.

In other instances, a word is analysed as belonging to two classes simply because it has the distribution of both classes, there being no difference in meaning between them. This overlap is common between temporal prepositions and conjunctions (with words like *tó* 'until' being able to introduce both noun phrases and clauses), between pronouns and determiners (with words like *né* 'this' being able to either head noun phrases or modify them), and between adjectives and manner adverbs.

3.2 Nominal and verbal classes

The major word classes are nouns, verbs and adjectives. The closed class of pronouns shares many characteristics with the open class of nouns, while verbs and adjectives have so much in common that they can be considered as belonging together in a larger superclass.

Nouns and pronouns typically function as heads of noun phrases; no other word classes can fulfil this function. Only nouns and pronouns can function as a preposed possessor within a noun phrase, and only they can be enumerated by numerals. Unlike nouns, pronouns (discussed in sections 5.2 and 5.3) cannot function as modifier within a noun phrase unless they are also classified as determiners.

Verbs usually function predicatively. While all verbs allow a subject, other arguments are restricted to certain subclasses of verbs. Noun phrase modifiers headed by verbs are usually introduced by a relative clause marker.

Adjectives function predicatively or attributively. When used attributively, there is usually no relative clause marker. Some adjectives can be fully reduplicated to derive an adverb or to indicate plurality of referents (see section 4.3). Many adjectives are also classed as manner adverbs.

Predicates headed by verbs and adjectives can typically be negated by *la* ‘not’, while those headed by nouns are not, being restricted to negation by the usually contrastive *laós* ‘indeed not’. Predicates headed by nouns are restricted in the tense–aspect marking they can take (e.g. in not occurring with *atu* ‘IRREALIS’). Adjectives and intransitive verbs are the base classes for verbal derivations with *ha-* (see section 4.2.1).

Quantifying adjectives (e.g. *barak* ‘many, much’, *uituan* ‘a few, a little’) are a small subclass of adjectives which have characteristics of both adjectives and quantifiers. In common with adjectives, they can occur predicatively or attributively, can be intensified, and can function as verb modifiers (e.g. *koalia barak demais* ‘talk much too’ = ‘talk too much’). In common with quantifiers (a subclass of determiners), they indicate quantity, and can readily function as noun phrase heads. The quantifying adjective *barak* ‘many, much’ can, like the quantifier *hotu* ‘all’ (section 5.3.6), float to post-verbal position when it modifies a subject.

3.3 Other word classes

Determiners (section 5.3) indicate definiteness and/or number within noun phrases. Unlike quantifying adjectives (e.g. *barak* ‘many’), determiners cannot function predicatively. Determiners include the following subclasses: demonstratives, interrogatives, singular *ida* ‘one, a’, plural *sira*, and quantifiers.

Numerals are discussed in section 5.7.1. They are the only class of words that can be used to form complex numerals. Native Tetun (as opposed to borrowed) numerals follow the numeral classifier within a numeral phrase, and can be reduplicated with numeral-specific meanings (section 4.3). Numerals mainly function attributively (without a relative clause marker) but can be predicative.

N numeral classifiers (section 5.7.2) are a small class of words which occur immediately preceding a numeral in a numeral phrase.

Adverbs modify constituents other than nouns. They cannot function as predicates, and cannot take arguments. With the exception of manner adverbs (all of which are also classified as adjectives), they cannot be negated or intensified. Some types of adverbs are discussed in chapter 8.

Tense–aspect markers, described in section 8.2, specify the tense or aspect of a clause. They cannot head a predicate.

Auxiliaries (section 8.3) immediately precede the verb (or other predicate head). They can usually stand alone as the predicate without a following verb if the context is understood. *Seidauk* ‘not yet’ is a tense–aspect marker functionally, but syntactically falls within the auxiliary class. Other auxiliaries deal with permission or obligation (e.g. *bele* ‘can’).

Prepositions (chapter 6) can all introduce peripheral phrases. Some can, in addition, introduce oblique arguments or noun phrase modifiers, or function as predicates. Prepositions require a noun phrase complement, which cannot be fronted or omitted.

Conjunctions are grammatical items which function to connect words or clauses. These can be divided into various subtypes.

1. Subordinating conjunctions (sections 6.6 and 10.6) introduce adverbial clauses, that is, full clauses functioning as adverbial modifiers within a superordinate clause.
2. Complementisers (section 10.5) mark a clause as the complement of a verb or noun.
3. Relative clause markers (section 10.8) introduce a relative clause, linking it to the noun phrase that it post-modifies.
4. Coordinating conjunctions (section 10.4) coordinate a range of constituent types, including noun phrases and clauses; they are placed between the constituents they coordinate.

Tags typically follow the final clause in a sentence, being linked to it intonationally. Each tag has an associated characteristic intonation pattern. The tag *ka* forms a yes–no question (section 7.6.2).

Interjections take no part in clausal syntax, and typically (except in the case of hesitation markers) constitute utterances (or at least intonation groups) on their own. They need not conform to normal Tetun phonological patterns. One can distinguish the following subclasses of interjections, primarily on semantic grounds.

1. Pro-clauses typically form single-word responses to a preceding utterance (e.g. *sín* ‘yes’, *lae* ‘no’).
2. Social formulae: Most greetings and other interjections of politeness are incorporated from Portuguese (e.g. *boa viajen* ‘have a good journey’, *bon dia* ‘good morning’, *deskulpa* ‘sorry’, *obrigadu/a*¹ ‘thank you’, *nada* ‘you’re welcome’ (a standard response to *obrigadu/a*), *kolisensa* ‘excuse me’, and *favór ida* ‘please’).
3. Exclamations express attitudinal rather than propositional meanings. They are nearly always utterance-initial, but can occur elsewhere as well, usually between (rather than within) clauses. While there is significant inter-speaker variation, somewhat standard exclamations include *ai!* to express difficulty, *eipá!* (with falling intonation on *pá*) to express frustration, *hei!* to attract someone’s attention from a distance, and *koitadu* ‘poor thing! what a pity!’ to express sympathy.
4. Hesitation markers (e.g. *bé*) occur fairly freely within utterances, and appear to fill gaps in speech while the speaker determines what to say. They are not used in writing.

¹ Some Tetun speakers distinguish *obrigadu* and *obrigada* according to the gender of the speaker, with men saying *obrigadu* and women *obrigada* (as per standard Portuguese). For many, however, the choice depends on the gender of the addressees, with *obrigadu* being used to men or mixed groups, and *obrigada* to women. Similarly, some speakers distinguish between singular *boa tardi* ‘good afternoon’ for greeting one person and plural *boas tardis* for greeting more than one; many people however do not consider the plural form (which is not found in standard Portuguese) to be fully acceptable.

4. Morphology

4.1 Overview

Tetun Dili has relatively little productive affixation. Many of the affixes found in Tetun Terik (van Klinken 1999a:58–98) either do not occur in Tetun Dili, or occur fossilised in only a few words. Of these, we discuss only *-n* as a relic marker of inalienable possession (see section 5.4.4).

In the verbal domain, causative verbs can be derived from intransitive verbs and adjectives by *ha-*; this is discussed in 4.2.1 and along with other causative constructions in section 9.6.2. There are also several low-frequency prefixes which derive intransitive verbs from transitive ones (see 4.2.4).

Verbal expressions can be derived from nouns by the prefix *ha-*, by zero derivation (or ‘conversion’, i.e. with no change in form), or by making the noun the object of *halo* ‘do, make’ (section 4.2.2). Abstract nouns are derived from verbs and adjectives by zero derivation.

The suffix *-dór*, bound root *-tén*, and noun *nain* are all used to derive adjectives or nouns denoting actors. The bound form *-nulu* ‘tens’ attaches to the numerals 2 to 9 to derive the numerals 20 to 90 (e.g. *hitu-nulu* ‘seven-tens’ = ‘seventy’; see section 5.7.1).

Full reduplication derives adverbs, as well as having other functions discussed in section 4.3. Compounding is a productive means of forming nouns.

4.2 Affixation and zero derivation

4.2.1 *Derivation from intransitive to causative verb*

The prefix *ha-* is used with intransitive verbs and adjectives to derive causative verbs. This is just one of the available methods for encoding causation in Tetun Dili, and is discussed in more detail in section 9.6.2.

Vi/Adj		Vt	
<i>lakon</i>	‘disappear’	<i>ha-lakon</i>	‘cause to disappear, abolish’
<i>manas</i>	‘hot’	<i>ha-manas</i>	‘heat’
<i>moe</i>	‘ashamed, embarrassed’	<i>ha-moe</i>	‘shame, embarrass’
<i>sai</i>	‘exit’	<i>ha-sai</i>	‘remove, extract’

4.2.2 Derivation from noun to verb

The prefix *ha-* also derives a small number of transitive verbs from nominal roots, where the verb bears some semantic relationship to the root noun.¹

Noun		Verb	
<i>domin</i>	'love'	<i>hadomi</i>	'love, pity'
<i>kraik</i>	'below, lower part'	<i>haraik</i>	'humble' (<i>haraik-an</i> 'humble oneself')

An alternative means of deriving verbs from nouns is simply by zero derivation (that is, there is no change in form).² This strategy is used, for example, to relate tools and the (usually transitive) verb for the use of those tools. A significant number of such derivations are from Portuguese nouns.³

	Noun gloss	Verb gloss
<i>bahat</i>	'chisel'	'chisel, carve'
<i>kadó</i>	'saw'	'saw'
<i>bomba</i>	'pump'	'pump up'
<i>telefone</i>	'telephone'	'phone, ring up'
<i>xavi</i>	'key'	'lock up'

Zero derivation is also used to derive verbs from abstract nouns (e.g. *sala* N 'error'; Vi 'err'). This is most obvious with Portuguese nominal loans which are used as verbs in Tetun (e.g. *servisu* N, Vi 'work'). There is considerable inter-speaker variation on such zero derivation, with some speakers, for instance, using *dieta* 'diet' and *jejún* 'fast (go without food)' as verbs, while others prefer the construction using *halo* 'do, make', discussed in the next paragraph (*halo dieta* 'do/make diet' = 'diet', *halo jejún* 'do/make fast' = 'fast').

A far more productive means of creating verbal expressions from abstract nouns is to make the noun the object of the verb *halo* 'do, make'. Recent loan words, such as English 'check in' below, are readily incorporated into this construction.

¹ These examples illustrate low-frequency morphophonemic processes that are not discussed in the phonology chapter due to the necessarily brief presentation, namely the fact that in some derivations using the prefix *ha-*, the final coda consonant and/or the initial /k/ of root-initial consonant clusters is deleted. Van Klinken (1999a:38, 63) comments briefly on the same phenomena in the Fehan dialect of Tetun Terik, where the *ha-* prefix is much more productive.

² An instance in which *ha-* is used by some conservative speakers but replaced with zero derivation by younger speakers is *hakail/kail* 'fish with a line', from *kail* 'fishing hook'.

³ It is the Portuguese nominal loans which show that the noun is primary in Tetun Dili. In the conservative Fehan dialect of Tetun Terik, in contrast, it is the verb that is primary, with nouns derived by partial reduplication (e.g. *kakorus* 'grater' from *korus* 'grate') (van Klinken 1999a:80).

Some speakers also use zero derivation for nouns denoting plant or animal products, to derive verbs meaning 'to produce that which is denoted by the root noun'; however this is not accepted by others. An alternative is to make the noun the object of the verb *fó* 'give' (e.g. *isin* N 'contents, bulb', Vi 'bear grain/fruit', *fó isin* 'bear grain/fruit'; *tolun* N 'egg', Vi 'lay an egg', *fó tolun* 'lay an egg').

Noun		halo noun	
<i>check in</i>	‘check in’	<i>halo check in</i>	‘check in’
<i>desizaun</i>	‘decision’	<i>halo desizaun</i>	‘decide, make a decision’
<i>funu</i>	‘war’	<i>halo funu</i>	‘fight’
<i>intervista</i>	‘interview’	<i>halo intervista</i>	‘interview, conduct an interview’
<i>tinan</i>	‘year’	<i>halo tinan</i>	‘have a birthday’

4.2.3 Derivation from verb or adjective to abstract noun

A significant number of native Tetun verbs and adjectives can be nominalised without change of form. Such deverbal nouns are usually (though not necessarily) possessed, with the possessor corresponding to the subject of the underlying verb. Reflexive *-an* can be part of the nominalisation (as in *haraik-an* below).

Verb/Adj		Example with abstract noun	
<i>diak</i>	‘good’	<i>sira nia diak</i>	‘their well-being’
		3P POS goodness	
<i>hakarak</i>	‘want’	<i>povu nia hakarak</i>	‘the desire/will of the people’
		the.people POS desire	
<i>hanoin</i>	‘think’	<i>hau nia hanoin</i>	‘my opinion’
		1S POS thought	
<i>haraik-an</i>	‘humble/lower-self’	<i>ita nia haraik-an</i>	‘our humility’
		1PI POS humility	
<i>moras</i>	‘sick, sore’	<i>se iha moras ruma</i>	‘if there is an illness’
		if EXIST sickness some	

4.2.4 Detransitivising derivation

The prefix *nak-* derives an intransitive verb with an undergoer subject from a transitive verb root (e.g. *nakduir* Vi ‘roll’ from *duir* Vt ‘roll’, *naklés* Vi ‘torn’, from *lés* Vt ‘tear’). The resultant verb does not imply an actor, and can be used equally for something that happens of its own accord (e.g. a cloth tore because it was old or caught on a thorn) or for an event that is (usually unintentionally) caused by someone. The form *nam-* appears to be a less common variant of *nak-* (e.g. *namkari* Vi ‘scatter’ from *kari* Vt ‘scatter’).⁴

- (4.1) *Rai komesa nak-doko.*
 earth start INTR-shake
 ‘The earth started to shake.’

⁴ In Tetun Terik, the form of the detransitivising prefix is *hak-*, with its 3S inflection *nak-* being fossilised for verbs whose subject is necessarily non-human. While the historically uninflected forms *hak-* and *ham-* are found in some Tetun Dili words which can take human subjects (e.g. *hakfuik* ‘whistle’, *hamlaha* Adj ‘hungry’), these forms are no longer synchronically analysable in Tetun Dili, and the prefixes are unproductive.

- (4.2) a. *Ita fakar bé.*
 1PI spill water
 ‘We spill the water (on purpose).’
- b. *Ita halo bé nak-fakar.*
 1PI make water INTR-spill
 ‘We cause the water to spill (e.g. accidentally).’

4.2.5 Derivation of actor nouns and adjectives

4.2.5.1 -dór

The suffix *-dór* attaches to a verb *X* to derive a noun meaning ‘person who habitually does *X*’. Although this is a Portuguese suffix, it occurs readily on Tetun roots (e.g. *hemu* ‘drink’ below). The feminine form ending in /a/ is restricted to words that are borrowed directly from Portuguese (e.g. *administradora* ‘mayor, administrator (female)’). So too is the use of *-dór* to derive words that do not denote persons (e.g. *komputadór* ‘computer’).

Verb		Noun	
<i>dukur</i>	‘sleep’	<i>dukurdór</i>	‘sleepyhead, person who sleeps a lot’
<i>hemu</i>	‘drink’	<i>hemudór</i>	‘drinker (of alcohol)’
<i>servisu</i>	‘work’	<i>servisudór</i>	‘one who is diligent, hard-working (but not necessarily competent)’

- (4.3) *Ita ema Timór mama-dór.*
 1PI person Timor chew.betel-AGENT.N
 ‘We Timorese are habitual betel-chewers.’

4.2.5.2 -tén

The bound root *-tén* attaches to root verbs, adjectives or nouns. The resultant adjective describes a person who habitually does the activity of the root, or who displays a characteristic identified by the root.⁵ The derivations are impolite, indicating that the description is considered a bad thing. This negative connotation presumably derives from the use of *tén* as a noun meaning ‘faeces’. It applies predominantly to Tetun roots, but also to Portuguese ones (e.g. *kafé* below).

Root		Adjective	
<i>baruk</i>	Adj ‘lazy’	<i>baruk-tén</i>	‘lazybones’
<i>bosok</i>	Vt ‘deceive, lie’	<i>bosok-tén</i>	‘liar’
<i>kafé</i>	N ‘coffee’	<i>kafé-tén</i>	‘coffee addict’
<i>lanu</i>	Adj ‘drunk’	<i>lanu-tén</i>	‘drunkard’

⁵ Note that although the resultant terms are adjectives, they are most easily translated by nouns in English.

- (4.4) *Nia nakar-tén la-halimar lós!*
 3S naughty-AGENT.A not-play true
 'He's terribly naughty!'

4.2.5.3 *nain*

The noun *nain* follows a verb X to form a nominal compound meaning 'one who is or does X'. It tends to denote someone who has a recognised right or ability to do something. This is presumably carried over from the use of *nain* as a noun meaning 'owner' or 'master' (e.g. *loja nain* 'shop owner' = 'shop owner, shop-keeper').

Verb		Compound	
<i>pinta</i>	Vt 'paint'	<i>pinta nain</i>	'painter'
<i>sala</i>	Vi 'err, do wrong'	<i>sala nain</i>	'sinner'
<i>servisu</i>	Vi 'work'	<i>servisu nain</i>	'competent hard worker'

- (4.5) *Nia ema kaben nain ida.*
 3S person marry master one
 'He is a married person.'

4.3 Reduplication

4.3.1 *Iteration*

Reduplication of nouns specifying units of time derives temporal adverbs meaning 'every unit-of-time' (e.g. *ful-fulan* 'monthly' from *fulan* 'month', *dadér-dadér* 'every morning' from *dadér* 'morning', *Janeiru-Janeiru* 'every January' from *Janeiru* 'January'). Some such derivations have taken on more general meanings (e.g. *ora-oras* 'often, off and on; soon' from *oras* 'hour', *bai-bain* 'daily, usually' from the bound root *bain* 'day').

Where a unit of time is specified by a noun and modifying numeral, reduplication of the numeral has the same effect of indicating 'every so-many units of time' (e.g. *minutu tol-tolu* 'minute RDP-three' = 'every three minutes').

4.3.2 *Collectivity*

Reduplication of numerals or of the quantifying adjective *uitoan* (or *ituan*, *oituan*) 'a little' indicates that the specified number of items is being considered as a group, as in (4.6) and (4.7). For *ida* 'one', reduplication thus indicates that each is considered individually, as in (4.8).⁶

- (4.6) *Imi tenki lao nain ru-rua.*
 2P must walk CLS:human RDP-two
 'You must walk in groups of two.'

⁶ The related form *ida-idak* means 'each'.

(4.7) *Depois mai, hán utoan-utoan deit.*
 afterwards come eat RDP-a.little just
 ‘Afterwards (we’d) come back, and eat just a little at a time.’

(4.8) *Sira mai ida-ida.*
 3P come RDP-one
 ‘They came one at a time.’

Some adjectives can be reduplicated when they describe plural referents; for some terms, such reduplication is particularly applicable when the referents are diverse (e.g. *selu-seluk* ‘others, various other’ from *seluk* ‘other’).⁷

(4.9) *Nia fó paun ki-kík mai hau.*
 1S give bread RDP-small come 1S
 ‘He gave me small pieces of bread.’

In colloquial speech, verbs are occasionally repeated to indicate that the activity is intense and/or long-lasting (e.g. *Ita eduka eduka eduka* ‘1PI train train train’ = ‘We discipline (them) over and over’). This strategy is however strictly avoided in writing.

4.3.3 Deriving adverbs from adjectives

Reduplication derives a few adverbs from adjectives. These include *di-diak* ‘carefully, thoroughly’ from *diak* ‘good’ (also an adverb ‘well, successfully’), and *lai-lais* ‘quickly’ from *lais* ‘quick’.

4.4 Compounds

4.4.1 Introduction

Tetun has many nominal expressions which are compound-like (in that they represent the conventional names for things and therefore appear to be lexicalised), but which follow normal grammatical rules for noun phrase formation, and are semantically regular. Phonologically, such compounds have a single primary stress just as single words do. Some have reduction of the initial member, the degree of reduction in some cases depending on the speaker or the degree of care taken in speaking (e.g. *lo(k)raik* ‘afternoon’ from *loro-kraik* ‘sun-low’, *mas(i)-midar* ‘sugar’ from *masin-midar* ‘salt-sweet’).

Syntactically, too, a compound is a single word. As such it is not separable; possessive compounds, for instance, do not have possessive marking (e.g. *ain-kabun* ‘calf (of leg)’ from

⁷ We are unable to confirm the suggestion by Hull (1996b:40) and Morris (1984b:xiv) that reduplicating adjectives indicates intensity in Tetun Dili. The main term which could be considered for such an analysis would be *bar-barak* from *barak* ‘many, much’. However even here the reduplication always adds an element of diffuseness, without necessarily intensifying the quantity; resultant glosses are ‘many and diverse’ or ‘much and aimless’ (e.g. *koalia bar-barak* ‘talk RDP-much’ = ‘ramble on, talk unceasingly with little purpose’). A marginal possibility is the adverb *tebe-tebes* ‘truly, extremely’ from the adjective and adverb *tebes* ‘true, truly’.

'leg-stomach, not **ain nia kabun* 'leg POS stomach'); coordinative compounds do not include a coordinator (e.g. *inan-aman* 'parents' from 'mother-father', in contrast to phrasal *inan ho aman* 'mother and father'); and noun–adjective compounds cannot be separated by a relative clause marker (e.g. *manu-metan* 'crow' from 'bird-black', in contrast to *manu nebé metan* 'bird REL black' = 'a bird that is black'). The members of the compound cannot be individually modified.

Although there is at present little consensus on Tetun orthography, there is in practice widespread agreement that some compounds must be written as one word (e.g. *aimoruk* 'medicine' from *ai-moruk* 'plant-bitter'), and some as two (e.g. *tua sín* 'vinegar' from 'wine sour'). To mark an intermediate level of cohesiveness, some writers use a hyphen (e.g. *lia-fuan* 'word' from 'language-round.entity', is written by some East Timorese as two words, and by some with a hyphen).

4.4.2 Nominal compounds

The vast majority of Tetun compounds are nominal, and consist of two terms. The various subtypes are discussed below.

Possessor-head constructions derive names for many body parts (e.g. *ibun kulit* 'lip' from 'mouth-skin'), animal products (e.g. *bani bén* 'honey' from 'bee-juice', *manu lian* 'sound of birds' from 'bird-voice'), and plant parts (e.g. *ai-funan* 'flower' from 'plant-flower'). Some common nouns in the second slot for plant parts are *hún* 'tree, base', *bén* 'sap, juice', *fuan* 'fruit', *musan* 'seed', and *tahan* 'leaf'. Some, such as *tarak* 'thorn' and *funan* 'flower', require that the possessor be explicit in context, either as a possessor (e.g. *sabraka tarak* 'orange thorn') or elsewhere in the clause (e.g. *Sabraka iha tarak* 'Orange(s) have thorn(s)'); if it is not clear or not important what plant the part is from, then the generic *ai* is used as possessor (e.g. *ai-tarak* 'thorn (in general)'). In addition, possessor-head constructions derive a wide range of other compounds (e.g. *foer fatin* 'rubbish bin' from 'rubbish place', *bé matan* 'spring' from 'water source', *tós nain* 'farmer' from 'garden owner', *Timór oan* 'native Timorese (person)' from 'Timor child/native').

Activity-place compounds provide conventionalised means of denoting places where certain activities are usually performed. The head noun is *fatin* 'place', and the activity is usually given by a verb (e.g. *moris fatin* 'birthplace' from 'be.born place', *harís fatin* 'bathroom / bathing area' from 'bathe place'). Alternatively a verb–object pair may be used to denote the activity (e.g. *tebe bola fatin* 'football field' from 'kick ball place', *fasi ropa fatin* 'washing area' from 'wash clothes place'). More complicated verb phrases (e.g. with adverbial modifiers) are however not used in this construction.

Head-modifier compounds allow a range of modifier types. Noun modifiers can relate to the head in diverse ways, including specifying source (e.g. *forai* 'peanut' from *fore-rai* 'legume-ground'), characteristic part (e.g. *nehek liras* 'flying ant' from 'ant wing'), likeness (e.g. *fore xikoti* 'snake bean' from 'bean whip'), and gender (e.g. *oan feto* 'daughter' from 'child woman/girl'). Adjectival modifiers are also common (e.g. *feto foun* 'daughter-in-law' from 'woman new'), with verbs occurring less frequently (e.g. *ai habit* 'peg' from 'wood squeeze').

Generic-specific compounds consist of two nouns, of which the first is a generic classifier, and the second has a more specific meaning. Many plant, bird, and fish names are compulsorily preceded by the nominal classifiers *ai* 'plant', *manu* 'bird' or *ikan* 'fish' respectively (e.g. *ai-teka* 'teak (tree)', *manu radi* 'duck'). For others the classifier is not used (e.g. *au* 'bamboo', *makikit* 'eagle'). For nouns denoting humans it is not uncommon to have a preceding *ema* 'person' (e.g. *ema durbasa* 'interpreter'). Disease names optionally have a preceding generic *moras* 'illness' (e.g. *moras tétanu* 'tetanus'). Generic-specific compounds vary in the degree of binding between the two terms. At one extreme are terms such as *ai-dila* 'pawpaw', which native speakers prefer to write as a single or hyphenated word, and which retains the generic *ai* even when one is referring to part of the plant (e.g. *ai-dila tahan* 'pawpaw leaf'). At the other extreme are terms with *manu* 'bird', which speakers prefer to write as two separate words, and from which the *manu* is omitted when the bird name is used as a modifier (e.g. *nán rade* 'duck meat' from 'meat duck').

4.4.3 Other

In addition to nominal compounds, there is a range of semantically coordinate compounds, where both terms come from a single semantic set. Such coordination is found for nouns (e.g. *maun-alin* 'brothers (and sisters)' from 'older.brother-younger.sibling', *aban-bainrua* 'the future' from 'tomorrow-day.after.tomorrow'), adjectives (e.g. *midar-sín* 'sweet and sour' from 'sweet-sour'), and verbs to derive adverbs (e.g. *bá-mai* 'to and fro' from 'go-come', *tún-sae* 'up and down, all over the place' from 'ascend-descend').

There are also various compound-like sequences of verbs, which are discussed as 'nuclear serialisation' in chapter 9. Amongst the more common of these are causative sequences beginning with *fó* 'give' or *halo* 'do, make'; these are discussed along with other causative constructions in 9.6.

Finally, there are a large number of idioms denoting human bodily and emotional states, which consist of a body-part noun and a following adjective or verb (e.g. *isin diak* 'well, healthy' from 'body good'). Most of these are not compounds when used adjectivally, since the two terms are separable. Nevertheless, some are compounded (e.g. *isin-rua* 'pregnant' from 'body-two'), and the remainder can be analysed as compounds when they are used as nouns (e.g. *isin manas* 'fever' from 'body hot'). For discussion of such expressions, see section 7.2.

5. Noun phrases

5.1 Basic noun phrase structure

The vast majority of noun phrases headed by common nouns fit the following pattern.

(Psr)	N	(NP)	(AdjP)	(Det*)	(PP)
			(NumP)		(RelCl)
			(Psr)		(Complement)

where Psr = possessor (section 5.4), NP = noun phrase modifier (section 5.5), AdjP = adjective phrase (section 5.6), NumP = numeral phrase (section 5.7), Det* = one or more determiners (section 5.3), PP = prepositional phrase (example 5.3), and RelCl = relative clause (section 10.8).

That is, possessors either precede or follow the head, while all other modifiers follow the head. Those lexical modifiers that are prototypically short come first, followed by determiners, followed in turn by longer modifiers. Excluding determiners and possessors, it is uncommon to have more than one modifier within a noun phrase, and more than two is very rare. Where qualitative and quantifying modifiers co-occur, the one with the largest scope comes last. This is usually the one expressing quantity (e.g. *buat foun barak* ‘thing new many’ = ‘many new things’), but may also be non-quantifying (e.g. *loron tolu primeiru* ‘day three first’ = ‘the first three days’). Some noun phrases with multiple modifiers are illustrated below, with the constituents labelled for easier comparison with the template.

(5.1) *hau nia uma né*
IS POS house this
Psr N Det
‘this house of mine’

(5.2) *NGO internasionál sira nebé hakarak atu ajuda*
NGO international PL REL want IRR help
N Adj PL RelCl
‘those international NGOs (non-government organisations) who want to help’

Many noun phrases consist of a single noun without modifiers; these are discussed in section 5.3.1.

It is of course also possible to have noun phrases headed by pronouns (e.g. *sé* ‘who’). In this case the most common modifier is *né* ‘this’ (e.g. *nia né* ‘3S this’ = ‘he/she whom we are talking about’). More complex sequences of modifiers are however also possible, as illustrated by (5.3).

- (5.3) *sira nain hát iha nebá né*
 3P CLS:human four LOC there this
 Pronoun NumP PP Det
 ‘those four of them over there’

Noun phrases carry no case marking, with the syntactic function of a noun phrase being indicated only by its position in the sentence. Nor is there marking of gender, with the exception that for some speakers Portuguese loans can retain their grammatical gender, under conditions discussed in section 5.6. Number marking (discussed in sections 5.3.1 and 5.3.5) is optional.

Coordination of noun phrases is discussed in section 10.4.

5.2 Pronouns

The personal pronouns are listed in Table 5.1.

Table 5.1: Personal pronouns

Pronoun	Gloss	Description
<i>hau</i>	1S	‘I, me’
<i>o</i>	2S	‘you (singular)’
<i>ita</i>	2S.HON	‘you (singular respectful)’
<i>nia</i>	3S	‘he, she, him, her, it’
<i>ita</i>	1PI	‘we, us (including addressee)’
<i>ami</i>	1PE	‘we, us (excluding addressee)’
<i>imi</i>	2P	‘you (plural)’
<i>sira</i>	3P	‘they, them’

Ita is used both for inclusive first person plural (‘we’) and as a respectful second person singular pronoun (‘you’). Even greater respect and formality is shown by *ita bót* (lit. ‘2S.HON big’). *O* ‘2S’ is used mainly for children; addressing adults as *o* indicates either great familiarity, or that the speaker is emphasising their own higher social status.¹

There is a strong tendency to use titles or names rather than pronouns when addressing people, or when referring to someone who is present at the time. This is felt to be more polite than using the pronouns. So, for instance, *Senhora mai hori-bain-hira?* ‘Mrs come at.(past)-

¹ While Catholics use *ita* and *ita bót* to address God, Protestants tend to use *o*, following the European Protestant tradition of addressing God by the more intimate, less formal, choice of second person pronoun.

day-how.many' is a polite way to ask 'When did you arrive?', if one is addressing a woman whom one calls *Senhora*.

Some people consider that the third person pronouns should be used only for persons, others extend them to animals, while yet others accept them for inanimate entities also.² In practice it is far more common to refer to non-humans by a noun, or simply to omit the noun phrase altogether if it is 'understood'. The exception is that *nia* '3S' is commonly used for non-humans when making generic statements, as in (5.4).

- (5.4) *Kalan nia namlaik tiha ona.*
 night 3S wither PRF ANT
 'By night-time it (a certain type of flower) withers.'

All non-personal pronouns can also function as determiners, and so are discussed in the following section (5.3). Reflexive and reciprocal constructions are discussed in sections 7.5 and 7.4 respectively.

5.3 Determiners

5.3.1 Introduction

The determiners can be divided into several semantic classes: demonstratives, interrogatives, singular *ida* 'one, a', plural *sira* and quantifiers. These will be discussed in turn.

Many noun phrases have no determiner. This is the normal structure in generic statements (e.g. *Tilun hodi rona* 'ear for hear' = 'Ears are for hearing'). It is also usual in noun phrases referring to entities which are unique within the context (e.g. *prezidenti* '(the) president', *fulan* '(the) moon', *fulan Agustu* '(the) month (of) August', *hau nia aman* '1S POS father' = 'my father'), although *né* 'this' (discussed below) can be added to emphasise continuity of the topic or to emphasise proximity to the speaker.

When a noun phrase has no indication of number, it can be interpreted (in so far as context allows) as either singular or plural.

- (5.5) *Hau rona asu hatenu.*
 1S hear dog bark
 'I hear the dog barking.' or: 'I hear a dog barking.' or: 'I hear dogs barking.'

5.3.2 Demonstratives

The demonstratives are *né* 'this', and *nebá* 'there' (from *né bá* 'this go').³ When used deictically, they indicate that the referents are close to or far from the speaker respectively.

² *Nia* '3S' is not, as in the Fehan dialect of Tetun Terik (van Klinken 1999a:115), extended to referring to places, times or events.

³ A minority of speakers additionally use *nemai* 'here' (from *né mai* 'this come'), for indicating proximity.

They occur in the standard expressions *iha né* ‘LOC this’ = ‘here’ and *iha nebá* ‘LOC there’ = ‘there’. While *nebá* ‘there’ is mostly used in locative phrases (e.g. *iha Dare nebá* ‘LOC Dare there’ = ‘over in Dare’), some also use it for distant times (e.g. *momentu nebá* ‘moment there’ = ‘at that time’), contrasting with *né* for the present (e.g. *agora né* ‘now this’ = ‘nowadays’).

Né (but not *nebá*) is also used anaphorically, to refer back to earlier-mentioned entities. In this function it readily follows proper nouns, as in (5.6), personal pronouns (e.g. *nia né* ‘3S this’ = ‘he/she whom we are talking about’), and the plural definite marker *sira* (e.g. *buat sira né* ‘thing PL this’ = ‘these things that we are talking about’). Anaphoric *né* also readily refers back to earlier actions or propositions (e.g. *tanba né* ‘because.of this’, and example (5.7)).

- (5.6) *Ami hela hamutuk ho malae UNAMET, ida naran Jose, ida naran*
IPE stay together with foreigner UNAMET one name Jose one name

Gabriel, ida naran Robert. Senhor Jose né mak aruma surat.tahan...
Gabriel one name Robert mister Jose this FOC tidy paper

‘We stayed with the foreigners from UNAMET (the United Nations body which oversaw the 1999 referendum): one named Jose, one named Gabriel, (and) one named Robert. It was Mister Jose who was tidying up the papers...’

- (5.7) *Uluk eskola mós hanesan né.*
formerly school also like this
‘In the old days school was also like this (that is, using the current system which I have just explained to you).’

Less commonly, *né* marks noun phrases that are definite not due to proximal deixis or to direct anaphora, but because their existence can be inferred from the existence of other mentioned entities. This is illustrated by example (5.8), where the existence of the egg white and yolk follows from the previous mention of *mantolun* ‘egg’.⁴

- (5.8) *Ita tau tán man-tolun ida, ha-sai tiha nia mutin né*
IPI put as.well bird-egg one cause-exit PRF 3S.POS white this

soe tiha, tau deit nia Kinur né.
throw PRF put just 3S.POS Yellow this

(From a recipe:) ‘We add an egg: having removed the egg white (and) thrown it out, we just put in the yolk.’

Some sentence-initial topics are marked by final *né*, with rising intonation and a following pause, as in (5.9). These include some subjects, fronted objects, and other fronted constituents (see sections 7.1.3 and 10.2). In casual speech (but not in formal speech or writing), such topics can have a sequence of one of the demonstratives (usually *né*) followed

⁴ This is referred to by Prince (1981) as ‘inferred’, and by Hawkins (1978:99) as ‘associative’ anaphora. Himmelmann (1996) points out that this function is characteristic of definite articles rather than demonstratives. It thus appears that *né* is part-way along the path from being a full demonstrative to becoming a definite article.

by a topic-marking *né*. The initial demonstrative is either deictic (as in (5.10) and *momentu nebá né* ‘time that this’ = ‘that distant time I am talking about’), or anaphoric, as in (5.11). The final *né* has rising intonation and is followed by a pause.

- (5.9) *Aimoruk né, sempre hetan reseita ka?*
 medicine this always get prescription or
 ‘With regard to medicine, do you always get a prescription?’
- (5.10) *Iha né né, karun la-halimar.*
 LOC this this expensive not-play
 (In the village we raise our own chickens.) ‘Here (in Dili, they) are terribly expensive.’
- (5.11) *Hau nia avó né né, nia...*
 1S POS grandparent this this 3S
 (Returning to earlier main character of discussion:) ‘My grandfather, he...’

There also appears to be an utterance-final *né* of frustration or disapproval, which mainly occurs in questions in informal interactions. This use is illustrated by the following example.

- (5.12) *Imi bá vadiu iha nebé mak foin mai né?*
 2P go hang.around LOC where FOC JUST come this
 ‘Where did you go wandering off to that you’ve only just come (home now)?’

5.3.3 Interrogatives

The interrogative pronouns are *sé* ‘who’, *nebé* ‘where, which’, *saida* ‘what’ and *sá* ‘what’. These are further discussed in 7.6.1. Of these, *nebé* ‘which’ (e.g. *kareta nebé* ‘vehicle which’ = ‘which vehicle’), and *saida* ‘what’ (e.g. *moras saida* ‘sick what’ = ‘what illness’) can function as determiners. *Sá* is idiosyncratic in that when it modifies a noun phrase, it must precede the noun; it then requests a selection from a limited set of options (*sá kor* ‘which colour’, not **kor sá*⁵). This construction is however not used by all speakers, with some preferring post-modifying *saida* or *nebé*.⁶

⁵ In the common question *Nia naran sá?* ‘3S name what’ = ‘What is his/her name?’, the *sá* is a pronominal predicate and not a determiner modifying *naran* (see section 7.3).

⁶ The interrogative pronouns can also be used indefinitely, as in Tetun Terik, although such usage is relatively rare. In this case the pronoun is often followed by *deit* ‘just’ (e.g. *sé deit* ‘who just’ = ‘whoever’; *Kuda saida deit, moris* ‘plant what just, live’ = ‘(You can) plant anything, (and it will) live’). Informally, *saida* ‘what’ can even be used as a verb (e.g. *Ami baku, ami saida* ‘1PE hit, 1PE what’ = ‘We hit (the children), or we do whatever (to them)’).

5.3.4 *ida* ‘one, a’

Ida ‘one, a’ is discussed here since it has a somewhat different distribution to other numerals. When modifying a noun phrase, it is often better translated as an indefinite article, since it commonly marks noun phrases with ‘new’ information. It is however also used for marking noun phrases with old information as being singular, as illustrated by the common sequence *ida né* ‘one this’ (e.g. *liurai ida né* ‘ruler one this’ = ‘this ruler’).

Ida may either precede or follow adjectival modifiers within a noun phrase. The adjective precedes *ida* if the noun–adjective pair represents a single idea, such as in compounds (e.g. *ema-bót ida* ‘person-big one’ = ‘an important person, VIP’, *língua ofisiál ida* ‘language official one’ = ‘an official language’). Most other adjective phrases consisting of a single adjective also precede *ida*. In contrast, multi-word adjective phrases nearly always follow *ida*, as in (5.13). This position places greater emphasis on the adjective, and so is appropriate, for instance, if a contrast is being made.⁷

- (5.13) *milagre ida bót tebe-tebes*
 miracle one big RDP-truly
 ‘a truly great miracle’

Although *ida* is rarely used with a numeral classifier, if it does have a classifier (thus forcing a numeric interpretation), it must follow the adjective, just as other numerals do. Unlike other numerals, *ida* frequently heads a noun phrase (e.g. *ida né* ‘one this’ = ‘this one’).

5.3.5 *sira* ‘PLURAL’ and Portuguese plurals

Sira is both the third person plural pronoun (‘they’) and a definite plural marker (5.14).

- (5.14) *Oan sira la iha.*
 child PL not EXIST
 ‘The children aren’t here. (They are all out.)’

As a plural marker *sira* is not restricted to humans (e.g. *buat sira né* ‘thing PL this’ = ‘the aforementioned things’), and is definite. Thus *profesora sira* ‘teacher.FEM PL’ = ‘the female teachers’ implies that all the female teachers under consideration are meant. It is not common to use *sira* with other markers of plurality, and some combinations of plural modifiers, such as of *sira* with numerals, are considered ungrammatical (e.g. **kadernu rua sira* ‘notebook two PL’). After a noun identifying one person, *sira* indicates ‘and associated persons’ (e.g. *João sira* ‘John and the people associated with him, such as his family’).

Some speakers use Portuguese plural forms of Portuguese nouns (e.g. *alunu-s* ‘student-PL’); this is particularly common in formal, heavily Portuguese-influenced contexts such as television and radio news broadcasts. At times one also hears the Portuguese plural and the Tetun definite plural used together (e.g. *refujiadu-s sira* ‘refugee-PL PL’ = ‘the refugees’).

⁷ The differences between the two orders are thus very similar to those between a simple adjective and an adjectival relative clause in Indonesian (Sneddon 1996:146–7). Adjectives following *ida* tend to be translated into Indonesian by a *yang* relative clause, and there is nothing to prevent them being analysed as a type of relative clause in Tetun (see section 10.8).

Some people however disapprove of using Portuguese plural forms, and many more disapprove of mixing of the two systems.

5.3.6 Quantifiers

Hotu ‘all’ has three meanings; as a quantifier it means ‘all’, as an adverb it means ‘also’, and as an intransitive verb it means ‘be finished’ (see section 8.2.2.4). When *hotu* ‘all’ quantifies the subject (5.15), it optionally floats to immediately following the verb (5.16), particularly if the verb is intransitive or has an object that cannot be quantified (so that *hotu* cannot be interpreted as quantifying the object). Otherwise *hotu* is usually the final element in the noun phrase (e.g. *povu sira hotu* ‘the.people/commoner PL all’ = ‘all the common people’).

(5.15) *Ami hotu hadér.*
 1PE all/also get.up
 ‘We all got up.’ or: ‘We too got up.’

(5.16) *Ami halai hotu.*
 1PE run all
 ‘We all ran away.’

When *hotu* quantifies the object, it optionally fills that same post-verbal slot (5.17).⁸ Alternatively, *hotu* ‘all’ post-modifies the object (e.g. *halo buat hotu* ‘do thing all’ = ‘do everything’).

(5.17) *Sira ha-tún hotu sasán.*
 3P cause-descend all goods
 ‘They lowered all the goods (from the top of the truck).’

The reduplicated form *hotu-hotu* (or *hot-hotu*) ‘all, truly all’ can modify or (less commonly) head a noun phrase. Unlike *hotu* it cannot float to post-verbal position, and does not have the alternative meanings of ‘also’ and ‘finished’.

(5.18) *Agora sira hotu-hotu osan la iha.*
 now 3P RDP-all money not have
 ‘Now they all have no money.’

Balu (for some speakers *balun*) functions as a determiner meaning ‘some (of)’ (e.g. *ema balu* ‘person some’ = ‘some of (the) people’) and as a pronoun ‘some’. In some contexts its interpretation defaults to ‘half’ (e.g. *fulan ida ho balu* ‘month one and half’ = ‘one and a half months’). It is used almost exclusively of count nouns.

There are in addition some quantifying terms which are not determiners. These include the quantifying adjectives (and adverbs) *barak* ‘many, much’, *uitoan* ‘a few, a little’ (e.g. *ema*

⁸ In this slot *hotu* can, with the appropriate tense–aspect, alternatively be interpreted as ‘finished’ (e.g. *hán hotu tiha* ‘eat finish PRF’ = ‘after eating’).

uitoan ‘person a.little’ = ‘a few people’, *fós uitoan* ‘uncooked.rice a.little’ = ‘a little uncooked rice’) and *tomak* ‘whole, entire, all’ (e.g. *Timór tomak* ‘Timor whole’ = ‘the whole of Timor’, *Timór-oan tomak* ‘Timor-native.of whole’ = ‘all Timorese’), the compounded numeral *ida-rua* (lit. ‘one-two’) ‘a few’, and, for some speakers (as well as the church register), an indefinite numeral *hira* ‘several’.

5.3.7 Other

The quantifying pronoun and determiner *ida-idak* ‘each’ is used almost exclusively in the context of possession, with each referent having their own possession (5.19).

(5.19) ... *tuir ita ida-idak nia fiar.*

follow 1PI each POS believe

‘(He tells us to pray) each according to our own faith.’

Hirak ‘certain’ is always used with a following definite determiner (*né* ‘this’, *nebá* ‘there’) or a relative clause. While it is mostly used in noun phrases with non-human referents (such as time periods, words, or situations), it can also be used with human referents. *Hirak* indicates that the referents form a group, whose members are not enumerated, but could in principle be known (i.e. it is ‘specific’). It is used far more commonly in writing than in speaking.⁹

(5.20) *Distritu hirak né hanesan representante ba distritu sanulu resin tolu.*

district certain this like representative to district ten extra three

‘These districts (which the writer had just listed) are representative of the thirteen districts.’

Ruma ‘some or other’ functions as noun phrase modifier only, and indicates an unknown or unspecified quantity or type. It occurs frequently in the standard phrases *dala ruma* ‘occasion some’ = ‘sometimes’, *ema ruma* ‘person some’ = ‘someone, somebody’, *fatin ruma* ‘place some’ = ‘somewhere’, and *buat ruma* ‘thing some’ = ‘something’.

(5.21) *Nia bele bót bá mós, nia halo sala ruma, ita tenki hanorin nia.*

3P can big go also 3S do wrong some 1PI must teach 3S

‘Even when she gets older, (if) she does something-or-other wrong, we must instruct her.’

Seluk ‘other, another’ on its own is indefinite, and neutral with respect to number (e.g. *ema seluk* ‘person other’ = ‘other people, someone else’). The combination *sira seluk* ‘PL other’ is definite plural ‘the others’ (e.g. *labarik sira seluk* ‘child PL other’ = ‘the other children’).

⁹ Our analysis here is slightly different to Hull’s (1999), who glosses *hirak* as ‘a few, several’ and *hirak né* as ‘these kinds of, such’.

5.4 Possessive and associative constructions

5.4.1 Overview

Tetun has two orders of possessor and possessum, as follows.

Possessor	((<i>ni</i>) <i>nia</i>)	Possessum
Possessum	Possessor	((<i>ni</i>) <i>nian</i>)

That is, the possessor is followed by the possessive marker (*ni*)*nia*¹⁰ when it precedes the possessum (e.g. *João nia uma* ‘John POS house’ = ‘John’s house’), and by (*ni*)*nian* when it follows the possessum (e.g. *uma João nian* ‘house John POS’ = ‘John’s house’). The form (*ni*)*nian* is also used when the possessum functions as predicate, as in (5.22), or as a headless noun phrase, as in (5.23).¹¹

(5.22) *Uma né João nian.*
house this John POS
‘This house is John’s.’

(5.23) *João nian iha ne bá.*
John POS LOC there
‘John’s (house...) is over there.’

After pronominal possessors, the short forms *nia* and *nian* must be used (e.g. *hau nia inan* ‘1S POS mother’ = ‘my mother’, *Né sé nian?* ‘this who POS’ = ‘Whose is this?’), with the exception that the 3S possessive form is obligatorily reduced from **nia nia* to *ninia* or (more commonly) *nia*, as in *nia oan* ‘3S.POS child’ = ‘his/her child’.¹² Note that this *nia* is not restricted to human referents as the personal pronoun *nia* tends to be (e.g. *nia ikun* ‘3S.POS tail’ = ‘its tail’, *nia rohan* ‘3S.POS edge’ = ‘its edge’).

After a non-pronominal possessor, either the shorter *nia/nian* or the longer *ninia/ninian* is acceptable; however in speaking the shorter form is far more common than the longer form, with the longer form mostly restricted to formal registers (e.g. *mentalidadi povu ninian* ‘mentality the.people POS’ = ‘mentality of the (common) people’). When non-pronominal possessors are marked as plural by *sira*, the *sira* precedes the possessive marking (e.g. *feto sira nia direitu* ‘woman PL POS right’ = ‘women’s rights’).

¹⁰ The *nia* is not colloquially shortened to *ni*, as is the case in Tetun Terik, and reported by Hull (1996b:59).

¹¹ This distinction is not rigidly maintained, particularly in writing, with some writers adding a final *-n* even when there is a following possessum (e.g. *ita nian lian* ‘1PI POS language’ = ‘our language’). Nevertheless, many speakers consider the addition of *-n* in this construction to be unacceptable.

¹² Liturgical Tetun follows Tetun Terik in not requiring *nia* after pronouns. This results in a contrast between everyday *ami nia aman* ‘1PE POS father’ = ‘our father (Dad)’ and liturgical *ami Aman* ‘1PE Father’ = ‘our Father (God)’.

There is a common construction in which a subject pronoun is apparently followed by a possessum (e.g. *Hau isin diak* ‘1S body good’ = ‘I am well’). This can however be analysed very differently, as a subject pronoun followed by a complex predicate. This construction is discussed in section 7.2.

There are restricted conditions under which the possessive linker is not required; see section 5.4.2 below for details.

5.4.2 Order of possessor and possessum

While there is some overlap in when the two orders of possessor and possessum can be used, in many cases one order is either required or preferred.

The first factor determining order is alienability, with inalienable possession requiring the possessum-final order. Inalienable possession in Tetun Dili includes the following semantic relationships:

1. Part-whole relationships. Many of these are expressible as compounds, with no possessive linker (e.g. *ibun kulit* ‘mouth skin’ = ‘lip’; see section 4.4.2). However parts of individual entities are expressed using the possessive marker (e.g. *ai né nia funan* ‘plant this POS flower’ = ‘this plant’s flower’). This includes body parts (e.g. *o nia án* ‘2S POS self’ = ‘your self’, *hau nia liman* ‘1S POS hand’ = ‘my hand’).
2. Location relative to an entity. The possessive marker (*ni*)*nia* is usually omitted if the possessor is an unmodified noun that does not refer to a human (e.g. *iha uma laran* ‘LOC house inside’ = ‘in the house’). The marker is however used when the possessor refers to a human (5.24) or is itself modified (5.25). For a list of common location nouns, see section 6.2.

(5.24) *iha João nia kotuk*
 LOC John POS back
 ‘behind John’

(5.25) *iha bélihun bót nia ninin*
 LOC pool large POS edge
 ‘at the edge of the large pool (of water)’

3. Kin relationships (e.g. *João nia inan* ‘John POS mother’ = ‘John’s mother’), with the exception of *família* ‘family’, which allows either order (e.g. *família Ana Maria nian* or *Ana Maria nia família* ‘the family of Ana Maria’).
4. naran ‘name’ (e.g. *tiu nia naran* ‘uncle POS name’ = ‘uncle’s name’).

When Tetun deverbal nouns are possessed, they similarly require that the possessor come first (e.g. *hau nia hanoin* ‘1S POS think’ = ‘my thoughts/opinions’). There is no overt nominalisation of the verb, and the possessor corresponds to the verb’s subject (so *Amérika nia tulun* ‘America POS help’ means ‘America’s assistance (to someone)’, not ‘(someone’s) assistance to America’). The underlying verb’s object, on those rare occasions when it is specified, is introduced by *ba* ‘to’ (5.26).

(5.26) *hau nia fiar ba Nai Maromak*
 1S POS believe to Lord God
 ‘my faith in God’

A relationship of physical possession is usually presented with the possessum-final order (e.g. *nia livru* '3S.POS book' = 'his/her book'), but the reverse is also acceptable (e.g. *kareta UNAMET nian* 'vehicle UNAMET POS' = 'UNAMET vehicle').

The second factor affecting order is the nature of the possessor. In everyday speech, pronominal possessors virtually always precede the possessum, and are often judged to be odd in the reverse order (e.g. *hau nia uma* '1S POS house' = 'my house'; *?uma hau nian*). Non-pronominal possessors referring to particular individuals also strongly prefer this order (e.g. *Maria nia kareta* 'Maria POS vehicle' = 'Maria's vehicle'). In the press, however, it is not unusual to have pronominal post-possessors (e.g. *ajuda sira nian* 'help 3P POS' = 'their help').

The alternative order, in which the possessum precedes the possessor, usually expresses a more general relationship of association between the possessor and possessum (e.g. *direitu feto nian* 'right woman POS' = 'women's rights', *reseita hahán nian* 'recipe/prescription food POS' = 'recipe', *feto Timór Loro-sae nian* 'woman Timor Sun-ascend POS' = 'the women of East Timor'). In this order, the term filling the 'possessor' slot is often descriptive rather than referring to a particular entity, and the relationship is not strictly one of possession. Such 'possessors' can be adjectival (e.g. *istoria uluk nian* 'story former POS' = 'old story/story about the past') or even a verb phrase (e.g. *tempu silu batar nian* 'time snap maize POS' = 'the time for harvesting maize').

This contrast between a possessum-final order (marking a possessive relationship), and a possessor-final order (marking an associative relationship) is shown clearly in the contrast between *militar nia sasán* 'military POS goods' which means 'the military's goods' (everything that the military actually owns, even including stationery), and *sasán militar nian* which refers to 'military goods' (everything normally associated with the military, even if actually owned by a civilian).

The possessor-final order is common in formal and fixed phrases (as noted by Hull (1996b:84)), such as from the fields of politics, law and economics (e.g. *lei Portugal nian* 'law Portugal POS' = 'Portuguese law', *Dioseze Dili nian* 'diocese Dili POS' = 'the Diocese of Dili'). This preference is perhaps partly a matter of register, with formal registers appearing to prefer possessor-final constructions.

In the possessor-final construction, the possessive marker (*ni*)*nian* is often omissible, resulting in the noun-modifier construction discussed in section 5.5. A contributing factor is that *nian* tends to be omitted when the referent is unique (e.g. *estadu Japaun* 'the Japanese government', *prezidenti CNRT* 'the president of CNRT'; cf. *visi-prezidenti CNRT nian* 'the vice president(s) of CNRT'). It is also more readily omissible in informal registers.

Where both orders are acceptable, the difference is sometimes a matter of focus. For instance, *samea nia beneno* 'snake POS poison' is the normal and strongly preferred order to refer to snake venom. The reverse (*beneno samea nian*) is however possible if the topic of discussion is *beneno* 'poisons', and one is contrasting snake venom with other poisons.

5.4.3 Combination of two 'possessors'

It is possible for a single noun phrase to include both a preceding and a following 'possessor', in which case the former designates a true possessor while the latter indicates a more general association.

- (5.27) *sira nia lei relijiaun nian*
 3P POS law religion POS
 ‘their religious law’

5.4.4 Genitive head-marking

In the conservative Fehan dialect of Tetun Terik, all vowel-final kin and body part terms take a genitive *-n* clitic when they have a preceding possessor (e.g. Fehan *ni ina-n* ‘3S.POS mother-GENITIVE’ = ‘her mother’) (van Klinken 1999a:142ff). All such kin terms have an invariant final /n/ in Tetun Dili (e.g. *inan* ‘mother’, *maun* ‘older brother’), as do the body part nouns (e.g. *kabun* ‘stomach’), suggesting that the genitive clitic has become fossilised with nouns that are more usually inalienably possessed.

There are two vowel-final kin terms borrowed from Portuguese which (for most, but not all speakers) take a final genitive *-n* when they are possessed, although not in any other construction. These are *tia* ‘aunt’ and *tiu* ‘uncle’ (e.g. *hau nia tiu-n* ‘1S POS uncle-GENITIVE’ = ‘my uncle’; cf. *tiu ida* ‘uncle one’ = ‘an uncle’). No other vowel-final Portuguese loans take genitive *-n* (e.g. *hau nia sobrinha* ‘1S POS niece’ = ‘my niece’).

Some speakers, particularly those with roots in the south of Timor, also apply genitive *-n* to possessed *rai* ‘land, country’ when they mean ‘country’ (e.g. *hau nia rai-n* ‘1S POS country-GENITIVE’ = ‘my country’ (e.g. Timor)), but not when they mean ‘land’ (cf. *hau nia rai* ‘1S POS land’ = ‘my land’ (e.g. the hectare I own)). Others consider the genitive *-n* here to be Tetun Terik, and do not use it.¹³

5.5 Noun phrase as modifier

A noun phrase modifying a noun can have a wide range of semantic relationships to it. One of the more common is specifying source location (e.g. *ema Los Palos* ‘person (from) Los Palos’). For some relationships, there is an alternative formulation in which the noun phrase is followed by possessive *nian* (section 5.4.2). Note that there are also many two-word noun–modifier sequences which are lexicalised as compounds (section 4.4.2).

5.6 Adjective phrase

Adjective phrases functioning as modifiers usually consist of a single adjective (e.g. *uma bót ida* ‘house big one’ = ‘a big house’). They can however also be intensified (e.g. *independénsia lo-lós liu* ‘independence RDP-true more’ = ‘more genuine independence’). See section 8.5 for a full list of intensifiers, as well as for comparatives and superlatives, and section 5.3.4 for a discussion of the order of adjectival modifiers and *ida* ‘one’.

¹³ The word for ‘language’ is another contender for genitive marking. All speakers agree that when it has a preceding possessor, there must be a final *-n* (e.g. *hau nia lian* ‘1S POS language’ = ‘my language’). However when it precedes a language name, some people use *lia* (as per Tetun Terik), and some *lian* (indicating that for them the originally genitive *-n* has become part of the noun).

Adjectives borrowed from Portuguese are borrowed in the singular masculine form, with two exceptions. Firstly, masculine *bonitu* ‘handsome’ describes males, while feminine *bonita* ‘pretty’ describes females. Secondly, where whole noun phrases are borrowed from Portuguese, these retain both the Portuguese order (with the adjective generally preceding the noun), and the Portuguese gender and number agreement on the adjective (e.g. *Estadu-s Unidu-s* ‘state.MASC-PL united.MASC-PL’ = ‘United States’, *primeiru ministru* ‘first.MASC minister.MASC’ = ‘prime minister (male)’, *primeira klase* ‘first.FEM class.FEM’ = ‘first grade (of school)’). In contrast, when borrowed adjectives modify Tetun nouns, they use the Tetun word order and the masculine form of the adjective (e.g. *loron primeiru* ‘day first’ = ‘first day’).¹⁴

The quantifying adjective *barak* ‘many, much’ is unusual in that when it quantifies the subject of an intransitive verb it can follow the verb (like *hotu* ‘all’, discussed in section 5.3.6) (5.28). Otherwise it follows the noun which it modifies (*osan barak* ‘money much’ = ‘much money’).

- (5.28) *Refujiadu-s sira mate barak iha Atambua.*
 refugee-PL PL die many LOC Atambua
 ‘Many refugees died in Atambua.’

5.7 Numeral phrase

5.7.1 Numerals

Tetun has a decimal numeral system. The basic numerals are as listed in Table 5.2. For zero and for numerals beyond a thousand, Portuguese or Indonesian numerals are used.¹⁵

¹⁴ Some speakers allow Portuguese ordinals to follow a Portuguese noun in accordance with normal word order rules in Tetun (e.g. *faze terseiru* ‘phase third’ = ‘third phase’), in which case they prefer the masculine ending. Others disallow this order.

In formal contexts, such as reading the news on television, some speakers retain number agreement on Portuguese adjective–noun pairs even when these are not fixed phrases (e.g. *partidu-s tradisionais* ‘party-PL traditional.PL’ = ‘traditional (political) parties’). This can lead to puzzlement on the part of Tetun speakers who have no Portuguese education.

¹⁵ Some Tetun dictionaries do list larger numbers, but these are not known by the vast majority of people, and in any case different sources disagree on their value. The best known, *tokon* is listed by Hull (1999) as a million, and is used with this value in some written materials; nevertheless Costa (2000) lists it as a hundred thousand, while most people only know it in the expression *tokon ba tokon* ‘numerous, zillions’.

Table 5.2: Numerals

N numeral	Gloss
<i>ida</i>	1
<i>rua</i>	2
<i>tolu</i>	3
<i>hát</i>	4
<i>lima</i>	5
<i>nén</i>	6
<i>hitu</i>	7
<i>walu</i>	8
<i>sia</i>	9
<i>sanulu</i>	10
<i>atus</i>	100
<i>rihun</i>	1000
<i>hira</i>	'how much?, how many?' ¹⁶

The formula for numerals from ten to just under a million is as follows, where at least one of the first three bracketings (i.e. for thousands, hundreds or tens) must be specified. In the formula 'Digit' covers the numerals from 1 to 9. The examples following the formula are lined up in columns for easier comparison.

(<i>rihun</i> Numeral) '1000'	(<i>atus</i> Digit) '100'	(Digit- <i>nulu</i>) '10s'	(<i>resin</i> Digit) 'extra'	
<i>rihun sanulu resin nén</i>				'16,000'
<i>rihun ida</i>	<i>atus sia</i>	<i>sia-nulu</i>	<i>resin tolu</i>	'1993'
	<i>atus walu</i>			'800'
	<i>atus ida</i>	<i>rua-nulu</i>	<i>resin lima</i>	'125'
		<i>sanulu</i>	<i>resin rua</i>	'12'

All numerals are regular according to the above formula, except that 'ten' is expressed by *sanulu*, and not **ida-nulu*.

Tetun numerals tend to be used only for small numbers, such as for the number of children in a family, one's age, or the time.¹⁷ Dates, prices and arithmetic are much more often given in Portuguese or Indonesian than in Tetun, while time is commonly specified in any of these three languages. While even children can count in all three languages, most Indonesian-

¹⁶ *Hira* is classed as a numeral because, unlike other quantifiers, it takes a numeral classifier when quantifying a human noun phrase (e.g. *ema nain hira?* 'person CLS:human how.many' = 'how many people?'; contrast *ema barak* 'person many' = 'many people').

Some, but not all, Dili speakers also use *hira* non-interrogatively as an indefinite 'several', as is the case in Tetun Terik and church Tetun.

¹⁷ The fact that Tetun is seldom used for large numbers is illustrated by the observation that even children who frequently handle a thousand rupiahs tend to be unsure of the value of Tetun *rihun* 'thousand'. Many adults have difficulty specifying years or prices in Tetun. However some now appear to be making a conscious effort to use Tetun numerals in these areas too.

educated people are comfortable only with Indonesian once the numbers reach a million or more.¹⁸

Tetun numerals follow the noun they modify (e.g. *uma rua* 'house two' = 'two houses'). Portuguese numerals modify only Portuguese nouns, in which case they follow the Portuguese order of having the numeral first. This construction is mainly found in telling the time, where Portuguese phrases such as *dez minutu-s* 'ten minute-PL' = 'ten minutes' are used interchangeably with Tetun *minutu sanulu* 'minute ten' and Indonesian *sepuluh menit* 'ten minute'. Similarly, Portuguese *duas ora-s* 'two.FEM hour.FEM-PL' = 'two o'clock' is equivalent to Tetun *tuku rua* 'o'clock two' or Indonesian *jam dua* 'o'clock two'.¹⁹

There are no Tetun ordinal numbers. For low ordinal numbers, Portuguese borrowings are sometimes used (e.g. *primeiru* 'first', *segundu* 'second', *terseiru* 'third'). An alternative is to use *númeru* 'number' (or Indonesian *nomor*) followed by a numeral (e.g. *númeru rua* 'number two' = 'second').

Approximation is expressed by a variety of means: a preceding *maiz ou menus* 'more or less' or *kala* 'perhaps'; a following *hanesan né* 'like this' or *resin* 'extra' (e.g. *sanulu resin* 'ten extra' = 'more than ten'); or by listing numerals in the approximate range (including by the fixed expression *ida-rua* 'one-two' = 'a few'; *tinan rua ka tolu* 'year two or three' = 'two or three years'). In addition there are the expressions *atus ba atus* 'hundred to hundred' = 'hundreds' and *rihun ba rihun* 'thousand to thousand' = 'thousands'.

5.7.2 Numeral classifiers

When enumerating humans it is preferable but not obligatory to use the human classifier *nain* (also a noun meaning 'owner') before numerals greater than one (e.g. *feto nain rua* 'woman CLS:human two' = 'two women', *ita nain tolu* 'we CLS:human three' = 'we three'). A classifier is rarely used with *ida* 'one', being permissible only if stressing that *ida* is 'one' (as opposed to some other numeral) and not indefinite 'a'. It is seldom used with large numbers such as a hundred.

The other classifiers in moderately common use are listed in Table 5.3, along with their primary meanings as nouns. These terms classify the entities by shape. Note that for the vast majority of nouns (e.g. for counting animals, tables or books) there is no classifier.

¹⁸ This creates significant problems for United Nations and other materials translated into Tetun, which tend to use either Portuguese *milhaun* or the alleged Tetun *tokon* for 'million'. Many people don't know either term, with some youth incorrectly equating *milhaun* with Indonesian *milyár* 'billion'.

¹⁹ The noun *oras* 'hour' (borrowed from the Portuguese plural *hora-s* 'hour-PL'), when used with a Tetun numeral, specifies a number of hours (e.g. *oras tolu* 'hour three' = 'three hours'), while with a preceding Portuguese numeral it specifies the time (e.g. *tréz oras* 'three o'clock').

Table 5.3: Numeral classifiers

Classifier	Noun meaning	Classifier for	Example
<i>lolon</i>	'trunk, stem'	long cylindrical objects	<i>sigaru</i> 'cigarette'
<i>tahan</i>	'leaf'	thin flat objects	<i>surat</i> 'card'
<i>fuan</i>	'fruit, heart'	whole roundish objects	<i>hás</i> 'mango'
			<i>paun</i> 'bread (roll/loaf)'
<i>musan</i>	'seed'	very small round objects	<i>aimoruk</i> 'tablet'
			<i>fatuk</i> 'pebble'

5.8 Complements of nouns

Some abstract nouns borrowed from Portuguese can take a complement clause or verb phrase. For some nouns (e.g. *oportunidadi* 'opportunity', *lisensa* 'permission', *direitu* 'right (to do...)', the complement is introduced by irrealis *atu*, *para* 'so that', or *para atu* (5.29). Other nouns (e.g. *opiniaun* 'opinion') take sentential complements introduced by *katak* 'that'. Apart from requiring an introductory *atu* or complementiser, these complements have the same structure as complements of verbs (see section 10.5). This construction is found more in writing than in speaking.

- (5.29) *Sira iha vontadi para atu aprende.*
 3P have desire so.that IRR learn
 'They have a desire to learn.'

6. Prepositions

6.1 Introduction

Prepositional phrases consist of a preposition followed by a noun phrase. They can function as oblique arguments, temporal or locational modifiers of clauses, modifiers of noun phrases, or even clausal predicates. Prepositions are an area of considerable inter-speaker variation and substantial influence from Portuguese, as reflected in the following discussion.¹

Some words which are analysed by other writers as prepositions (e.g. *hakat* ‘cross, across’ in Hull (1999)) are analysed by us as verbs occurring in serial verb constructions. These are presented in section 9.3.

6.2 *iha* ‘LOC’

The locative preposition *iha* is both very common and general in its meaning, being translatable according to context as ‘in, at, on, from, to’. It is thus assigned the general gloss ‘LOC’. This preposition commonly introduces oblique locative arguments (e.g. *bá iha Dili* ‘go LOC Dili’ = ‘go to Dili’), peripheral location (6.1) and temporal phrases (6.1).

- (6.1) *Hau moris iha Zumalai iha dia déz de Abril,...*
1S born LOC Zumalai LOC day ten of April
‘I was born in Zumalai on the tenth of April, (1975).’

Location within or relative to an entity is expressed by a possessed location noun (e.g. *iha hau nia kotuk* ‘LOC 1S POS back’ = ‘behind me’). The more common nouns used in this slot are listed in Table 6.1, although there are inter-speaker and inter-regional differences in the use and interpretation of some of these terms. Note that some of these location terms are also body part names.² There are in addition some location terms which are rather more restricted in their context, such as *iha...tutun* ‘at the summit (of mountain), point, end’ and *iha...hún* ‘at the base (of a plant), foot (of a mountain)’.

¹ There is no borrowing of prepositions or conjunctions from Indonesian.

² The use of the same terms for body parts and locatives is of course common, as demonstrated by Bowden (1992) for Oceanic languages.

Table 6.1: Location nouns

Location phrase	Location phrase meaning	Noun meaning
<i>iha...oin</i>	'in front of'	'face'
<i>iha...kotuk</i>	'behind'	'back'
<i>iha...laran</i> ³	'in, within, inside; amongst; during (time)'	'the seat of emotions, thoughts, and character'
<i>iha...klaran</i>	'in the middle of; between'	
<i>iha...lét</i>	'between; among, amongst'	
<i>iha...sorin</i>	'beside'	'side'
<i>iha...sorin bá</i>	'on the other side of'	'side go'
<i>iha...sorin mai/né</i>	'on this side of'	'side come/this'
<i>iha...sorin/liman lós</i>	'on the right side of'	'side/hand right'
<i>iha...sorin/liman karuk</i>	'on the left side of'	'side/hand left'
<i>iha...leten</i>	'on, on top of, above'	'upper'
<i>iha...kraik</i> ⁴	'at the base of; under, below'	'lower'
<i>iha...okos</i>	'underneath, under, below'	
<i>iha...ninin</i>	'at the edge of'	'edge'

When the possessor noun phrase consists of a single non-human noun, the location is usually immediately juxtaposed to the possessor noun, without an intervening possessive *nia* (e.g. *iha Dili laran* 'LOC Dili inside' = 'inside Dili').⁵ The relationship is explicitly marked as possessive by *nia* when the possessor refers to a person (e.g. *iha hau nia sorin* 'LOC 1S POS side' = 'beside me', *iha João nia oin* 'LOC John POS face' = 'in front of John'),⁶ or when the possessor is itself modified (e.g. *iha kadeira rua nia klaran* 'LOC chair two POS middle' = 'between the two chairs'). The location nouns can also be used without a possessor; in this case there is no indication of what the location is relative to (e.g. *iha kotuk* 'LOC back' = 'at the back').

The use of these location terms is not restricted to complements of *iha*, being found, for instance, also in complements of the preposition *hosi* 'from', and in objects of verbs of motion (e.g. *tama uma laran* 'enter house inside' = 'enter the house'). There is thus no reason to analyse these location terms as postpositions (or as forming complex prepositions with *iha*), as is the case in analyses of some other Austronesian languages (Bowden 1992:45).

³ The opposing concept 'outside (the house)' can be expressed by *iha liur* 'LOC outside'. However many speakers cannot use the construction **iha...liur* to mean 'outside of...'; instead, *iha xikra nia liur* 'LOC cup POS outside' would mean 'the outside surface of the cup', while *iha odamatan liur* 'LOC door outside' would mean 'at (the) outside door' (with *liur* as a modifier).

⁴ This expression is disputed, with some claiming it means 'at the base of...', others 'under, below...', and others not accepting *iha...kraik* at all with a possessor noun. (Note that all accept *iha kraik*, with no possessor noun, as meaning 'down below, lower down, downhill'.) It seems that those who do accept possessed *kraik* require an explicitly possessive *nia* (e.g. *iha meja nia kraik* 'LOC table POS lower' = 'under the table').

⁵ The result is a noun-noun sequence as found in compounds (see section 4.4), and indeed a few are perceived as compounds by native speakers (e.g. *iha ai-laran* 'LOC wood/plant-inside' = 'in the jungle').

⁶ This rule does not apply in the church register, with the published gospels frequently defining location relative to people without using *nia* (e.g. *iha Ita sorin kuana* 'LOC 2S.HON side right' = 'on Your right side').

Note that the location terms are only used of entities that are perceived to have an inherent front and back, such as people, houses and televisions. One cannot (as in English) use these expressions to specify a location relative to other types of entities, such as trees. The left and right sides too are determined relative to the entity itself; hence the left side of the house is that which is to one's left if one is inside the house facing the front.

As shown in (6.1) above, *iha* can be used to introduce temporal phrases, as well as locative ones. This is true also of the expression *iha...nia laran* 'in, within, inside; amongst' which is also used with the temporal meaning of 'during' (e.g. *iha Dezembru nia laran* 'LOC December POS time' = 'in/during December'). This meaning can also be expressed with the use of the preposition *duranti* 'for (period of time), during, when', discussed in section 6.6.3.

6.3 *ba, mai* 'to (location, person), for (person)'

The terms *bá/ba* 'go/to' and *mai* 'come' are largely parallel. Both are verbs, deictic particles, and prepositions. While as a preposition *ba* is not stressable, as a verb and deictic it can be stressed, and hence carries an acute accent ('*bá*') in this book.⁷

When used as verbs or as deictic particles (see section 9.3), *bá* and *mai* are clearly distinguished from one another, with *mai* indicating movement towards where the speaker is now, and *bá* indicating movement away, or along some other plane.⁸ The goal of both verbs is either presented as a direct object (e.g. *bá Dili* 'go to Dili') or introduced by *iha* (e.g. *bá iha Dili* 'go LOC Dili' = 'go to Dili').

As prepositions, *ba* and *mai* introduce goal location (6.2), recipient (6.3), addressee (6.4), and beneficiary (6.5) noun phrases. In this function, however, the semantic distinction between *ba* and *mai* is not as rigidly maintained. As with the verbs, *ba* is always used when the complement noun phrase is second or third person, while complements of *mai* are restricted to those which include the speaker (e.g. *hau* '1S', *ita* '1PI', and *ami* '1PE'). Nevertheless, it is not uncommon to use *ba* instead of *mai* with first person pronouns also (e.g. *fó ba hau* 'give to 1S' = 'give it to me'), as is the norm in Tetun Terik. Speakers of Tetun Dili disagree as to how acceptable this incursion is. If the prepositional phrase is predicative, however, *ba* is necessarily used instead of *mai* (6.6).

(6.2) *Dala barak hau lao ba foho.*
 occasion many 1S walk to mountain
 'Often I go to the mountains.'

(6.3) a. *Nia fó livru mai ami.*
 3S give book come 1PE
 'He gave us a book.'

⁷ East Timorese rarely make this distinction orthographically, spelling both as 'ba'.

⁸ When used in temporal expressions, *bá*, as part of the sequence *liu bá* 'pass go' is associated with past time (e.g. *fulan rua liu bá* 'month two pass go' = 'two months ago'), while *mai* is associated with future time (e.g. *iha fulan Agustu mai* 'LOC month August come' = 'this coming August').

b. *Sira fó livru ba nia.*
 3P give book to 3S
 ‘They gave him a book.’

(6.4) a. *Nia husu kadeira mai hau.*
 3S ask chair come 1S
 ‘He asked me for chairs.’

b. *Hau husu ba o, “...”*
 1S ask to 2S
 ‘I ask you, “...”’

(6.5) *Buat saida mak imi halo mai hau, hau la haluha.*
 thing what FOC 2P do come 1S 1S not forget
 ‘Whatever you do to/for me, I won’t forget.’

(6.6) *Kamizola né ba hau.* not: **Kamizola né mai hau.*
 T-shirt this for 1S
 ‘This T-shirt (is) for me.’

The preposition *ba* is also used sentence-initially to mean ‘for, as for’, indicating who the following statement is claimed for. Note that in this context *ba* cannot be replaced by *mai* when it has a first person complement.

(6.7) *Ba ema seluk hau la hatene, maibé ba hau,...*
 for person other 1S not know but for 1S,
 ‘For other people I don’t know, but for me, (the most important issue at present is how to train East Timorese to live better).’

6.4 *hosi* ‘from’, *tó* ‘until’

The preposition *hosi* ‘from, originate from’ is used to express source location. In a sentence such as (6.8), its meaning is ambiguous between one of origin (e.g. ‘She originates from Baucau’), or one of more immediate source (e.g. ‘She has just travelled from Baucau’).

(6.8) *Nia mai hosi Baucau.*
 3S come from Baucau
 ‘She comes from Baucau.’ or: ‘She has come here from Baucau.’

In addition, *hosi* introduces the ‘source’ for a range of verbs, including verbs of making (e.g. *halo hosi rai* ‘make from earth’ = ‘made from dirt’), and receiving (6.9). It is occasionally used as an alternative to *iha* for introducing static location, as in (6.10).

(6.9) *Nia simu osan hosi Maria.*
 3S receive money from Maria
 ‘She received money from Maria.’

- (6.10) *Sira hamrik hosi kotuk.*
 3P stand from back
 ‘They stood behind (the people).’

The preposition and transitive verb *tó* ‘until, reach’ is one means of introducing goal location; this is exemplified along with its temporal use in section 6.6.6. Goal location is however more commonly introduced by *ba* ‘to’ and *mai* ‘come’, discussed in the preceding section.

6.5 *ho, hodi, lori* ‘with’

The preposition *ho* ‘with’ introduces a wide range of oblique arguments, including accompaniment (6.11), instrument (6.12), manner (6.13), and means of transport (6.14).⁹

- (6.11) *Hau lao ho nia.*
 1S walk with 3S
 ‘I walk with him.’
- (6.12) *Nia lao ho liman.*
 3S walk with hand
 ‘He walks on his hands.’
- (6.13) *Tama ho disiplina!*
 enter with discipline
 ‘Enter in a disciplined manner!’
- (6.14) *Hau bá ho bís.*
 1S go with bus
 ‘I go by bus.’

An alternative means of introducing instruments is to use *lori* and *hodi*, which are verbs meaning ‘carry, take, bring’ as well as instrumental prepositions. Since these occur in serial verb as well as prepositional constructions, they are discussed in section 9.4.

6.6 Temporal prepositions

6.6.1 Introduction

All of the terms discussed in this section are both prepositions (i.e. they can introduce noun phrases) and conjunctions (i.e. they can introduce clauses and verb phrases). Although there are a significant number of such terms discussed here, they are used relatively infrequently. Alternatives for indicating temporal relationships are:

⁹ *Ho* is also a coordinator meaning ‘and’ (see section 10.4.2). In Tetun Terik, *hó* is a transitive verb meaning ‘accompany’; in Dili, its verbal use has been taken over by the Portuguese loan *akompanha*.

A Portuguese loan *sein* ‘without’ is used as the antonym of *ho* ‘with’ by some speakers.

1. the preposition *iha* ‘at (time), LOC’, and *iha...laran* ‘during’ (see section 6.2);
2. temporal conjunctions which are not classed as prepositions (see section 10.6.2); these all mean something like ‘when, whenever’;
3. tense–aspect markers (see section 8.2);
4. temporal connective adverbs (e.g. *depois* ‘afterwards, then’; see section 10.7).

6.6.2 ‘before’

To express the meaning ‘before (time)’, the Portuguese loan *antes* (or *antes de*, for speakers of Portuguese) may be used (e.g. *antes konvensaun* ‘before (the) convention’). The alternative *molok* ‘before’ is used mainly in church and other formal contexts (e.g. *molok atu hán* ‘before IRR eat’ = ‘before eating’).

6.6.3 ‘during, for’

Duranti (or *durante*) ‘for (period of time), during, when’ (from the Portuguese preposition *durante*) is primarily used in formal and written contexts. The time period is usually specified as a number of units of time, sometimes followed by *nia laran* ‘POS inside’, as in (6.15). Alternatively the time is given by a noun phrase or (less commonly) clause specifying a simultaneous period, as in (6.16a) and (6.17). A further alternative is to use the construction *iha ... laran*, illustrated in (6.16b) (see section 6.2).

(6.15) a. *Hau hanorin duranti oras rua (nia laran).*
 1S teach during hour two POS inside

b. *Hau hanorin oras rua nia laran.*
 1S teach hour two POS inside

‘I teach for two hours.’

(6.16) a. *duranti Agustu nia laran*
 during August POS inside

b. *iha Agustu nia laran*
 LOC August POS inside

‘during August’

(6.17) *duranti Indonesia nia tempu*
 during Indonesia POS time
 ‘during the Indonesian period’

6.6.4 ‘after’

The concept ‘after’ may be expressed by the preposition and conjunction *depois* (or, particularly for Portuguese speakers, *depois de*) (e.g. *depois de konsulta populár* ‘after

consultation popular’ = ‘after the popular consultation’). It is however more usual to use *depois* as a connective adverb meaning ‘afterwards, then’.

An alternative to *depois* is *liu* ‘pass, after’, which is particularly used in church and by Tetun Terik speakers (e.g. *liu semana tolu* ‘pass week three’ = ‘after three weeks’). *Liu* here appears not to be fully prepositional since it frequently co-occurs, like verbs, with the perfective marker *tiha* (e.g. *liu tiha fulan hát* ‘pass PRF month four’ = ‘after four months’). It is also an intransitive verb meaning ‘pass’. In this capacity *liu* can take a temporal phrase as its subject, indicating that the specified time has passed (e.g. *tuku nén liu ona* ‘o’clock six pass ANT’ = ‘(it is) past six o’clock’).¹⁰

6.6.5 ‘from, since’

Although *komesa* is primarily (as in Portuguese) a verb meaning ‘start’, it can also be used as a preposition (6.18) or a conjunction with the meaning ‘from (time, phase)’. In the church register, the Tetun Terik *hahú* is preferred; it too is a verb meaning ‘start’ as well as a temporal preposition and conjunction.

- (6.18) *Nuné komesa sinku anu-s tó kinzi anu-s, ita hanorin sira...*
 so start five year-PL until fifteen year-PL 1PI teach 3P
 ‘So, from five to fifteen years (of age), we instruct them (appropriately to their age).’

Dezde ‘since (time)’ derives from the Portuguese preposition *desde* (e.g. *dezde Domingu* ‘since Sunday’, *dezde hau sei kík* ‘since 1S still small’ = ‘since I was young’).

While some speakers allow *hori* to be a productive preposition with the Tetun Terik meaning of ‘since’ (e.g. *hori Domingu* ‘since Sunday’), such use is both rare and disputed, and for many *hori* is restricted to a finite set of past-time expressions. The most common of these are:

<i>hori-bain-hira</i>	‘when?’ (question about the past)	(lit. ‘at-day-how.many’)
<i>hori-bain-rua-k</i>	‘two days ago’	(lit. ‘at-day-two-k’)
<i>hori-kalan</i>	‘last night, yesterday evening’	(lit. ‘at-night’)
<i>hori-ohin, hor-ohin</i>	‘a little while ago (today)’	(lit. ‘at-just.before’)
<i>hori-seik</i>	‘yesterday’	(* <i>seik</i> is not a word)
<i>hori-uluk</i>	‘very long ago, in the distant past’	(lit. ‘at-formerly’)

A meaning of ‘since’ can be derived by adding *kedas* ‘IMMED, in advance, immediately’. Thus *hori-uluk* means ‘long ago’, while *hori-uluk kedas* means ‘since a long time ago’ (i.e. starting a long time ago, and continuing until now).

6.6.6 ‘until’

The preposition and temporal conjunction *tó* means ‘until, up to’. With time complements, its interpretation depends on context: if the complement represents a point of time, it means

¹⁰ We have no evidence for *tuir* ‘follow’ meaning ‘after (time)’, as listed in Hull (1999), although it is used in our data to mean ‘according to’ or ‘following (someone’s opinion)’ (section 6.7).

‘until’ (6.19), while if it represents a duration, *tó* is interpreted as ‘for (the period of time)’ (6.20). A common construction is for *tó* to indicate how long a preceding situation continued, while also presenting the time at which the following situation began (6.21).

(6.19) *Maibé tó agora ami seidak koalia intensivu liu.*
 but until now IPE not.yet talk intensive more
 ‘But until now we haven’t discussed (it) in greater detail.’

(6.20) *Hau bá iha ailaran tó tinan rua.*
 IS go LOC jungle until year two
 ‘I went into the jungle for two years.’

(6.21) *Nuné tó dadér rai loron, aviaun...*
 like.this until morning land day aeroplane
 ‘So it went on until in the morning a helicopter (was about to land on the air field).’

With other quantitative complements, *tó* means ‘up to and including’.

(6.22) *I klasifikasaun tuir mai: tolu-nulu resin lima tó hát-nulu resin lima.*
 and classification follow come three-tens extra five until four-tens extra five
 ‘And the following classification is: (from age) thirty-five to forty-five.’

As a preposition, *tó* is also sometimes used to introduce goal location (6.23). This relates to its use as an intransitive verb meaning ‘arrive, suffice’ (e.g. *Nia tó iha Dare* ‘3S arrive LOC Dare’ = ‘He arrived in Dare’), and its (predominantly Tetun Terik) use as a transitive verb meaning ‘reach, arrive at’.

(6.23) *Kortina né mós nak-lés ba rua, hosi leten tó kraik.*
 curtain this also INTR-tear to two from top until bottom
 ‘The curtain also tore in two, from top to bottom.’

Also reflecting its history as a verb is the fact that, unlike other prepositions, *tó* can be negated by the verbal negator *la* ‘not’.

Portuguese *até* ‘until (time)’ is used primarily in standard Portuguese leave-takings (e.g. *até logu* (often pronounced with a single stress as *atelogu*) ‘until later, see you later’, from *até logo*). A minority of speakers use it with Tetun complements in leave-takings (e.g. *até aban* ‘until tomorrow, see you tomorrow’). An even smaller minority extend it to mean ‘until’ outside of leave-takings.

6.7 Other prepositions

Besik ‘close to, near (location)’ can introduce location for a preceding verb phrase (e.g. *hamrík besik kareta* ‘stand near (the) car’). It is also a verb with the same meaning. When used predicatively, it can be intransitive (e.g. *aldeia nebé besik* ‘village REL close’ = ‘village that is nearby’), have a direct object (e.g. *besik uma* ‘close.to home’), or have a complement introduced by *ba* ‘to’, *iha* ‘LOC’ or *ho* ‘with’. Additionally, *besik* can, like *kuazi* ‘almost,

nearly', be used adverbially to mean 'almost' (e.g. *besik semana rua* 'almost week two' = 'almost two weeks').¹¹

Kona-ba 'about, concerning (a topic)' (lit. 'touch to') introduces a topic of speech or thought.¹² The synonym *sobre* is occasionally used by Portuguese speakers.

- (6.24) *Hau hanorin kona-ba agrikultura.*
 1S teach about agriculture.
 'I teach agriculture.'

Hanesan 'like' indicates likeness in some respect. As a preposition, it follows a verb phrase to introduce a peripheral phrase (6.25). When used predicatively, *hanesan* 'be like, be similar' is fully verbal, being able to be negated by *la* 'not', and occurring both transitively and intransitively. It is also used with a sense of 'for example, for instance, like' (6.26), introducing a wide range of constituent types (including noun phrases and clauses).

- (6.25) *Nia matenek, maibé nia halo-an hanesan ema bulak.*
 3S intelligent but 3S do/make-self like person crazy
 'He's intelligent, but he acts like a crazy person.'

- (6.26) *Hanesan ema nia ruin tohar, nia bele halo tomak fali.*
 like person POS bone break 3S can make whole again
 (Explaining the work of a healer:) 'For instance if a person has a broken bone, he can make it whole again.'

Nudar 'as, in the capacity of' generally reminds the addressee of that part of the referent's identity that is particularly relevant to the current situation. It is primarily used in church and formal contexts.

- (6.27) *Nudar feto Timór Loro-sae, nia la túr nonók.*
 as woman Timor sun-ascend 3S not sit quietly
 (From a funeral speech honouring East Timorese female martyrs:) 'As an East Timorese woman, she didn't sit quietly.'

Entre 'between, among, amongst' introduces co-participants for an abstract noun (e.g. *kolaborasaun entre sira* 'collaboration between them'). It is, like *hosi* 'from', also used sentence-initially to specify a group from which one is to select members to talk about (6.28). This Portuguese loan is used primarily by more Portuguese-influenced speakers. Note that

¹¹ The antonym of *besik* 'close to' is *dók* 'far'; however *dók* cannot be used as a preposition or transitive verb, its complement instead being introduced by *hosi* 'from' (e.g. *hela dók hosi uma* 'stay far from home').

¹² Some people use *kona* for this function, particularly in writing, but *kona-ba* is strongly preferred by many speakers. It is possible though rare for *kona-ba* to take an adverbial modifier (e.g. *kona mós ba...* 'touch also to...' = 'also concerning...'); this suggests that *kona* still has some verbal properties, and that *kona-ba* is not fully lexicalised as a preposition. Native speakers differ as to whether they consider *kona-ba* to be one word or two, although it is generally written as two. We have no evidence for a complex preposition *kona-tuir*, listed in Hull (1996b:96) as 'about'.

locative ‘between’ (e.g. ‘I sat between Betty and Alex’) is translated by *iha...klaran* (see section 6.2), not *entre*.

- (6.28) *Entre feto sira né, iha Salome,...*
 amongst woman PL this EXIST Salome,
 ‘Amongst these women, there was Salome, ...’

Konformi (or *konforme*) ‘according to, depending on (a person, a source of data, a person’s desires or abilities...)’ is in Portuguese a preposition, but functions more widely in Tetun as both a preposition and a conjunction (e.g. *konformi ami bele* ‘according to what we were able’). It also stands alone to mean ‘it’s up to (you)’, for instance in answer to a request for advice.

Tanba and *tán* ‘because’ are primarily conjunctions (see section 10.6.3), but can also introduce noun phrases (e.g. *tanba né* ‘because.of this’, *tanba lia bobar* ‘because.of word encircle’ = ‘because of lies’).

Tuir ‘according to (someone’s opinion/desire, an information source, criterion, ability, a person, a timetable...)’ frequently specifies the source for a statement (e.g. *tuir hau nia hanoin* ‘follow 1S POS think’ = ‘in my opinion’, *tuir hau nia observasaun* ‘follow 1S POS observation’ = ‘from my observations’). It is also a transitive verb meaning ‘follow, obey, participate in (meeting, exam, group...)’.

Portuguese *de* ‘of’ is used in Portuguese-style dates to separate the day and month, or sometimes the month and year (e.g. *dia trinta de Agustu* ‘day thirty of August’ = ‘30th August’). It is also used in phrases from Portuguese (e.g. *ministru de edukasaun* ‘minister of education’), but not in phrases that include non-Portuguese words.

7. Basic clauses

7.1 Verbal clauses

7.1.1 Basic clause order

Tetun verbal clauses have subject–predicate order (7.1); the exception is presentative existential clauses, in which the verb can be initial (see sections 7.1.5 and 7.1.6). Unlike Tetun Terik, Tetun Dili has no subject marking on verbs.

- (7.1) *Labarik né sei kík.*
child this still small
'The child was still small/young.'

The default order for transitive clauses is subject–verb–object (7.2). The object can however be fronted to before the subject (see section 7.1.3), or, in restricted circumstances, occur between the subject and verb (section 7.1.4).

- (7.2) *Ami lori kuda ida.*
1PE bring horse one
'We brought one horse.'

Oblique arguments are usually introduced by prepositions, typically following the object slot. This includes goal (7.3), source, recipients (7.4), addressees, instruments, and so on. Instruments and some direction phrases can alternatively be introduced by serialised verbs, discussed in sections 9.4 and 9.3. Complement clauses, described in section 10.5, also occur in this post-object position.

- (7.3) *Sira halai ba Dare.*
3P run to Dare
'They ran (away) to Dare.'

The most common means of expressing a recipient is to introduce it by the prepositions *ba* 'to (away from speaker, not towards speaker)' or *mai* 'come (towards speaker)', placing the prepositional phrase after the object, as in (7.4). The preposition can sometimes be omitted (e.g. *Sira fô kanuru ami* '3P give spoon 1PE' = 'They gave us spoons'); however this is

considered to be less polite, and speakers differ on whether they find it acceptable at all. The object can alternatively follow the recipient, particularly if the object noun phrase is long, as in (7.5). In this case, too, the preposition is omissible, at least in some instances (e.g. *fó nia susubén* ‘give 3S milk’ = ‘give her milk’).

(7.4) *Sira fó kanuru mai ami.*
3P give spoon come 1PE
‘They gave us spoons.’

(7.5) *Nia fó ba sira ukun rasik an.*
3S give to 3P rule own self
‘He gave them self-rule.’

Tetun has no system of voice; that is, there is no passive construction. There are however several means of creating passive-like clauses. One is to use a detransitivising prefix (e.g. *nak-duir* Vi ‘roll’ from *duir* Vt ‘roll’; see section 4.2.4).

The second is to downgrade the prominence of the subject by either omitting it (7.15) or using a generic *ema* ‘person, people, someone’ (7.6). The prominence of the object can be upgraded by fronting it, as in the second clause of (7.7) (see section 7.1.3).¹

(7.6) *Ema sunu bispu nia uma.*
person burn bishop POS house
‘People/someone burned the bishop’s house.’

(7.7) *Balu mate, balu kikoan sira né ema sama.*
some die some little PL this person tread.on
‘(In the stampede) some (people) died, some—the little ones—were trampled.’

Finally, an abstract noun can be made the object of *hetan* ‘get’ or *simu* ‘receive’ (e.g. *hetan ajuda hosi...* ‘get help from...’); this uncommon construction is found mostly amongst Portuguese-influenced speakers.

(7.8) *Ami hetan masakre hosi Darah Merah.*
1PE get massacre from blood red
‘We were massacred by the Darah Merah (a militia group whose name in Indonesian means ‘Red Blood’).’

7.1.2 Subject and object

Subjects can be recognised primarily by their position preceding the predicate, except in the specific circumstances discussed in sections 7.1.5 and 7.1.6, and in equative clauses,

¹ Native Tetun speakers sometimes translate passive clauses in other languages by clauses with fronted objects and focused subjects, for instance translating ‘That dog was hit by Jose’ as *Asu nebá, Jose mak baku*. ‘dog there Jose FOC hit’. It is apparently this that leads Martins (1998) to label such constructions ‘passive’. The construction is, not surprisingly, rare.

where the order of the two noun phrases is to some extent reversible (see section 7.3). Other characteristics unique to subjects are that they control reflexivisation (section 7.5), and are obligatorily omitted in reduced complement clauses (section 10.5.3) and before the second verb in serial verb constructions (chapter 9).

Semantically, subjects of intransitive clauses may be either actors (e.g. *Sira halai* ‘3P run’ = ‘They ran’) or undergoers (e.g. *Bé nak-fakar* ‘water INTR-spill’ = ‘The water spilled’). Subjects of transitive clauses are always more agentive than the objects.

While subjects are prototypically definite, they need not be (7.6). One result of the large influx of abstract nouns from Portuguese is that (in contrast to Tetun Terik) abstract nouns can readily function as subjects, particularly in written genres (van Klinken 1999b).

- (7.9) *Demokrasia maka hanorin nuné.*
 democracy FOC teach like.this
 ‘It is democracy that teaches this.’

Objects too are recognised primarily by their default position following the verb; alternatively they are fronted clause-initially, or (under restricted circumstances) incorporated into the verb. Semantically, objects always represent the undergoer.

Both subjects and objects are readily omissible in contexts where the referent is either recoverable from context or is irrelevant. There appear to be no grammatical rules constraining the interpretation except in the case of subjects which are obligatorily omitted in serial verb constructions or in reduced complements. Instead, the intended referents of omitted arguments must be inferred pragmatically.

The following extract from a spoken recipe illustrates frequent omission of the subject (‘we’) and object (the ingredients). While the inferred subject remains *ita* ‘we’ (here used in the indefinite sense of formal English ‘one’) throughout most of the clauses, the subject of the final verb is instead the food.

- (7.10) *Depois ita tau mina; ita sona uluk lís;*
 afterwards 1PI put oil 1PI fry first garlic
 ‘Then we put in oil; first we fry the garlic;

depois tau tomati bá, halo bén tiha, tau nán ona.
 afterwards put tomato go make liquid PRF put meat ANT
 then (we) add tomato to (it), (and after) making it soft/like liquid, add meat.

Tau ba taxu, depois fila, fila tó tasak.
 put to pan afterwards turn turn until cooked
 (We) put (it) in the pan, then stir (it)—stir (it) until (it) is cooked.’

7.1.3 Object fronting

The object can be fronted to before the subject (if any).

Fronting is often for reasons of contrast with other members of a set which have been either explicitly mentioned or implicitly evoked. It is common but not necessary to mark such fronted contrastive noun phrases with *mak* (see section 7.8).

- (7.11) *UNAMET sira hili deit ema CNRT mak servisu hamutuk ho sira,*
UNAMET PL choose just person CNRT FOC work together with 3P

I pro-otonomi sira la hili.
and pro-autonomy 3P not choose

‘UNAMET (the UN body overseeing the 1999 popular consultation on self-determination) people chose only CNRT (the pro-independence National Council of Timorese Resistance) people to work with them, and didn’t choose pro-autonomy people.’

- (7.12) *Hau hatene tuir hanorin né halo nusá.*
1S know follow teach this do how
(A teacher recalls when Indonesian was first made the language of education:) ‘I (already) knew how to teach.’

Só deit lingua né mak hau la hatene.
only just language this FOC 1S not know
It was only the language that I didn’t know.’

The object can also be fronted to mark it as topical. Such topics are often but not necessarily marked as anaphoric by *né* ‘this’. If the identity of the subject is irrelevant, the subject is often either omitted (7.15) or is generic *ema* ‘person’ = ‘someone/people’ or *ita* ‘1PI’. *Ita* is in this context often used indefinitely, like formal English ‘one’ or informal English ‘they’ (7.14).

- (7.13) ... *Nán ita koa halo kabelak bo-bót; depois ita tetak liu deit.*
meat 1PI cut make flat.segment RDP-big afterwards 1PI tenderise more just
‘(When we start making *bifī*, we must already have meat, flavourings, and so on.)
The meat we cut into big pieces; then we further tenderise it.’

- (7.14) *Buat nebé mak nia hakarak, ita lakohi fó.*
thing REL FOC 3S want 1PI not.want give
(About overstrict parents:) ‘The things which he/she (the child) wants, we refuse to give.’

- (7.15) ..., *ami hotu kontenti, maibé odamatan janela taka metin.*
1PE all happy but door window close tight
‘(When we heard that the pro-independence side had won), we were all happy, but the doors and windows were kept tightly closed (as we were afraid).’

7.1.4 Subject–object–verb (object incorporation)

An object can also be placed between the subject and the verb (i.e. subject–object–verb), a position also found under similar conditions for complements of *bele* ‘can’ and *hatene* ‘know, know how to’ (see sections 8.3.2 and 10.5.3). This construction occurs relatively infrequently, and is found only in informal speech. It is restricted to clauses expressing a contrast, such as in countering a claim to the contrary (e.g. *Hau Inglés hatene!* ‘Is English know’ = ‘I do know English!’), contrasting with an earlier situation, or contrasting what this subject referent does with what others do (7.16). The object noun phrase is usually a single, unmodified, non-referential noun. The construction appears to be object-incorporation.

- (7.16) ... *Só ida né mak keiju hán.*
 only one this FOC cheese eat
 ‘(We don’t really eat cheese.) It is only this one (my little daughter) that is a cheese-eater.’

7.1.5 Existential and possessive

The word *iha* is a general locative preposition meaning ‘at, in, on...’ (see section 6.2), an existential verb meaning ‘exist, there is/are’, and a transitive verb meaning ‘have’.²

For existential *iha*, the single argument either precedes or follows *iha*. The argument must follow *iha* if it is modified by a relative clause (7.17). This can usually be translated as ‘there BE someone/something that...’. This is also the usual order for indefinite arguments, whose existence is being predicated (7.18). With or without a relative clause, this construction typically introduces new, usually minor, participants into a discourse.³

- (7.17) *Iha distritu lima mak iha ona baunilha.*
 EXIST district five FOC have ANT vanilla
 ‘There are five districts which have vanilla (as a cash crop).’

- (7.18) *Depois komesa iha funu.*
 afterwards start EXIST war
 ‘Then there started to be fighting.’

When the argument is definite (and therefore more prototypically subject-like), the alternative order in which the argument precedes *iha* is strongly preferred. In this case, the

² A close relationship between locatives, existentials, and possessives has been noted cross-linguistically by Clark (1978), amongst others.

Note that *Nia iha uma* ‘3S have/LOC house’ could be interpreted as either ‘She has a house’ or ‘She is at home’. They differ, however, in that the possessive interpretation is negated by *la*, which negates verbs and adjectives (*Nia la iha uma* ‘She does not have a house’), while the locative interpretation is negated by *laós*, which is primarily used for negating constituents other than verbs and adjectives (*Nia laós iha uma* ‘She is not at home’).

³ This is consistent with the strong cross-linguistic tendency for presentational clauses to be verb-initial, thus cancelling the expectation that the single argument noun phrase is ‘given’ (Clark 1978; Givón 1978:295).

verb is normally translatable as ‘be present, be here’, predicating presence rather than existence. This order is primarily used for irrealis clauses, with 80% of the examples from the oral corpus being negative (7.19), conditional, or questions.

- (7.19) *Apá la iha.*
Dad not EXIST
‘Dad is not here.’

This is because most positive statements specifying that something is present encode its location as well (7.20), and so use *iha* transitively (as a preposition).

- (7.20) *Apá iha merkadu.*
Dad LOC market
‘Dad is at the market.’

When specifying that something is available to, or possessed by, some unspecified person, either order is possible, with the argument-final order (7.21) being the most common. This order could be interpreted syntactically either as existential, or as possessive with an omitted subject.

- (7.21) *Se iha osan, diak liu sosa kaixa.*
if EXIST/have money good more buy box
‘If there is money, it is better to buy a box (as that is cheaper than buying individual items).’ or: ‘If you have money,...’

- (7.22) *Ami bá, hahán la iha.*
IPI go food not EXIST
‘(When) we went, there was no food.’ or: ‘..., we had no food.’

For possessive *iha*, the usual order of constituents is subject–verb–object (7.23). A far less common order is subject–object–verb (7.24); some people, however, consider this order to be unacceptable in Tetun Dili.

- (7.23) *Nia la iha servisu.*
3S not have work
‘He doesn’t have work.’

- (7.24) *Sira kilat la iha.*
3P gun not have
‘They (UNAMET police) didn’t have guns.’

7.1.6 Other verb–subject clauses

There are a number of other intransitive verbs which can, like *iha* ‘EXIST’, be followed by non-definite subjects which present new, minor participants. This is found in writing (including Bible translation) and formal speech, but very little in everyday speech. It appears

to be calqued on Portuguese, for which the corresponding verbs can all occur clause-initially.⁴ The verbs which occur in this construction include *mosu* ‘appear’ (7.25) (Portuguese *aparecer*), *akontese* ‘happen’ (7.26) (Portuguese *acontecer*), *moris* ‘live’ (Portuguese *viver*), *falta* ‘be absent, be missing’ (7.27) (Portuguese *faltar*), and *hela* ‘remain’ (Portuguese *ficar*).

Note that all these verbs usually take a preceding subject, with the subject–verb order being strongly preferred for definite subjects (7.28) and in ordinary speech.

(7.25) *Ohin loron mosu violénsia barak tebes iha Timór.*
today day appear violence much true LOC Timor
‘These days much violence has appeared/started in Timor.’

(7.26) ... *tán atu akontese buat ruma la diak ba nia oan feto ka?*
because IRR happen thing some not good for 3S.POS child female or
‘(Is she afraid) because something bad might happen to her daughter?’

(7.27) *Falta ida tán karik.*
missing one more maybe
‘Perhaps there is one more (card) missing.’

(7.28) *Sira barak falta.*
3P many missing
‘Many of them are absent (from school).’

7.1.7 Verbs taking adjectival complements

Tetun has no copulas. There are however some Portuguese loan verbs which take adjectival complements.⁵

Sai ‘become’ (from Portuguese *sair* ‘come out’) takes as complement either a noun phrase (e.g. *sai prezidenti* ‘become president’) or an adjective phrase (e.g. *sai diak liu tán* ‘become good further more’ = ‘become even better’).

Senti ‘feel’ allows adjectival complements (e.g. *senti hamlaha* ‘feel hungry’, *senti moras* ‘feel sick’).

7.2 ‘Body-good’ expressions

Character, emotions, health and physical attributes of people are often expressed by ‘body-good’ expressions, which consist of a noun followed by a single-word predicate (e.g. *isin diak*

⁴ Further support for the calquing theory comes from the fact that this construction is not found in Tetun Terik.

⁵ These could be classed as ‘semi-copulas’ (Hengeveld 1992). Apart from allowing adjectival complements, they differ from transitive verbs in two ways. Firstly, transitive verbs generally have selectional restrictions between the verb and the object, while these semi-copulas have selectional restrictions between the subject and the complement. Secondly, the complement cannot be fronted.

'body good' = 'healthy, well'). The noun usually denotes a body part, and the predicate is typically an adjective.

Expression	Gloss	Meaning
<i>ain ás</i>	'leg tall'	'tall (of person)'
<i>aten barani</i>	'liver brave'	'very courageous'
<i>isin manas</i>	'body hot'	'have a fever'
<i>laran diak</i>	'inside good'	'kind-hearted, generous'
<i>laran moras</i>	'inside sick'	'jealous; sad'

Many such expressions are standardised; there is however also considerable inter-speaker variation in the expressions that are used and in their interpretation. For instance, *matan moris* (lit. 'eye live') for many means 'vigilant, alert'; some however use this expression to describe giggling 'boy-crazy' girls, a concept for which others prefer the expression *oin moris* (lit. 'face live'). Some expressions can have both a literal and a metaphorical interpretation, again with inter-speaker variation on whether either or both is understood. So, for instance, *liman át* (lit. 'arm bad') is by some interpreted only literally to mean 'having an injured or crippled arm', while others interpret it metaphorically as 'habitually thieving'.

While emotions such as happiness and sadness are in daily speech mostly attributed to the *laran* 'inside' (e.g. *laran kontenti* 'inside happy' = 'happy'), just as in Tetun Terik, some Portuguese-influenced speakers and the church register tend to attribute them instead to the *fuan* 'heart' (e.g. *fuan kontenti* 'happy'). This is a calque on the Portuguese uses of *coração* 'heart'.

Most body-good expressions behave in some contexts like single-word adjectives (i.e. as compounds), but are in other contexts clearly two separate words. This 'split personality' is shown by alternatives in three areas of grammar, which are considered in more detail below: (1) in whether the body part noun heads the subject of a clause or is part of the predicate; (2) in whether modifiers are placed within or before the body-good expression; and (3) in whether the expression behaves like a relative clause or like an adjective when it modifies a noun. A minority of body-good expressions, however, are always treated grammatically as single-word adjectives; these are thus fully compounded (e.g. *isin rua* 'body two' = 'pregnant').

When used predicatively, there are two grammatical options. The most common is for the noun phrase referring to the person to be subject, and the body-good expression to function as predicate, as in (7.29). The less common is for the body part to be subject, with the person as its possessor, and the adjective on its own to head the predicate, as in (7.30).

(7.29) *Hau laran susar tanba hau nia inan moras.*
 1S inside difficult because 1S POS mother sick
 'I am distressed because my mother is ill.'

(7.30) *Hau nia laran né susar tebe-tebes.*
 1S POS inside this difficult RDP-truly
 'I was truly distressed (fearing what my captors would do to me).'

Under the influence of Portuguese, emotion and health predicates can be introduced by *senti* 'feel'. Here, too, the subject usually refers to the person while the complement of *senti* is the entire body-good expression (7.31). Alternatively, the subject can be the body part noun phrase, while the complement of *senti* is simply the adjective (7.32).

(7.31) *Nia kolega sira senti laran susar tebes.*
 3S.POS friend PL feel inside difficult truly
 'His friends felt truly distressed.'

(7.32) *Ita nia kabun-laran senti moras hotu.*
 IPI POS stomach-inside feel sick all
 'One's (lit. 'our') stomach feels completely sore.'

When the entire body-good expression is used predicatively, as in (7.29), there is considerable variation on where to place those auxiliaries and modifiers that slot between the subject and the predicate in a verbal clause. In some cases they precede the body part noun, as does *bele* 'can' in (7.33); this suggests that the expression as a whole is analysed as the predicate head, and hence as a compound adjective. In other cases, however, the modifiers intervene between the noun and the adjective (7.34), preventing a compound analysis.

(7.33) *Sira iha neba, sira bele laran moras malu.*
 3P LOC there 3P can inside sick RECIP
 'Once they are there, they could get jealous of one another.'

(7.34) *O isin para bot!*
 2S body so big
 'You're so fat!'

The preferred position appears to differ from speaker to speaker, and from modifier to modifier. For negation, for instance, some allow the negator *la* (which elsewhere negates verbs and adjectives only) to precede the adjective but not the noun (e.g. *Labarik isin la diak* 'child body not good' = 'The child was unwell'). Some, however, accept it before the noun, thus analysing the noun-adjective pair as a complex adjective (e.g. *la laran ksolok* 'not inside content' = 'not content'). Yet another alternative is to avoid the issue by using the contrastive negator *laos* (which can negate any constituent type) before the noun (e.g. *Nia laos isin diak* '3S indeed.not body good' = 'He is not well').

When used as noun phrase modifiers, body-good expressions frequently occur in the adjective phrase slot, without a relative clause marker and optionally followed by determiners (e.g. *ema ain at ida* 'person leg bad one' = 'a cripple'). However it is also common to have a relative clause marker, which while possible for adjectives is much more common for clauses (e.g. *ema nebe laran at* 'person REL inside bad' = '(a) person who is nasty').

Many body-good expressions can also be used as nouns, with abstract meanings (e.g. *laran kontenti* 'inside happy' = 'happiness, contentment'). In this case, the two words making up the expression cannot be separated, thus allowing the sequence to be analysed as a compound.

- (7.35) *Isin manas la iha?*
 body hot not exist
 'Is there no fever?'

7.3 Other non-verbal clauses

In non-verbal clauses, the predicate is juxtaposed to the subject. There is no copula.

The usual order for clauses with nominal predicates is for the old information to be given in the first noun phrase (provisionally analysed as subject), and for the new information to follow (7.36). Nominal predicates can (like most other noun phrases) consist simply of a possessor phrase (7.37).

- (7.36) *Né hau nia alin.*
 this 1S POS younger.sibling
 'This is my younger sibling.'

- (7.37) *Oklu né hau nian.*
 spectacles this 1S POS
 'These spectacles are mine.'

In questions, however, the questioned constituent may either occur in second position (7.38), or be fronted, in which case it is focused by *mak* (7.39).

- (7.38) *Emá né sé lós?*
 person this who true
 'Who is this person really?'

- (7.39) *Ida nebé mak o nia uma?*
 one which FOC 2S POS house
 'Which is your house?'

Numeral phrase predicates can specify the number of the subject (7.40a), or its cost or some other measure. The former is however more commonly expressed by a possessive or existential construction, as in (7.40b), while for the latter some prefer to precede the cost with the noun *folin* 'price, value'.

- (7.40) a. *Hau nia pontu-s rua.*
 1S POS point-PL two
 'My points are two.'

- b. *Hau iha pontu-s rua.*
 1S have point-PL two
 'I have two points (to make).'

A number of prepositions can introduce prepositional phrase predicates, as in (7.41).

- (7.41) “*Nia iha nebé?*” “*Nia iha Same.*”
 3S LOC where 3S LOC Same
 “Where is he?” “He is in Same.”

Predicates headed by *naran* ‘name’ usually consist simply of *naran* followed by a proper name, as in (7.42a). Unlike verbal predicates, they cannot be negated by *la* ‘not’, and are usually not preceded by a relative clause marker when they modify a noun phrase (7.43). Evidence that *naran* in (7.42a) is not part of the subject comes from two sources. Firstly, when *naran* is possessed, it requires explicit possessive marking using *nia* (see section 5.4), as in (7.42b). Secondly, it is possible to put an adverb between the subject and *naran* (e.g. *Hau mós naran Maria* ‘1S also name Maria’ = ‘I too am called Maria’). Although *naran* is in this construction part of the predicate, there is an equally common construction in which it is explicitly possessed as part of the subject, with the proper name on its own constituting a nominal predicate, as in (7.42b).

- (7.42) a. *Hau naran Maria.*
 1s name Maria

- b. *Hau nia naran Maria.*
 1s pos name Maria

‘My name is Maria.’

- (7.43) *Depois mak kose aimoruk ida naran Vaseline ba ligadura mós.*
 afterwards FOC rub medicine one name Vaseline to bandage clean
 ‘Then rub a medicine/ointment called Vaseline onto a clean bandage.’

7.4 Reciprocal

Reciprocals are formed by placing *malu* ‘RECIP, each other’ in the object, addressee or recipient slot of a clause. It occurs in two main constructions.

In the first, the subject has plural referents (regardless of whether it is syntactically marked as plural). The reciprocal marker indicates that one or more or all of the referents of the subject are doing the action identified by the verb to the others. There is no indication as to whether the referents are acting simultaneously (as would logically be required by *hasoru malu* ‘meet RECIP’), or not (as one would expect of *husu ba malu* ‘ask to RECIP’). This construction requires that the status of the referents be comparable with respect to the specified activity. So, for instance, *Sira baku malu* ‘they hit RECIP’ could be used of a husband and wife hitting each other, or of one of the partners hitting the other, but would not be appropriate for talking about parents hitting little children.

In cases where one of the participants is presented as the instigator, the subject refers to the instigator, the verb is followed by *malu*, and the secondary participants are introduced by *ho* ‘with’. The distinction is illustrated by (7.44).

(7.44) a. *João istori malu ho Maria.*
 John quarrel RECIP with Maria
 ‘John quarrelled with Maria (he started it).’

b. *João ho Maria istori malu.*
 John and Maria quarrel RECIP
 ‘John and Maria quarrelled (no indication as to who started it).’

The subjects of reciprocal clauses are usually human, but need not be (e.g. *Rezultadu besik malu* ‘result close.to RECIP’ = ‘The results were very close’).

There are a number of common expressions in which *malu* cannot be interpreted literally as a reciprocal, either because the preceding term is intransitive, or because the preceding transitive verb takes a different semantic class of object to that represented by *malu*. The following examples illustrate this.

- *diak malu* (lit. ‘good RECIP’) ‘be on good terms with one other’. Note that the adjective *diak* cannot be used transitively. For instance one cannot say **João diak Maria* ‘John good Maria’ to mean that John and Maria are on good terms with each other.

(7.45) *João ho Maria diak malu hela.*
 John and Maria good RECIP CONT
 ‘John and Maria get on well.’

- *hadau malu* (lit. ‘snatch RECIP’) ‘compete for resources’. Outside of reciprocal constructions, *hadau* means ‘snatch, grab, seize’, and its object specifies that which is seized (e.g. *hadau rai* ‘seize (someone else’s) land’). When used with *malu* to mean ‘compete with one another for resources’, *hadau* is unusual in that an object noun phrase specifying what is competed for can follow *malu*, as in (7.46).

(7.46) *Sira hadau malu kanuru.*
 3P snatch RECIP spoon
 ‘They fight each other for the spoon.’

- *hán malu* (lit. ‘eat RECIP’), when used with a human subject means ‘fight (by any means), argue’.

(7.47) ... *tauk atu partidu sira hán malu, i bele mosu fila fali “guera sivil”.*
 afraid IRR party PL eat RECIP and can appear return again war civil
 ‘... are afraid that the parties will fight each other, and that “civil war” will arise again.’

- *hán malu* (lit. ‘eat RECIP’), when the subject refers to what has been said, means ‘agree’.

(7.48) *Sira nia liafuan né la hán malu.*
 3P POS word this not eat RECIP
 ‘Their statements didn’t agree.’

- *diskuti malu* (lit. ‘discuss/argue RECIP’) ‘discuss, argue with each other’. Outside of reciprocal constructions, the object of *diskuti* specifies what is being discussed, not who it is being discussed with.

(7.49) ..., *hau sei diskuti malu ho milísia sira.*
 1S still discuss RECIP with militia PL
 ‘(Having come out onto the verandah), I argued with the militia.’

- *fahe malu* (lit. ‘divide/separate RECIP’) and *haketak malu* (lit. ‘separate RECIP’) ‘separate from each other’. In other contexts, the object of *fahe* or *haketak* specifies what is being separated or divided, while the subject specifies the agent.

(7.50) *Agora ita rua tenki fahe malu. Hau bá loro-sae; o bá rende bá.*
 now 1PI two must separate RECIP 1S go sun-ascend 2S go surrender go
 ‘Now we two must separate. I’ll go east; you go and surrender.’

7.5 Reflexives

Reflexives in Tetun indicate that the subject referent does something to himself/herself. Tetun Dili reflexives are formed in two related ways. In the first, the noun *án* ‘self’ heads a noun phrase, and is possessed by a pronoun which has the same person and number as the subject. While most such noun phrases function as direct object (7.51), *án* can also occur in addressee or beneficiary prepositional phrases (7.52). In the second means of forming reflexives, *-an* (an unstressed variant of *án*) is a clitic attached to the verb; this type is restricted to reflexivisation on the object.⁶ Both kinds of reflexives are illustrated by the synonymous examples in (7.51).

(7.51) a. *Nia oho nia án rasik.*
 3S kill 3S.POS self own

b. *Nia oho-an rasik.*
 3S kill-self own

c. *Nia oho rasik nia án.*
 3S kill own 3S.POS self

‘He himself killed himself.’

(7.52) *Ita uza ba ita nia án.*
 1PI use for 1PI POS self
 ‘We use (the car) for ourselves.’ (i.e. The car is for our private use.)

⁶ Tetun Terik also uses *-an* for reflexives (e.g. *ha-dók-an* ‘cause-far-REFLEXIVE’ = ‘distance oneself’), but there it is not a noun meaning ‘self’, and is never possessed.

Adverbs and tense–aspect words can either precede or follow a possessed *án*, but necessarily follow cliticised *-an*, as illustrated by the position of *rasik* ‘own, oneself’ in (7.51). Possessed *án* can, like other objects, be fronted to before the subject (although this is very rare), but cliticised *-an* can not. A verb and cliticised *-an* are thus inseparable, and in writing are often joined by a hyphen. The exception in which an adverb can intervene is the well-known catch-cry *ukun rasik an* ‘rule own self’ = ‘be independent’ (7.54).

There are many fixed idioms consisting of a transitive verb followed by *-an*, which do not have a variant with a possessed *án*. Such idioms include *foti-an* ‘lift self’ = ‘boast’, *bók-an* ‘move self’ = ‘move (even of inanimate entities)’, and *haraik-an* ‘lower self’ = ‘be humble, humble oneself’. Clitic *-an* is also used for Portuguese loans which in Portuguese are reflexive, such as *esforsa-an* ‘strive’ (from Portuguese *esforçar-se*), and *oferese-an* ‘volunteer’ (from Portuguese *oferecer-se*). Verb-*-an* idioms (but not reflexives with possessed *án*) can readily head noun phrases, a construction primarily found in written and formal registers (7.53, 7.54).

(7.53) *Ita hatudu ita nia haraik-an ba sira.*
 1PI show 1PI POS lower-self to 3P
 ‘We show them our humility.’

(7.54) *Ukun rasik an né laós mamar.*
 rule own self this indeed.not soft
 ‘Independence is not easy.’

Clitic *-an* can also follow some intransitive verbs to indicate that the action happened of its own accord, without anyone causing it.

(7.55) *Odamatan né nak-loke-an.*
 Door this INTR-open-self
 ‘The door opened by itself.’

7.6 Interrogatives

7.6.1 Information interrogatives and their answers

In information interrogatives, a questioned noun phrase usually takes the same place in the sentence as the answer would. The interrogative pronouns are *sé* ‘who’ (7.56), *saida* ‘what’ (7.57), *sá* ‘what’, and *nebé* ‘where, which’. The following examples illustrate questioning of the subject (7.56), object (7.57), possessor (7.58), and an oblique argument (7.59).

(7.56) “*Sé mak fó hán imi?*” “*Ami rasik mak buka hahán.*”
 who FOC give eat 2P 1PE own FOC seek food
 ‘Who fed you (when you first became a refugee)?’ ‘It was we ourselves who sought food (i.e. we had to feed ourselves).’

- (7.57) *O hakarak saida?*
 2S want what
 ‘What do you want?’
- (7.58) *Surat né sé nian?*
 letter this who POS
 ‘Whose is this letter?’
- (7.59) *Ita koalia ho sé?*
 2S.HON speak with who
 ‘Who were you speaking with?’

For ‘what’, *saida* (or *seda*, lit. ‘what-one’) is much more common than its synonym *sá*. *Saida* functions as both pronoun (7.57) and determiner (e.g. *kareta saida* ‘vehicle what’ = ‘what vehicle’). *Sá* too is both pronoun and determiner. As a pronoun, it occurs primarily in a few fixed expressions, including ...*naran sá?* ‘...name what’ = ‘What is...called?’, and ...*halo sá/saida?* ‘...do what’ = ‘What is...doing?’⁷ In contrast to *saida*, when *sá* functions as determiner, it precedes the head noun, and indicates a selection from a restricted list of options (e.g. *sá tinan?* ‘which year?’, *sá kór?* ‘which colour?’).

As a pronoun, the main contexts for *nebé* ‘where’ are *iha nebé* ‘LOC where’ to question location, *bá/ba nebé* ‘go/to where’ to question destination, and *hosi nebé* ‘from where’ to question source. Less commonly, or when it functions as determiner, *nebé* means ‘which’, asking to select one answer from an already-narrowed-down set. Question (7.60), for instance, would be appropriate if one knew that one of the vehicles in a lot was to be used, but were unsure of which.

- (7.60) “*Kareta nebé mak ita atu sae?*” “*Maria sira nian.*”
 vehicle which FOC IPI IRR ascend Maria PL POS
 ‘Which vehicle will we ride in?’ ‘Maria (and family)’s.’

Questioned objects can, like other objects, optionally be fronted, although this is not common. Other questioned participants (e.g. addressee, recipient or beneficiary) are virtually never fronted, with speakers disagreeing as to whether such fronting is in fact possible. Questioned subjects (7.56) and fronted objects are nearly always explicitly focused by a following *mak* (see section 7.8); speakers vary in their judgments as to whether such focusing is obligatory or just strongly preferred.⁸

Quantity is questioned by *hira* ‘how many, how much’, which fits into the same slots as numerals, and takes a numeral classifier when enumerating humans (7.61). There is no general term to question duration (‘how long’); instead one guesses which units of time would be appropriate, and asks how many of these units it took (e.g. *fulan hira* ‘month how.many’ = ‘how many months’, *loron hira* ‘day how.many’ = ‘how many days’).

⁷ However *sá* is used much more widely in Tetun Terik and in the church register.

⁸ A common questioned subject is of course *sé* ‘who’; it is possible that the strong preference for following it with *mak* is to avoid ambiguity with *se* ‘if’.

- (7.61) *Iha ema nain hira mak halo tós?*
 EXIST person CLS:human how.many FOC make garden
 ‘How many people are there working the garden?’

Questioned adverbial constituents vary in their placement within the clause. Words questioning reason (e.g. *tansá* ‘why’) are nearly always fronted, as is *bain-hira* ‘when (future)’. Location questions use prepositional phrases incorporating the pronoun *nebé* ‘where’ (e.g. *iha nebé* ‘LOC where’ = ‘where’); these appear never to be fronted. Other adverbial question words are freer in the choice between clause-initial and clause-final position, with there sometimes being a difference in meaning between the two positions (as for *nusá* ‘why; what’s up’, discussed below). Note that even though questioned adverbial constituents may be fronted, the answer is normally in the default clause-final position (7.65).

When a questioned adverbial constituent is fronted there is a strong preference for explicitly focusing it with a following *mak* (7.62) (see section 7.8).⁹ The exception is (*halo*) *oinsá* ‘how’, which readily occurs without *mak* (7.66) (7.70).

The adverbial question terms are as follows. All consist of two or three morphemes.

- *bain-hira* ‘when (future)’ (7.62) is literally ‘day-how.many’, where *bain* is a bound root.
- *hori-bain-hira* ‘when (past)’ (7.63) is literally ‘at.(past.time)-day-how.many’
- *tansá* ‘why’ is literally *tán-sá* ‘because-what’.
- *tanba sá* ‘why’ is literally *tán-ba sá* ‘because-to what’.
- *nusá* in clause-initial position means ‘why’ (7.65); it is not used in formal situations. Literally the term is *nu-sá* ‘like-what’, where *nu* is a bound root (which in Tetun Terik is a preposition meaning ‘like’).
- *nusá* in clause-final position means ‘what’s up, how is it’. It is a very general request to talk about a situation.
- *halo nusá*, usually in clause-final position, means ‘how’, asking about means or method (7.64). It literally means ‘do like-what’.
- *oinsá* in clause-initial position means ‘how, by what means/method’ (7.66) (7.70). It is literally *oin-sá* ‘face/type-what’.
- *oinsá* in clause-final position can have the same meaning as clause-initial *oinsá* (7.71), or can be a very general request to talk about a situation, meaning something like ‘how is it’ (7.67).
- *halo oinsá* ‘how, what (would one) do’ (7.68) is literally ‘do how’.

- (7.62) *Bain-hira mak Senhór mai?*
 when FOC Mr come
 ‘When will he (“Senhór”) come?’

⁹ There may be register differences in the usage of *mak*. For instance, in our corpus, all clause-initial *bain-hira* ‘when’ and *nusá* ‘why’ are followed by *mak* except for some examples from the Catholic Bible translation, in addition to elicited examples given to prove that *mak* is not obligatory.

- (7.63) *Senhora mai hori-bain-hira?*
Mrs come at.past-when
'When did you ("Senhora") come?'
- (7.64) *Né lés halo nusá?*
this tear do how
'How do you tear this (drink carton)?'
- (7.65) "*Nusá mak Leni tanis?*" "Nia tanis tán atu hán dosi."
why FOC Leni cry 3S cry because want eat cake
'Why is Leni crying?' "She is crying because she wants to eat cake."
- (7.66) *Oinsá ita bót sira haré ninia saúdi?*
how 1S.HON big PL see 3S.POS health
'How do you look after her health?'
- (7.67) *Imi nia servisu né oinsá?*
2P POS work this how
'How is your work?' (e.g. Is it good? Where do you work? What do you do?)
- (7.68) *Halo oinsá mak sira bele mai? La iha meius ida.*
do how FOC 3P can come not EXIST/have means one
'How could they come? They have no means (to do so).'

Indirect questions (surrounded in the examples below by square brackets) have the same format as main-clause interrogatives, including the preference for focusing fronted constituents with *mak*.

- (7.69) ... *ita bele hatene [sé mak irmán Ana].*
1PI can know who FOC sister Ana
'(I will review Ana's life so that) we can know who Sister Ana (really) was.'
- (7.70) *Nia buka hatene [oinsá buat né atu sai diak liu tán].*
3S seek know how thing this IRR become good further more
'He is trying to understand how this thing (growing rice) can be (done) even better.'
- (7.71) *Hau la hatene [problema né akontese oinsá].*
1S not know problem this happen how
'I don't know how this problem occurred.'

7.6.2 Polar and alternative questions and their answers

Polar ('yes–no') questions can be formed by just adding question intonation to a declarative clause (7.72), by adding the question tag *ka* (lit. 'or') (7.73), or *ka lae* 'or not'

(7.74),¹⁰ or by specifying both the positive and the negative options in full (7.75). While all these options create unbiased information questions, *ka* (but not *ka lae*) can also be used for critical rhetorical questions (7.76). Final *ka lae* is also used to create sentences which by their structure and intonation appear to be questions but which in fact state obvious truth and expect no answer (7.77).

(7.72) “*Nia la tanis?*” “*Tanis.*”
 3S not cry cry
 “Didn’t she cry?” “She did.”

(7.73) “*Apá iha né ka?*” “*La iha. Nia bá iha tós.*”
 Dad LOC this or not EXIST 3S go LOC garden
 “Is Dad here?” “He’s not. He went to the gardens.”

(7.74) “*Ita diak ka lae?*” “*Diak hela, obrigadu.*”
 2S.HON good or not good CONT thank.you
 “Are you well?” “I’m well, thank you.”

(7.75) *O atu bá rende ka o la bá rende?*
 2S IRR go surrender or 2S not go surrender
 ‘Do you intend to go and surrender, or not?’

(7.76) *Imi beik ka?!*
 2S stupid or
 ‘Are you stupid?!’

(7.77) *Sira hatene ona Bahasa Indonesia ka lae?*
 3P know ANT language Indonesia or not
 ‘They (high school students) already know Indonesian, don’t they?’

Questions with positive polarity can be answered by *sín* ‘yes’ (from Portuguese) or *lae* ‘no’; however it is more common to (additionally or instead) echo part of the question that agrees with the answer, including at least the auxiliary or predicator, as in (7.73) and (7.74). For negative polarity questions, the answer is normally an echo (7.72). When the question concerns whether something has already occurred, the negative reply is *seidauk* ‘not yet’ rather than *lae* ‘no’ (7.78).

(7.78) “*Sira bá tiha ona?*” “*Seidauk.*”
 3P go PRF ANT not.yet
 “Have they gone yet?” “Not yet.”

¹⁰ The question tags have variants *ga* and *ga lae*. In other contexts, however, the coordinator *ka* does not have a variant *ga*.

7.7 Commands and invitations

Most sentences used for commands have no overt indication of imperative mood. Rather, if the actor is the addressee(s), and the tense/aspect is consistent with this being interpreted as a command (e.g. by not specifying past time), then a sentence can be interpreted as some sort of request. The subject is frequently omitted (7.79), but need not be (7.80). The aspectual terms *lai* 'FIRST, before doing something else' (section 8.2.2.6) and *ona* 'ANT, already' (section 8.2.2.2) are quite common in commands, with *lai* being less insistent.

- (7.79) *Dada ís makás!*
 pull breath strong
 'Take a deep breath!'

Bá (lit. 'go') can be used at the end of a command or invitation for the addressee to do something without the speaker.

- (7.80) *Imi hán bá!*
 2P eat go
 'You (plural) eat up!'

Prohibitions normally use *la bele* 'not can' = 'cannot' (e.g. *La bele tama* 'not can enter' = 'Don't enter!'), or *lalika* 'need not'. The church register uses both *la bele* and the Tetun Terik term *keta* 'PROH, don't'. In Dili *keta* is much less common, with *la bele* being preferred by most speakers. All three options are further discussed in section 8.3.

Invitations to do something with the speaker can be prefaced by *mai ita* 'come 1PI'.

- (7.81) *Mai ita komesa ona.*
 come 1PI start ANT
 'Let's start.'

7.8 Focus (*mak*)

The focus marker *mak* (or *maka*, the longer form being preferred in writing, but not used much in speaking) usually focuses a fronted constituent, which it immediately follows. It can however, focus a following, non-fronted constituent. Finally, it is common in introducing the second of two clauses, with a meaning of 'and then, only then'. Each of these uses will be considered in turn.

Common focused constituents include fronted questioned constituents (see section 7.6.1) and relativised constituents. About a quarter of relative clauses with the relative clause marker *nebé* have *mak* immediately following *nebé*. Here *mak* means that the referent of the noun phrase is known, as illustrated by the contrast in (7.82). Note that in Tetun Terik, *mak* is itself the relative clause marker, with focus constructions such as those discussed below being interpreted as cleft constructions (van Klinken 1999a:318, 194). It is possible that this history of being a relative clause marker still affects the semantic interpretation of *mak* in Tetun Dili relative clauses.

- (7.82) *milagre sira nebé mak Moises halo*
 miracle PL REL FOC Moses do
 ‘the miracles that Moses did’ (we know about them, and could in principle count them and list them)

milagre sira nebé Moises halo
 miracle PL REL Moses do
 ‘miracles that Moses did’ (however many there were; we couldn’t list them)

Outside of relative clauses and questions, *mak* is usually contrastive, particularly for subjects and fronted objects. It often explicitly contrasts the referent of the focused constituent with other entities with which it forms a set, as illustrated by (7.83) and (7.84). Meanwhile, that which follows the *mak* is uncontroversial. Even where the contrast is not explicit, the implied contrast means that the focused constituent is interpreted as exclusive, as illustrated by examples (7.85) and (7.86).

- (7.83) *Hau mak hakarak, la iha ema ida obriga hau.*

1S FOC want not EXIST person one oblige 1S
 ‘It was I who wanted (to help); there was no-one (who) forced me to.’

- (7.84) *Ami kompriende balu-balu, barak mak ami la kompriende.*

1PE understand RDP-some much FOC 1PE not understand
 ‘We understand parts (of this newspaper article), but there is a lot that we don’t understand.’

- (7.85) a. *Hau mestri.*

1S teacher.MASC
 ‘I am a (male) teacher.’

- b. *Hau mak mestri.*

1S FOC teacher.MASC
 ‘I am the teacher.’

- (7.86) a. *Nia kík liu.*

3S small more
 ‘He is very small’ or: ‘He is smaller.’

- b. *Nia mak kík liu.*

3S FOC small more
 ‘He is the smallest.’

Although *mak* usually focuses the preceding phrase, it can also focus a following noun phrase. This is mainly found in equative clauses, where the initial noun phrase presents topical or otherwise known information, and the focused second noun phrase presents new information. In example (7.87), for instance, the *mak* in the question focuses the fronted question constituent, while the *mak* in the answer focuses the following answer. Non-fronted

objects can similarly be focused by a preceding *mak* (7.88). The positioning of *mak* before rather than after the focused constituent in these cases is presumably a result of word order constraints, since *mak* is unable to occur clause-finally.

(7.87) “*Ida nebé mak o nia uma?*” “*Hau nia uma mak neba.*”
 one which FOC 2S POS house 1S POS house FOC there
 ‘Which one is your house’ ‘My house is there.’

(7.88) *Hili mak nebé?*
 choose FOC which
 ‘Which (one) do you choose?’

Mak is also used to introduce the second of two clauses, with a meaning ‘and then, only then’. Here the clause following *mak* can readily introduce new information, but the *mak* still focuses on the preceding clause in that only after the situation described in the first clause has come about can that described in the second clause occur (7.89). *Mak* may similarly follow reason clauses, *depois* ‘afterwards, then, after that’ (*depois mak* being a common sequence meaning ‘and only then’), and time phrases (7.90). The latter (with a future time being understood) is presumably the source for leave-taking formulae with initial *mak(a)* (7.91). Some speakers further use sentence-initial *mak(a)* as a type of future time marker outside of leave-takings, but such usage is uncommon.

(7.89) *Kuandu diak ona mak foin fila; kuandu seidak diak, seidak bele fila.*
 when good ANT FOC JUST return when not.yet good not.yet can return
 ‘Only once we are well can we return (home from the healer’s house); if we are not yet well, we cannot yet return.’

(7.90) *Kinta bain-hitu mak nia ba.*
 Thursday day-seven FOC 3S go
 ‘It is next Thursday (7 days away) that she goes.’

(7.91) *Mak ita sei hasoru malu.*
 FOC 1PI FUT meet RECIPIENT
 ‘Until we meet again.’

7.9 Peripheral constituents

Time is specified by prepositional phrases (see section 6.6), adverbial clauses (section 10.6.2), noun phrases with a temporal noun as head (e.g. *momentu né* ‘moment this’ = ‘at that time’, *aban* ‘tomorrow’, *Domingu* ‘Sunday’), or temporal adverbs (e.g. *loro-loron* ‘RDP-day’ = ‘every day’). Clause-initial time generally presents a setting for what follows (7.92). Alternatively, it may be contrastive, and focused by *mak*, as is the case with *loron* in (7.93). When the time is new information, it tends to occur clause-finally (7.94, 7.95). Short temporal expressions can also occur between the subject and the verb, as per *kalan* in (7.93), and *loro-loron* in (7.94).

- (7.92) *Aban dia vinti oitu, hotu-hotu bele tún ona Dili.*
 tomorrow day twenty eight RDP-all can descend ANT Dili
 ‘Tomorrow, on the twenty-eighth day (of the month), everyone can go down to Dili.’
- (7.93) *Iha foho, mane sira kalan bá toba iha ailaran, loron mak mai uma.*
 LOC mountain man PL night go sleep LOC jungle day FOC come house
 ‘In the rural areas, at nights the men sleep in the jungles, and only come home during the day (because every night the military look for them).’
- (7.94) *Hau loro-loron tama tuku walu, sai tuku hát.*
 1S RDP-day enter o'clock eight exit o'clock four
 ‘Every day I start (work) at eight o'clock, and leave at four o'clock.’
- (7.95) *Ami hela iha neba tinan rua.*
 1PE stay LOC there year two
 ‘We stayed there two years.’

Peripheral location is specified by prepositional phrases introduced by *iha* ‘LOC’. It usually occurs after the clause core, that is, after any object or oblique arguments (7.96). It may however be initial, particularly if it provides information about the setting, such as *Iha foho* in (7.93). When location is questioned (with *iha neba* ‘LOC where’ = ‘where’) it appears to always be final.

- (7.96) *Hau fó alfabetizasaun ba povu sira iha neba, durante tréz anu-s.*
 1S give literacy to the.people PL LOC there during three year-PL
 ‘I taught literacy to the people there, for three years.’

8. Clausal modification

8.1 Introduction

In this chapter we look at those constituents of clauses and verb phrases which are not predicate heads, arguments, or peripheral constituents.

We look first at a range of tense–aspect markers, then at auxiliaries of permission and obligation (e.g. *bele* ‘can’), negation, intensifiers and the related means of forming comparatives and superlatives, and manner modifiers.

There are in addition other types of adverbs which will not be considered further in this grammar. Many adverbs are borrowed from Portuguese, including both well-known ones (e.g. *derepenti* ‘suddenly’), and adverbs which are used mainly by Portuguese-oriented speakers (e.g. *diretamenti* ‘directly’, *presizamenti* ‘precisely’). Those adverbs which are used to mark relationships between clauses are discussed in section 10.7. Adverbs of frequency (e.g. *kala-kalan* ‘RDP-night’ = ‘nightly’, *dala ruma* ‘occasion some’ = ‘sometimes’) tend to occur initially or preceding the verb. Some have fixed positions (e.g. *sempre* ‘always, often’ precedes the verb). Adverbs of approximation (e.g. *kala* ‘perhaps’, *besik* ‘almost’, *kuazi* ‘almost’) precede the head (e.g. *mais ou menus tolu-nulu* ‘more or less three-tens’ = ‘approximately thirty’).

8.2 Tense–aspect

8.2.1 Introduction

Tetun verbs are not marked for tense, aspect or mood. Temporal information can instead be specified through direct expressions of time, such as *aban* ‘tomorrow’, and *fulan Janeiru* ‘month January’ = ‘(in) January’. Once the hearer knows that the speaker is speaking about a certain time, that information does not need to be repeated, and subsequent sentences, when taken out of context, could usually be used equally well of any other time. Thus *Hau bá* ‘I go’, can, depending on context, be interpreted as ‘I went’, ‘I am going’, or ‘I will go’.

Many Tetun verbs do not carry the sorts of aspectual meanings that would normally be associated with their translation equivalents in other languages. For instance, stative verbs and adjectives such as *diak* ‘good’ can be interpreted either as indicating that the referent is in that state (‘be good’), or is entering that state (‘become good’). The same is true of activity verbs such as *túr* ‘sit’, which can be interpreted as either ‘sit down’ (i.e. start sitting) or ‘be

sitting'. Similarly, accomplishment verbs such as *oho* 'kill' and *hamoe* 'embarrass' are neutral with respect to whether the action has successfully reached its end-point (i.e. the victim is dead or feels embarrassed) or whether this point was not reached. When there is no explicit evidence to the contrary, one assumes that the victim of *oho* 'kill' is indeed dead. However most speakers insist that this expectation can readily be cancelled, as in example (8.1).

- (8.1) *Horiseik hau oho fahi ida, maibé nia halai liu...*
 yesterday 1S kill pig one but 3S ran further
 'Yesterday I killed a pig, but it ran away. (I don't know whether it lived or died.)'

Such distinctions can instead be revealed through a range of tense–aspect markers, discussed in this section. Such tense–aspect markers are not grammatically obligatory, and many sentences lack them. Some markers occur immediately before the verb, others occur after the verb, and yet others can be placed either after a transitive verb or clause-finally after its object. All these markers can have as their time of reference either the time of speaking ('now'), or the time that is being spoken about ('then'). So, for instance, *Hau hán hela* 'I eat CONT' can be interpreted as 'I was eating (at that time)', or 'I am eating (now)', or '(when) I will be eating'.

8.2.2 Focus on the completed event

8.2.2.1 *tiha* 'PERFECTIVE'

Perfective *tiha* (usually pronounced *tia*, although always written *tiha*) occurs in the post-verb slot. In statements with non-stative verbs it indicates that the activity of the verb has been completed (8.2). Frequently it occurs with transitive accomplishment verbs, indicating that the end-point of the activity has indeed been reached (8.3).

- (8.2) *Lapizeira monu tiha hosi meza.*
 pen fall PRF from table
 'The pen fell from the table.' or: 'The pen has fallen...'
- (8.3) *Sira oho tiha nia.*
 3P kill PRF 3S
 'They killed him. (He died.)'

With stative verbs or adjectives (with which *tiha* occurs only infrequently), *tiha* indicates that the subject referent has entered that state (e.g. *Nia diak tiha* '3S well PRF' = 'He became well, he was healed').

Tiha is also used (by some, though apparently not all, speakers) in commands and instructions (e.g. *Soe tiha deit!* 'throw PRF just' = 'Just throw (it) out!'), and in reported commands (8.4).

- (8.4) *Nia haruka sira túr tiha ba rai.*
 3S order 3P sit PRF to earth
 'He ordered them to sit on the ground.'

Some speakers, particularly from the south coast, commonly use *tiha* in backgrounded clauses (e.g. *Hán tiha hotu*, ... 'eat PRF finish' = 'After eating, ...'), in the expression *liu tiha* 'pass PRF' = 'after' (8.5), and in other expressions meaning 'after that', such as *né liu tiha* 'this pass PRF' and *hotu tiha* 'finish PRF'. This is in accord with its use in Tetun Terik.

- (8.5) *Liu tiha loron ida sira komesa fahe fós.*
 pass PRF day one 3P start distribute uncooked.rice
 'After a day, they started to distribute rice (to the refugees).'

When a clause with *tiha* is negated, the meaning is 'no longer'.

- (8.6) *Nia la iha tiha.*
 3S not EXIST PRF
 'She is no longer here.'

8.2.2.2 *ona* 'anterior'

The anterior aspectual marker *ona* 'ANT, already' either precedes or follows the object noun phrase in a transitive clause. In all other clause types, it immediately follows the predicate head.

For non-stative verbs, *ona* usually indicates that the event has already happened, and that its happening is still relevant. So, *Nia mai ona* '3S come ANT' = 'He has come' does not indicate when he came, but does imply that he is still here (or was still here at the time one is talking about); that is, the situation resulting from the completion of the action still holds. Some speakers also use *ona* in informal speech to mark past time in general.

For states, *ona* indicates that the subject referent has entered that state. There is no indication as to how or when this state began (e.g. *diak ona* 'well ANT' = 'be already well, have become well', *kaben ona* 'married ANT' = 'be married'). It also frequently follows time phrases, indicating that the specified amount of time has passed (e.g. *kleur ona* 'long.time ANT' = 'it has been a long time', *fulan hira ona?* 'month how.many ANT' = 'how many months have passed? / how many months old?').

- (8.7) *Ba sira nebé ke iha SMP ona, sira hatene ona bahasa Indonesia.*
 for 3P REL REL LOC junior.high ANT, 3P know ANT language Indonesia
 'For those who are in junior high school, they already know Indonesian (so we use that to teach them Portuguese).'

Ona can also be used in statements of what must be done, or in invitations. In this case it adds a sense of imminence ('do it now!') (e.g. *Mai hán ona!* 'come eat ANT' = 'Come and eat now!'). Some speakers use it (as per Tetun Terik) in clauses with future time reference to indicate that the event depicted in the clause is certain and imminent (e.g. *Ema atu fán hau*

ona ‘person IRR sell 1S ANT’ = ‘Someone will soon betray me’); others do not readily accept such use.

Ona frequently occurs in backgrounded conditional and temporal clauses (8.8) and in relative clauses (8.7).

- (8.8) *Kuandu iha eskola laran ona, Tetun né, ami la koalía.*
 when LOC school inside ANT Tetun this 1PE not speak
 ‘Once we were in school, we didn’t speak Tetun. (But back at home we did.)’

When a clause with *ona* is negated, the resultant meaning is ‘no longer, not again’; that is, it indicates that the situation described in the clause was true in the past, and could be expected to be true now, but isn’t.

- (8.9) *Servisu la iha ona.*
 work not exist ANT
 ‘There is no longer work.’

8.2.2.3 *tiha ona* ‘PERFECT’

The common sequence *tiha ona* indicates that an activity has finished and still has effect (i.e. it is perfect aspect) (8.10).¹ It is primarily (approximately 75%) used with intransitive clauses. It is not often used with states; when it is, the focus is on the entering of that state (8.11).

- (8.10) *Hau lalikan bá tán, tanba horiseik hau bá tiha ona.*
 1S need.not go again because yesterday 1S go PRF ANT
 ‘I don’t need to go again, because I’ve already gone yesterday.’

- (8.11) *kanek nebé diak tiha ona maibé sei iha fitar*
 wound REL good PRF ANT but still have scar
 ‘a wound that has healed but still has a scar’

Tiha ona is used relatively little in narrative. When it is, it tends to indicate that the events of the clause had already happened at the reference time (8.12).

- (8.12) ... *Tanba ami nia uma ahi hán hotu tiha ona.*
 because 1PE POS house fire eat all PRF ANT
 ‘(When we returned to Dili we moved into another house.) Because our house had been completely burned.’

¹ Eccles (1998) presents an insightful discussion of *tiha*, *ona* and *tiha ona*, based on an analysis of traditional stories published in Morris (1984a). He concludes that *tiha ona* focuses simultaneously on the completion of a process (as per *tiha*) and on the resulting new situation (as per *ona*).

8.2.2.4 *hotu* 'be finished'

Hotu is an intransitive verb meaning 'be finished' (8.13), as well as a quantifier meaning 'all' and an adverb 'also' (see section 5.3.6 for illustrations of both).

- (8.13) *Timór nia funu seidak hotu.*
 Timor POS war not.yet finish
 'Timor's struggle is not yet finished.'

It is also used immediately following non-stative durative verbs to indicate that the action specified by that verb has finished. (In all examples in the corpus, the initial verb is intransitive.)

- (8.14) *Ita koalia hotu mak ida fali koalia.*
 1PI speak finish FOC one instead speak
 'Only once we finish speaking does another takes a turn at speaking.'

It seems that *hotu* is still verbal in this position, and not fully grammaticised into a completive marker.² This is indicated by its ability to take tense–aspect marking of its own, such as *seidak* 'not yet' (8.15), *tiha* 'PRF' (e.g. *harís hotu tiha* 'bathe finish PRF' = 'after bathing'), perfect *tiha ona* (e.g. *halo hotu tiha ona* 'do finish PRF ANT' = 'had finished (it)'), and *lai* 'FIRST' (e.g. *Servisu hotu lai* 'work finish FIRST' = 'Finish working (for now; we'll continue later)').

- (8.15) *Sira koalia seidak hotu.*
 3P speak not.yet finish
 'They have not yet finished speaking.'

8.2.2.5 *foin* 'only just, very recently'

Foin 'JUST, only just, very recently' occurs in pre-verbal position. With punctual verbs it indicates that the activity has very recently been completed (e.g. *Nia foin tó* '3S JUST arrive' = 'He has only just arrived'). For states it indicates that the state has very recently been entered (8.16). For non-stative durative verbs (e.g. *Hau foin harís* '1S JUST bathe'), context determines whether *foin* is interpreted as indicating that the activity has just finished ('I have just bathed'), or that it is only just in progress ('I am just in the process of bathing'); the latter interpretation (which is less likely) can be forced by marking the verb as continuous (8.17). The sequences *foin atu* 'JUST IRR' indicates 'just about to' (8.18).

- (8.16) *Nia foin tinan lima.*
 3S JUST year five
 'She is only just five years old.' or: 'She has only just become five years old.'

² Development from a verb meaning 'finish' to a completive marker is common cross-linguistically (Bybee, Perkins and Pagliuca 1994:74).

(8.17) *Ami foin janta hela.*
 1PE JUST eat.dinner CONT
 'We are only just eating dinner.'

(8.18) *Nia foin atu sai.*
 3S JUST IRR exit
 'He is just about to leave.'

Foin is often used in a clause introduced by *mak*, where a preceding clause or temporal phrase specifies something that must happen first before that specified in the *mak* clause can or will happen.

(8.19) *Ami bá hela iha ne bá mak ami foin hela hakmatek.*
 1PE go stay LOC there FOC 1PE JUST stay peaceful
 'Only when we went and stayed there did we live in peace.'

It also occurs in several adverbial time expressions, including *foin daudauk né* 'JUST PROG this' = 'recently', and *foin agora* 'JUST now' = 'only now'.

8.2.2.6 *lai* 'FIRST, before doing something else'

Lai 'FIRST, before doing something else' occurs primarily with non-stative verbs, in either the post-verb or post-object slot, with no apparent difference in meaning. It indicates that the specified activity must be done, before some other (explicit or implicit) event occurs (8.20). It is mainly used in explanations, procedures, and the expression of future plans; it is not used for sequencing events in a narrative.

(8.20) *Hau husu tán ida lai mak ita bá hán.*
 1S ask as.well one FIRST FOC 1PE go eat
 'I'll ask one more (question), and then we'll go and eat.'

(8.21) *Ha-tún tiha labarik sira lai!*
 cause-descend PRF child PL FIRST
 'Make the children get out (of the car, before attempting the ascent).'

Lai is also commonly used in polite invitations, requests and commands; its politeness is in some contexts aided by the implication that the compliance need only be a temporary interruption before something else happens. For instance, *Ita para lai* '1PI stop FIRST' = 'We'll stop now' implies that after stopping, we'll later resume again, so that stopping is not such an imposition.

A negated clause with *lai* means 'not yet', and is appropriate for a temporary prohibition.

(8.22) *Ropa né la bele uza lai, tanba sei bokon hela.*
 clothes this not can use FIRST because still wet CONT
 'These clothes mustn't be used yet, as they are still wet.'

8.2.2.7 *kedas* 'IMMED, in advance, immediately'

Kedas (or *kedan*) occurs immediately after the verb, before the object (if any). It can indicate that the activity specified in the clause happens immediately after the reference time (8.23). For states, it indicates immediate entry into the state (e.g. *Nia diak kedas* '3S well IMMED' = 'He was immediately well/healed').

- (8.23) ...; *hanesan moras kík, bele bá kedas uma.*
 like sickness small can go IMMED home
 '(If one is very sick, one must stay with the healer a week); if the illness is minor, one can go home immediately.'

Alternatively, *kedas* can indicate that the activity is purposely done, in preparation for some other (sometimes implicit) event.

- (8.24) *Hau tein kedas nán, para ita mai fali, hán deit.*
 1S cook IMMED meat so.that 1PI come back eat just
 'I will cook the meat in advance, so that when we come back, we can just eat.'

Following a time phrase, *kedas* indicates 'that very time, no later than that' (e.g. *kalan né kedas* 'night this IMMED' = 'that very night', *uluk kedas* 'formerly IMMED' = 'even back then').

When a time phrase including the bound form *hori* 'at (past time)' is followed by *kedas*, the meaning is 'since that time, from that time until now'. For instance, *hori-uluk kedas* (lit. 'at-long.ago IMMED') means 'since long ago'; in contrast *hori-uluk* on its own (or, when its position in the clause allows, followed by *ona* 'ANT') means simply 'long ago'.

8.2.3 *Focus on the middle: continuity*8.2.3.1 *hela* 'CONTINUOUS'

Hela is an intransitive verb meaning 'reside, stay (in a place)' (e.g. *Sira hela iha Dili* '3P stay LOC Dili' = 'They live in Dili'). This notion of continuation carries over when it follows another verb as continuous aspect marker.³ With non-stative durative verbs, *hela*, like *daudauk* (see below), adds progressive aspect (e.g. *Sira hán hela* '3P eat CONT' = 'They are/were eating'). It also occurs with stative verbs and adjectives (e.g. *Hau diak hela* '1S well CONT' is a stronger affirmation of health than *Hau diak* '1S well' = 'I am well').

- (8.25) *Kuandu hau tó mai, nia halo hela kafé.*
 when 1S arrive come 3S make CONT coffee
 'When I arrived here, she was making coffee.'

³ We use this term following Bybee et al. (1994:126f). According to their definitions, 'progressive' aspect views an action as ongoing at the reference time. It is typically used for actions that require a constant input of energy to be sustained. 'Continuous' aspect is more general than progressive since it can be used with stative predicates as well as dynamic ones.

Apart from being used to indicate what is happening now, *hela* is frequently used in biclausal constructions in which the subordinate clause event is simultaneous with the main clause event. It is thus common in complements of verbs of perception (8.26), in ‘while’ clauses, and in main clauses where the simultaneous event is given by a time phrase or a subordinate clause (8.27).

(8.26) *Hau haré nia lao hela.*
 1S see 3S walk CONT
 ‘I saw her walking.’

(8.27) *Bain-hira hau haré nia, nia lao hela.*
 when 1S see 3S 3S walk CONT
 ‘When I saw her, she was walking.’

With punctual verbs, *hela* shifts the meaning from the activity to the resultant state. This occurs for intransitive verbs (8.28). It is also used on occasion to focus attention on the resulting state of a transitive verb, when the object is fronted and the subject is omitted; this gives a passive-like sentence (8.29).

(8.28) *Bero ida para hela iha neba.*
 boat one stop CONT LOC there
 ‘A boat was stopped/anchored there.’

(8.29) *Sira haré kuda ida kesi hela.*
 3P see horse one tie CONT
 ‘They saw a horse (which was) tied up.’

8.2.3.2 daudauk ‘PROGRESSIVE’

Daudauk (or *dadauk*, *dadaun*, *daudaun*) marks progressive aspect.⁴ It is found in the post-verb slot. It occurs with non-stative durative verbs which involve steady progress towards an end-point (e.g. *lao* ‘walk’, *hanorin* ‘teach’, *hán* ‘eat (a meal)’), and indicates that the activity is in progress at the reference time. Like *hela*, it can be used in biclausal constructions to indicate that the events of one clause are progressing at the same time as the events of the other clause.

(8.30) *Loro-matan sae daudauk ona.*
 sun-source ascend PROG ANT
 ‘The sun is already rising (i.e. it is part-way up).’

The contrast between progressive *daudauk* and continuous *hela* is illustrated by the alternatives considered when translating a story of Jesus’ disciples stripping wheat grains to

⁴ These forms appear to be derived by reduplicating *dauk* or *daun*. However the only other context for this root is in *seidauk* ‘not yet’ (where *sei* means ‘still’), and its rare synonym *la dauk* ‘not yet’.

snack on as they walked through a wheat field. In this context, *Sira korut hela trigo* ('3P strip CONT wheat') would simply indicate that at the time they were stripping grains of wheat; the natural assumption is that they did this in the one spot over a period of time. In contrast, *Sira korut daudaun trigo* ('3P strip PROG wheat') suggests that they were stripping the wheat grains while walking through the field; that is, there was progress towards an end-point (here, reaching the end of the field).

Common time expressions based on it are *agora daudauk (né)* 'now PROG (this)' = 'these days', and *foin daudauk (né)* 'JUST PROG (this)' = 'recently, a very short time ago'.

8.2.3.3 *sei* 'still', *seidauk* 'not yet'

Sei has two distinct meanings, 'still' and 'FUTURE, will' (discussed in 8.2.4.1 below).⁵ For both, it precedes the predicate head, which is usually a verb, prepositional phrase, or time phrase.

Sei 'still' occurs with states (including time periods such as *kalan* 'night') and non-stative durative verbs, indicating that the specified situation, having begun, still holds now, although it is expected to finish at some time in the future. It frequently co-occurs with continuous *hela* (8.32).

(8.31) *Agora nia sei moris. Nia sei iha Alas.*
 now 3S still live 3S still LOC Alas
 'He (the old man) is at present still alive. He is still in Alas.'

(8.32) *Nia sei hanorin hela.*
 3S still teach CONT
 'She is still teaching (at the moment).'

The negative of *sei* 'still' is *seidauk* 'not yet', which indicates that the situation has not yet begun, although it is expected to do so. *Seidauk* can occur with punctual as well as durative verbs and states.

(8.33) *Nia atu servisu maibé nia idadi seidauk tó.*
 3S IRR work but 3S.POS age not.yet suffice
 'He wants to work but isn't yet old enough.'

Unlike *sei* and other tense–aspect markers, *seidauk* can occur without any predicate head, if it can be understood from context what it is that hasn't yet started. It thus behaves grammatically like the auxiliaries of permission and obligation, such as *bele* 'can' (section 8.3).

(8.34) "Hudi né isin ona ka seidauk?" "Seidauk."
 banana this bear.fruit ANT or not.yet not.yet
 "Has this banana tree borne fruit yet?" "Not yet."

⁵ Eccles (1999:30) notes that a common component of both meanings is that the conclusion of the process is yet to be reached.

8.2.3.4 *nafatin* 'continue'

Nafatin 'continue, keep on' follows the verb and may either precede or follow any object or oblique argument. It occurs with both stative and active predicates, and indicates that an action or state which began earlier continues. It can indicate that the continuation is 'forever, perpetually', particularly in the church register.

- (8.35) *Ami fila mós, sira sei ajuda nafatin ami.*
 1PE return also 3P still help continue 1PE
 'Even when we were returning (from being refugees), they (the Indonesian church) still continued helping us.'
- (8.36) *Se CNRT hakarak lao ba oin nafatin, tenki halo restruturasaun...*
 if CNRT want walk to front continue must do restructuring
 'If the CNRT wants to continue to progress, it must restructure (itself)...

When *nafatin* occurs in a negative clause, it indicates that the negative situation continues (8.37).

- (8.37) ..., *maibé imi la komprende nafatin deit.*
 but 2P not understand continue just
 '(I have taught you over and over) but you continue to not understand.'

8.2.4 Future⁶8.2.4.1 *sei* 'FUTURE'

The second use of *sei* is to present something as definitely happening in the future. It is thus particularly appropriate in discussing future plans, and in promises, threats, and prophecies. Unlike *sei* 'still', it frequently occurs with punctual verbs.

- (8.38) *O kuandu halo hanesan né beibeik, ema sei baku o.*
 2S when do like this continually person FUT beat 2S
 'If you keep acting like this, someone/people will hit you.'

Nevertheless, out of context, ambiguity between 'still' and 'will' is possible with durative verb phrases. This can be resolved in favour of a 'still' interpretation by adding continuous *hela* or continuative *nafatin*, while a future interpretation is strongly encouraged by adding a future time phrase such as *aban* 'tomorrow', or serialising the verb to a preceding *bá* 'go' or *mai* 'come' (thus indicating that the motion is required before the activity can even begin).

⁶ Another word sometimes claimed to be a future marker is *ba* (which is also a preposition meaning 'to'). Eccles (1999:32f), for instance, identifies it as an 'approximative marker', in sentences like *Aban hau ba selu nia* 'tomorrow 1S go pay 3S' = 'Tomorrow I'm going to pay him.' We analyse such sequences as serial verbs in which the *bá* literally means 'go' (see section 9.2), because these constructions are used only when an actual act of going precedes the activity specified by the following verb. In practice, however, such constructions do often indicate futurity, by virtue of this fact that the going must precede the activity. Thus *Hau bá toba* '1S go sleep', for instance, means that I must go first and then sleep, thus putting the sleep into the future.

- (8.39) *Hau sei halo kafé.*
 1S still/FUT make coffee
 1. 'I'm still making coffee (so can't do anything else yet).'
 2. 'I will make coffee.'

The distinction between the two uses of *sei* is shown most clearly by the fact that they are negated entirely differently. The negation of *sei* 'still' is *seidak* 'not yet' (8.2.3.3), while the negation of future *sei* is *sei la* 'will not'.

- (8.40) *Ami sei la uza dolar.*
 1PE FUT not use dollar
 'We will not use (American) dollars (in future, even though the United Nations administration uses them now).'

8.2.4.2 *atu* 'IRREALIS'

Pre-verbal *atu* 'IRREALIS, want to, intend to, be going to, about to' usually indicates a desire, intention or imagination of a future event. It occurs almost exclusively with non-stative verbs (8.41), although it is possible with states as well (e.g. *antes atu bót* 'before IRR big' = 'before (they) are grown-up', *hakarak atu hatene* 'want IRR know' = 'want to know'). Unlike *sei* 'FUT', it is compatible with the future event not actually coming about (8.42). In fact, using *atu* in clauses with past time reference is a common way of marking counter-factuality (i.e. indicating that what was anticipated to happen did in fact not happen).

- (8.41) *O lori aikose mai; hau atu sunu ahi.*
 2S bring match come 1S IRR light fire
 'Bring the matches here; I'm about to light the fire.'
- (8.42) *Hau atu bá misa, maibé la biban ona, tanba misa tama tiha ona.*
 1S IRR go mass but not manage ANT because mass enter PRF ANT
 'I meant to go to (Catholic) mass, but can't do it any more, because mass has already started.'

Although *atu* usually indicates an intention, it can also mark immediate future, occurring with complements which are not necessarily desired, but which are nevertheless about to happen (e.g. *Nia atu mate* '3S IRR die' = 'He is about to die'). With this meaning of 'nearly', it does not require human (or agentive) subjects, and can also precede measures (e.g. *iha 1978, atu 1979* 'in 1978, almost 1979').

Atu is frequently used in purpose clauses, and in complements of a wide range of verbs (and nouns) which deal with volition (e.g. *hakarak* 'want', *lakohi* 'not want'), ability (e.g. *posibilidadi* 'possibility'), and fear (e.g. *barani* 'brave, dare'), amongst others. In most cases the *atu* is syntactically optional, retaining some of its meaning of 'want to, intend to'. While

it may be part-way towards becoming a semantically bleached complementiser, it is not at that stage now.⁷

Sometimes (much more commonly in writing than speaking), *atu* in purpose or complement clauses occurs in the pre-subject position characteristic of conjunctions and complementisers. This position is not permissible in main clauses (e.g. one cannot say **Atu hau bá Baucau aban* ‘IRR 1S go Baucau tomorrow’).

(8.43) *Atu ita nia ain la bubu, ita la bele hamrík kleur liu.*
 IRR 1PI POS leg not swell 1PI not can stand long more
 ‘In order that one’s legs won’t swell (during pregnancy), one must not stand up for long.’

(8.44) *Agora mós sira husu atu ita rona sira.*
 now also 3P ask IRR 1PI hear 3P
 ‘Now too, they ask that we listen to them.’

8.3 Deontic modality

8.3.1 Introduction

Deontic modality deals with permission and obligation. It is expressible in Tetun Dili by the auxiliaries *bele* ‘can’, *lalika(n)* ‘need not’, *keta* ‘PROHIBITIVE’ and *tenki* ‘must’. All typically immediately precede a verb. With the exception of *tenki* ‘must’, these auxiliaries can also occur without a verb if the context is understood (e.g. *Lalikan!* ‘no need (to do that)’).

8.3.2 *bele* ‘can, may’

Bele ‘can, may’ indicates ability (8.45) or permission to do something. Permissive *bele* can be used to soften imperatives (8.46).

(8.45) *Nia bele kura ema nia moras.*
 3S can cure person POS sickness
 ‘He can cure people’s illnesses.’

(8.46) *Ita bele fila ona.*
 2S.HON can return ANT
 (Doctor to patient:) ‘You may go now.’

There are several respects in which *bele* is like a verb and unlike other auxiliaries. Firstly, it is negated by *la* ‘not’. Negative *la bele* ‘can not, may not, must not’ is usually used in

⁷ The semantic distinction between complements with and without *atu* requires further research. As an initial approximation, some speakers suggest that *Hau hakarak bá Japaun* ‘1S want go Japan’ = ‘I would like to go to Japan’ carries no implication that I actually have any plan to do go to Japan. However, with the addition of *atu* (*Hau hakarak atu bá Japaun*), there is an implication that I am at least looking into the possibility.

prohibitions (e.g. *La bele fuma!* ‘not can smoke’ = ‘Don’t smoke!’) or more general statements of what should not be done (8.47). Less often, it indicates inability.

- (8.47) *Feto la bele lao kalan.*
 woman not can walk night
 ‘Women must not walk at night.’

A further similarity to complement-taking verbs is that it is possible to negate either *bele* (e.g. *la bele ba* ‘not can go’ = ‘cannot go’) or the rest of the clause (e.g. *bele la ba* ‘can not go’ = ‘it is possible/allowed to not go’). Finally, like *hatene* ‘know’ (section 10.5.3), it is possible in informal speech for *la bele* to follow its ‘complement’ when it is short (usually a single verb). In this case the construction necessarily indicates inability, or, rather, an inability to do something well (8.48). The contrast between pre-verbal and post-verbal *la bele* is illustrated by (8.49).

- (8.48) *Hau lao ikus liu, tanba hau lao la bele.*
 1S walk last more because 1S walk not can
 ‘I’ll walk last (in the group), as I can’t walk well.’

- (8.49) *Nia la bele halai.*
 3S not can run
 ‘She’s not allowed to run.’ or: ‘She is unable to run (e.g. due to a broken leg).’

Nia halai la bele.
 3S run not can
 ‘She cannot run well.’

Bele (and *la bele*) can occur without a main verb if the context makes sufficiently clear what it is that is possible or allowed.

- (8.50) *“Ita bôt nia oan bele kose nehan ka lae?” “Seidauk bele.”*
 2S.HON big POS child can brush tooth or not not.yet can
 ‘‘Can your child brush her teeth?’’ ‘‘(She) is not yet able.’’

In purpose clauses, future expectation is frequently marked by pre-verbal *bele*. This usage presumably comes about because the main clause indicates that the conditions for the situation in the purpose clause to be possible or permissible have been fulfilled.⁸

⁸ Purpose clauses are generally introduced by *para* ‘so.that’, *atu* ‘IRR’ or *hodi* ‘PURPOSE’. The sequences *para bele* ‘so.that can’ and *atu bele* ‘IRR can’ (8.51) are thus common in purpose clauses. We do not however analyse either of these as complex purposive conjunctions (as does Hull (1999) for *atu bele*), since *bele* can readily be preceded within the purpose clause by a subject (8.44, 8.52) or negator.

- (8.51) *Lokraik hau fila fali mai halo servisu ruma atu bele manán osan.*
 afternoon 1S return back come do work some IRR can win money.
 ‘In the afternoons I came back (from school) and did some work in order to earn money.’
- (8.52) ..., *ami né tenki koalia Portugés né, para ami bele hatene lai-lais.*
 1PE this must speak Portuguese this so.that 1PE can know RDP-quick
 ‘(At school), we had to speak Portuguese, so that we would learn it quickly.’

8.3.3 *lalika(n)* ‘need not’

Lalika(n) ‘need not, should not’ immediately precedes a verb. It is used primarily in advice-giving, with the clause always having a human subject (whether explicit or understood), and a time reference that is not in the past (being future, present, or generic advice). While *lalika* can mean ‘need not (but may if so desired)’, it is frequently used, particularly with a second person subject, to mean ‘should not, don’t’. Example (8.53), for instance, can be interpreted as ‘You need not go to school today, but may go if you wish’, but is more likely to be intended as ‘Don’t go to school today’.

- (8.53) *O ohin lalika bá eskola.*
 2S today need.not go school
 ‘Don’t go to school today.’ or: ‘You need not go to school today (but may go if you wish).’
- (8.54) *Pai! Ami fasi kareta ka lalika?*
 Dad 1PE wash car or need.not
 ‘Dad! Shall we wash the car or don’t we need to/have to?’

Lalika can also be used as a single-word utterance (optionally with tense–aspect markers such as *ona*) to mean ‘no need’, or ‘don’t!’, for instance in response to an offer of help.

8.3.4 *keta* ‘PROHIBITIVE’

The term *keta* may be used in prohibitions (e.g. *Imi keta tauk* ‘2P PROH afraid’ = ‘Don’t be afraid’). It is common in the church register, as it is in Tetun Terik. Outside of this register, it is much less common, with *la bele* ‘not can’ = ‘cannot’ being preferred by most speakers. It can be used without a following verb (e.g. *Keta lai!* ‘PROH FIRST’ = ‘Don’t (do it) yet!’).⁹

In addition, *keta* is an adverb meaning ‘perhaps’; with this meaning it occurs either clause-initially or after the subject, and is frequently paired with its synonym *karik*.

⁹ In the church register, *keta* also occurs in negative purpose clauses to mean ‘will not’, just as it does in Tetun Terik (e.g. *Haré bá, atu ema keta lohi imi* ‘see go, IRR person PROH deceive 2P’ = ‘Look out, so that people will not deceive you’).

- (8.55) *Keta nia iha merkadu karik.*
perhaps 3S LOC market perhaps
'Perhaps he is at the market.'

8.3.5 *tenki* 'must'

The auxiliary *tenki* 'must, have to' is a common Portuguese loan (from *tem que* lit. 'it.has that'). It usually indicates obligation (8.56), but can also be used in the sense of 'inevitable, of necessity' (8.57).

- (8.56) *Se o moras karik, o tenki hemu aimoruk.*
if 2S sick perhaps 2S must drink medicine
'If you are sick, you must take medicine.'
- (8.57) *Nia hanoin tiha ona katak, iha loron ida, Indonesia tenki tama Timór.*
3S think PRF ANT that LOC day one Indonesia must enter Timor
'He (a journalist) had already thought that one day Indonesia would inevitably enter Timor.'

Some speakers additionally use the loan *tenki sér* (from Portuguese *tem que ser*), which tends to indicate stronger obligation than does *tenki*.

8.4 Negation

There are two basic negators in Tetun, *la* and *laós*.

The (unstressable) general negator *la* 'not' negates a following verbal or adjectival constituent, to which it attaches phonologically (e.g. *la kleur* 'not long.(time)', *la mai* 'not come'). Several negative words consist of *la* followed by an apparent root whose meaning is not known; these include *laós* 'indeed.not' and *ladún* 'not very' (discussed below), *lalikan* 'need not' (section 8.3.3), and *lakohi* 'not want, refuse' (8.63). Speakers consider these to constitute single words. A few other particularly common sequences are felt by some, though not all, speakers to constitute single words (especially *la-diak* 'not-good' and *la-bele* 'not-can' = 'cannot, must not').¹⁰

Laós (or *lahós*, with either stress on the /o/, or relatively equal stress on the two vowels) negates a wide range of clause constituents, including subject, object, all types of predicates, time phrases, prepositional phrases, and dependent clauses. Like *la* (from which one would assume it is derived, except that **ós* and **hós* are not recognised as roots), it occurs immediately before the negated constituent. *Laós* is the only means of negating constituents other than verbal and adjectival ones. It is strongly contrastive (particularly before verbs and adjectives, where *la* is the default); the corresponding true option is usually explicitly

¹⁰ We have no evidence for *la...la...* being used to mean 'neither...nor...', as mentioned in Hull (1999).

presented, either immediately before or immediately after the negated constituent (8.58). The sequence *laós A deit, maibé B* ('not A just, but B') means 'not just A, but also B' (8.59).

- (8.58) *Hori-uluk kedan, ema Timór sira né hakarak ukun-an Duni.*
 at.past-formerly IMMED person Timor PL this want rule-self Indeed
 'Ever since the distant past, the Timorese have indeed wanted independence.'

Laós foin agora, lae! Hori-uluk kedan.
 Indeed.not JUST now no at.past-formerly IMMED
 Not just now, no! Ever since the distant past.'

- (8.59) ... *Laós nia deit, maibé ita nia irman N—*
 indeed.not 3S just but 1PI POS sister (name)

ho nia oan nain lima.
 and 3S.POS child CLS:human five

'(She was a martyr for the cause.) Not just her, but also our sister N— and her five children.'

Ladún 'not very', precedes the adjective or verb which it modifies, and reduces its force (e.g. *ladún bót* 'not.very big').

Lae 'no' is a negative pro-clause. It is commonly used as the second alternative in yes–no questions (e.g. *Diak ka lae?* 'good/well or no' = 'Are (you) well?' or 'Is (it) good?'), and after the conjunction *se* 'if', with *se lae* 'if no' meaning 'otherwise, else'. *Lae*, or more commonly *lae ida*, is also used as a negative answer to a yes–no question, to contradict a previous statement, or to reinforce a previous negative statement that uses the forceful negator *laós*, as in the second sentence of (8.58).

Nunka 'never' (from Portuguese *nunca*) negates a following verbal or adjectival constituent (8.60). The less common *nunka-mais* 'never ever, never before' (from Portuguese *nunca mais* 'never again') is stronger (8.61).

- (8.60) *Servisu nain né nunka deskansa.*
 Work master this never rest
 'This diligent worker never rests.'

- (8.61) *Ita nunka-mais koalía nuné!*
 1PI never.ever speak like.this
 'We never talk like this!' (saying /tiha/ as per the usual spelling, instead of /tia/ 'PRF')

In speaking (but never in writing) negation is very often paired with an *ida* (lit. 'one') after the predicate. This is most common with *la* (8.62), but also occurs with other negative terms such as *lakohi* 'not want' (8.63), *nunka* 'never' (8.64), and *seidauk* 'not yet'. This *ida* is incompatible with tense–aspect marking (e.g. *ona* 'ANT', *tiha* 'PRF', *atu* 'IRR'). It occurs either with a single-word verbal or adjectival predicate, as in (8.62), or at the end of a post-

verbal argument (object noun phrase or post-posed subject of *iha* 'exist'), as in (8.65). In the latter case, the *ida* can normally be interpreted literally as 'one, a'. It is however possible for *ida* to occur in the object of a negative clause even if it could not occur in the positive counterpart (8.66) It seems plausible that this common indefinite marking on the post-verbal argument of negative clauses is the basis for the use of *ida* in negative clauses even when there is no post-verbal argument. This construction is not found in Tetun Terik.

- (8.62) *Sira la tiru ida.*
 3P not shoot one
 'They didn't shoot (at us).'
- (8.63) *Hau lakohi bá ida.*
 1S not.want go one
 'I don't want to go.'
- (8.64) *Hau nunka mama Ida.*
 1S never chew.betel One
 'I never chew betel.'
- (8.65) *Imi la komprende buat ida.*
 2S not understand thing one
 'You don't understand anything.'
- (8.66) *Hau la bá Timór ida.*
 1S not go Timor one
 'I didn't go to Timor.'

8.5 Intensifiers, comparative, superlative

Intensifiers follow the verb, adjective, or verb phrase which they modify. The common intensifiers are listed in Table 8.1. There is considerable inter-speaker variation as to which intensifiers are preferred under which circumstances.

Table 8.1: Intensifiers

Intensifier	Meaning	Example
<i>demais</i>	'too (much), excessive' (from Portuguese <i>demais</i>)	<i>malirin demais</i> 'too cold'
<i>Liu tán</i>	'even more' (lit. 'more in.addition')	<i>diak liu tán</i> 'even better' (than before/others...)
<i>Liu</i>	'more, extremely, completely, most' (Also Vt 'pass')	<i>barak liu</i> 'very many, most, the majority'
<i>tebe-tebes</i>	'truly, extremely' (lit. 'RDP-truly') (primarily formal)	<i>susar tebe-tebes</i> 'very difficult'
<i>tebes</i>	'indeed, truly' (primarily formal)	<i>presiza tebes</i> 'really need'
<i>Lós</i>	'very, truly' (Also Adj 'straight, true, right')	<i>barak lós</i> 'very many'
<i>atu-mate</i> ¹¹	'very' (lit. 'IRR-die') (not formal; especially for negative qualities such as hungry, sad)	<i>susar atu-mate</i> 'very difficult'
<i>la-halimar</i>	'very, really' (lit. 'not play/joke') (informal)	<i>matenek la-halimar</i> 'really intelligent'

There is in addition an informal intensifier which precedes the term it modifies, namely *para* 'so' (e.g. *Nia isin para bót!* '3S body so big' = 'He's so fat!').

Comparatives and superlatives are formed with *liu* 'more, extremely, most'. For comparatives in which the subject is compared with something else, the object of comparison is usually introduced directly by *liu* (8.67). For more complex comparatives, such as stating that one action or situation is better than another, the object of comparison is introduced by *duké* (from Portuguese *do que*), as in (8.68).

(8.67) *Hau bót liu nia lori tinan nén.*
 1S big more 3S take year six
 'I am older than him by six years.'

(8.68) *Sira hakerek Tetun kapás liu duké koalia.*
 3P write Tetun beautiful more than speak
 'They (Australians) write Tetun better than they speak it.'

While *liu* is neutral with respect to the distinction between 'more, very' and 'most', a superlative interpretation can be forced by emphasising the uniqueness of the description. This can be because the description is polar rather than a gradient (e.g. *ikus liu* 'last more' = 'very last'; one cannot be 'more last'), or because context makes it clear that this is 'more' than all others. Such an interpretation can be encouraged by focusing either the entity or the description with *mak* (see section 7.8).

¹¹ This is usually written '*atu mate*', but the 'u' is normally not pronounced, leading to a pronunciation distinction between *tauk at-mate* 'afraid IRR-die' = 'very afraid, terrified' and *tauk atu mate* 'afraid to die'.

- (8.69) *Uluk né malária mak barak liu.*
 formerly this malaria FOC much more.
 ‘At that time, it was malaria that was the most (common disease brought to our clinic).’

When *liu* is used with negated verbs the resultant meaning is ‘not at all’ (e.g. *la hatene liu* ‘not know more’ = ‘not know at all’). Negated adjectives with *liu* can however instead mean ‘not completely, not overly’ (e.g. *la bokon liu* ‘not wet more’ = ‘not completely wet’).

8.6 Manner modifiers

Manner modifiers follow the verb they modify. Single manner adverbs tend to precede any object noun phrase (8.72), but they can follow it instead, as do longer manner modifiers (8.74).

Most manner adverbs are also classified as adjectives (e.g. *diak* ‘well, successfully; good’, *makás* ‘strongly, loudly, fast...; strong, loud, fast...’). Such adverbs (but not the reduplicated ones discussed below) can be modified by an intensifier (8.70; see section 8.5), or can be negated by *la* ‘not’. The only common example of such negation is *la diak* (lit. ‘not good/well’); this expression, however, indicates inability, rather than its literal translation of ‘not well’ (8.71a). (Ironically, ‘not well’ is translatable by a final *la bele*, literally ‘not can’ = ‘cannot’, as in (8.71b); see section 8.3.2).

- (8.70) *Padre sira kritika át liu!*
 Father PL criticise bad more
 The priests really criticise (that book)!’

- (8.71) a. *Nia halai la diak.*
 3S run not good
 ‘He cannot run (at all).’

- b. *Nia halai la bele.*
 3S run not can
 ‘He cannot run well.’

A few manner adverbs are derived by reduplicating an adjective (8.72). The most common is *di-diak* ‘carefully, thoroughly’ from *diak* ‘good’ (also an adverb ‘well, successfully’) (e.g. *Lao di-diak!* ‘walk RDP-good’ = ‘Walk carefully!’).

- (8.72) *Hau atende lai-lais nia.*
 1S attend RDP-quick 3S
 ‘I’ll quickly attend to her (my daughter who is wanting attention).’

Manner adverb phrases can alternatively be introduced by *halo* ‘do, make’ (8.73, 8.74).

- (8.73) *Tenki estuda malu halo di-diak.*
must study RECIP make RDP-good
'(Courting couples) must study each other well (to see if the other is suitable for marriage).'
- (8.74) *Sira tolok ami halo át lós.*
3P insult IPE make bad very
'They insulted us very severely.'

Alternatively, manner can be specified by a prepositional phrase introduced by *ho* 'with'.

- (8.75) *Nia koalia ho lian makás.*
3S speak with voice loud
'He spoke with a loud voice.'

9. Serial verbs and causative constructions

9.1 Introduction

Tetun Dili has a number of serial verb constructions, which have features common to serial verb constructions cross-linguistically.¹ They consist of two verbs which co-occur in a single clause, with a single intonation contour. The verbs share tense, aspect, negation and illocutionary force. Neither verb is subordinate to the other, nor modifies the other. Each verb shares at least one argument with the other. In particular, the subject of the second verb is interpreted as being identical to the subject, object or recipient of the first verb, depending on the construction. In most of the constructions, one of the two verb slots is restricted to a limited set of verbs, while the other is more open.

Some serial verb constructions are ‘nuclear’. That is, they are tightly-bound compound-like sequences, with the two verbs together forming a single clause nucleus; nothing can intervene between the two verbs, and the transitivity of the whole is determined by the first verb. The others are ‘core’ layer serial verbs, in which each verb can potentially introduce its own object, and adverbs can potentially occur between the two verbs.²

The common types of shared-subject serialisation in Tetun Dili are core motion–action serialisation, nuclear and core motion–direction serialisation, and core instrument serialisation. Causative constructions, in which the subject of the second verb is identified with the object of the first verb, are mainly formed with *halo* ‘make, do’ or *fó* ‘give’ as the initial verb. These are discussed along with morphological and periphrastic causatives in section 9.6.

In addition to having verb sequences that are analysed as serialisation, Tetun Dili also has a verb *hotu* ‘finish’ which functions as an aspectual marker (section 8.2.2.4), and *dehan* ‘say’ which serves to introduce direct quotes (section 10.5.2)

¹ Such features have been discussed in many publications, including Crowley (1987), Durie (1997), Foley and Olson (1985) and Schachter (1974).

² This distinction between serialisation within the clause nucleus and the clause core (Foley and Olson 1985; Foley and Van Valin 1984:77ff; Van Valin 1993:106ff) has been found helpful for the analysis of serial verbs in many other Austronesian languages also (e.g. Crowley 1987; Early 1993; Grimes 1991:391, 399).

9.2 Motion–action serialisation

A very common type of core-layer verb serialisation consists of *bá* ‘go’ or *mai* ‘come’ followed by another verb. This construction indicates that the motion happened first, followed by the activity denoted by the second verb, and generally in order that the activity of the second verb can be carried out. For instance, *Hau bá hán* ‘1S go eat’, if used of the past, could be translated as ‘I went and ate’ or ‘I went to eat (and in fact did eat)’. This construction is common because there is a strong tendency for Tetun Dili speakers to explicitly note motion towards or away from the speaker. For instance, *Nia sae bero* ‘3S ascend boat’ = ‘He got into the boat’ is appropriate if the subject was already more or less in position to get into the boat. If he had to go to the boat first, however, *Nia bá sae bero* ‘3S go ascend boat’ is more appropriate. Adverbs can follow either verb in this construction, as illustrated by *fali* ‘again, back’ in (9.1) and (9.2).

- (9.1) *Ami mai fali sae kareta.*
 IPE come again ascend car
 ‘We came back and got into the car.’

- (9.2) *Ami mai hán fali.*
 IPE come eat again
 ‘We again came and ate.’

9.3 Motion–direction serialisation

The direction of motion can be specified by a direction verb (*tún* ‘descend’, *sae* ‘ascend’, *tama* ‘enter’, *sai* ‘exit’) immediately following an intransitive motion verb (e.g. *halai sai* ‘run exit’ = ‘run outside’), in what is nuclear layer serialisation.

- (9.3) *Ami halai sae tó Dare neba.*
 IPE run ascend until Dare there
 ‘We ran up there to Dare.’

Direction of motion can also be specified by a following *bá* ‘go’ or *mai* ‘come’ (e.g. *tama bá* ‘enter go’ = ‘go in’, *tama mai* ‘enter come’ = ‘come in’), which in this use function as deictic markers indicating movement away from or towards the speaker respectively. These deictics are only loosely tied to the verb; they follow the direction verbs (e.g. *monu tama bá* ‘fall enter go’ = ‘fall in (away from speaker)’), any intervening adverbs (e.g. *fila fali mai* ‘return again come’ = ‘come back here’), and source location (9.4). They precede prepositional phrases denoting destination, however, as in (9.5).

- (9.4) *Nia monu hosi leten mai.*
 3S fall from top come
 ‘He fell from above.’

- (9.5) *Ami sae bá tó foho leten.*
 1PE ascend go until mountain top
 'We climbed up to the summit (away from here).'

The two deictics occur not only with intransitive motion verbs, but also with transitive verbs which indicate a change of location of the object, such as verbs of bringing or giving (e.g. *lori osan mai* 'bring money come' = 'bring money here'). In this case they indicate the direction in which the object referent is moving. They are further extended to verbs of speaking (e.g. *hatete mai* 'tell come' = 'tell me/us', *hatete bá* 'tell go' = 'tell him/her/them/you').³

Finally, there are a range of other transitive and intransitive verbs which occur in construction with a preceding verb of posture or motion, and which in some way further specify its direction, path or position.⁴ These include:

- *haleu* 'surround, encircle, around' (e.g. *Sira hamrík haleu ami* '3P stand surround 1PE' = 'They stood around us').
- *liu* 'pass' (e.g. *lao liu bero né* 'walk pass boat this' = 'walk past the boat').
- *liu hosi* 'via' (lit. 'pass from') (e.g. *lao liu hosi uma* 'walk pass from house' = 'walk via the house').
- *tuir* 'follow' (e.g. *lao tuir bé ninin* 'walk follow water edge' = 'walk along the shore').
- *hakat* 'cross' (e.g. *halai hakat estrada* 'run cross road' = 'run across the road').
- *hasoru* 'meet, oppose', *kontra* 'oppose' (from the Portuguese preposition *contra*) (e.g. *lao hasoru nia* 'walk meet 3S' = 'walk towards her'). These occur in construction with non-motion verbs also (e.g. *funu hasoru sira* 'war oppose 3P' = 'fight against them').

9.4 Instrumental serialisation

The transitive verbs *lori* and *hodi* both have as their primary meaning 'carry, bring, take', and both can be used in conjunction with a preceding or following verb phrase to introduce

³ Phrase-final *bá* and *mai* are thus less explicit alternatives to prepositional phrases introduced by the related prepositions *ba* 'to (not towards speaker)' and *mai* 'to (towards speaker)' (e.g. *hatete mai hau* 'tell come 1S' = 'tell me'). These are discussed in section 6.3.

In Tetun Terik, phrase-final *bá* and *mai* are also found in non-motion prepositional phrases (e.g. Fehan dialect *iha Dili bá* 'LOC Dili go' = 'over in Dili'). The closest equivalent in Tetun Dili are however the deictics *nebá* 'there' and *né* 'this' (e.g. *iha Dili nebá* 'LOC Dili there' = 'over in Dili').

⁴ That the second verbs in these constructions are verbs and not prepositions is shown by the fact that they can take manner modifiers, their complement is not required, and they cannot introduce a peripheral phrase.

This construction falls within the construction type referred to by Durie (1997) as 'coincident motion or posture' serialisation, in which the second verb involves an action undertaken while moving or maintaining a posture as indicated by the first verb.

the instrument.⁵ Many speakers prefer *lori* while some (particularly from the south coast) prefer *hodi*.

Where *lori* or *hodi* precedes the other verb, as in (9.6), the first verb introduces the instrument with which the action denoted by the second verb is performed. This is the order frequently found in instrumental serialisations cross-linguistically (e.g. Durie (1997), Foley and Olson (1985:53)).⁶ In this construction, however, *lori* and *hodi* appear not to be fully verbal, in that (particularly for *lori*, although speaker judgments differ), the object of the initial verb cannot readily be fronted or omitted. It is thus possible that both verbs are in the process of change towards becoming prepositions.

(9.6) *Sira lori tudik sona malu.*

3P take knife stab RECIP

'They use knives to stab each other.' or: 'They stab each other with knives.'

In the other order, illustrated by (9.7) and (9.8), the instrumental phrase occurs in the position typical of prepositional phrases in Tetun. Here *lori* and *hodi* appear to be fully prepositional, with their objects unable to be fronted or omitted (e.g. one cannot say **Tudik né nia sona karau lori* 'knife this 3S spear buffalo with').

(9.7) *Bibi tuku hodi dikur.*

goat butt with horn

'The goat butts with its horns.'

(9.8) *Sira funu lori fatuk.*

3P war use stones

'They fight with stones.'

There are three indications that *hodi* is further along the path of grammaticalisation towards being a preposition than *lori* is.⁷ Firstly, *lori* is used far more frequently than *hodi* as a main verb meaning 'carry, bring, take'. Secondly, the limited textual evidence suggests that *lori* occurs more frequently in the serial verb construction (as in (9.6)), while *hodi* more commonly occurs in the prepositional phrase position (as per (9.7)). Finally, it appears that *lori* can only introduce instruments that actually move (such as tools, arms and legs), while *hodi* can (at least for some speakers) also be used of seeing with one's own eyes (e.g. *haré hodi matan* 'see with eye') or hearing with one's own ears. Note that for the latter function, the indisputably prepositional *ho* 'with' is acceptable to all speakers.

⁵ A more common means of introducing instruments is the preposition *ho* 'with' (see section 6.5); in this case the normal position for the instrument phrase is following the verb (e.g. *hakerek ho lapizeira* 'write with (a) pen').

⁶ Note that while Durie, Foley and Olson present this order as the standard for instrumental serialisation cross-linguistically, there are some languages in eastern Indonesia, including Taba and Tidore, which have the opposite order (John Bowden, pers. comm.).

⁷ This discrepancy also occurs in the Fehan dialect of Tetun Terik. There, *lori* is a verb only, and is not used to introduce instruments, while *hodi* is analysed both as a verb, and (in instrumental constructions) as a prepositional verb (i.e. part-way between being a verb and preposition) (van Klinken 1999a:273ff).

9.5 Other shared-subject

There are less common serial verb sequences which do not fit the above categories. In these, two consecutive verbs share a subject (e.g. *Sira lao hakruk deit* ‘3P walk bend.over just’ = ‘They walked bent over’). If both are transitive, they also share an object (e.g. *lés fahe paun né* ‘tear divide bread this’ = ‘tear this bread and divide it’). The second verb may indicate a following action (as per *lés fahe* above) or a simultaneous one (as per *lao hakruk* above).⁸

9.6 Causation

9.6.1 Introduction

Causative constructions in Tetun Dili can be formed in many ways: with the derivational prefix *ha-*, with one of a number of serial verb constructions, or with a non-serialising periphrastic construction using *halo* ‘make, do’. We discuss each of these possibilities in turn.

9.6.2 *ha-* prefixation

As mentioned in 4.2.1, the prefix *ha-* is attached to root adjectives and intransitive verbs to form causative transitive verbs. The resulting derivation nearly always takes a person as subject, and an inanimate object, or an object referring to an animal, or to a person who is for some reason not able to resist (e.g. a prisoner being put into prison).

Vi/Adj		Vt	
<i>lao</i>	‘walk’	<i>ha-lao</i>	‘conduct (meeting, work, etc.)’
<i>mate</i>	‘die’	<i>ha-mate</i>	‘extinguish, turn off, deaden’
<i>mós</i>	‘clean’	<i>ha-mós</i>	‘clean, erase, peel, bathe’
<i>tama</i>	‘enter’	<i>ha-tama</i>	‘insert, cause to enter’
<i>tún</i>	‘descend’	<i>ha-tún</i>	‘lower’

- (9.9) ... *hodi ha-mate lai-lais tiha ahi.*
 to cause-die RDP-quick PRF fire
 ‘(Wrap the person who is on fire in a blanket) to quickly extinguish the fire.’

Speakers vary considerably in the number of roots to which they prefix *ha-*, and in whether it also derives verbs for involuntary activities (e.g. *aimoruk bele ha-menus moras né* ‘medicine can cause-less pain/sickness this’ = ‘medicine can reduce the pain’). This is largely

⁸ The example *lao hakruk* ‘walk bent over’ above, and the few other examples from the corpus which share the subject only, fit Durie’s (1997) ‘coincident motion or posture’ serialisation, in which the second verb involves an action undertaken while moving or maintaining a posture as indicated by the first verb. The transitive example *lés fahe* ‘tear divide’, and a number of others like it, are what Durie calls ‘synonymic serialisation’ where two verbs with similar meanings and identical argument structures are serialised. There is insufficient evidence to determine whether these two serialisation types would account for all the data in what we for now simply list as ‘other shared-subject’ serialisation.

due to the varying degrees of influence of Tetun Terik, in which *ha-* is far more prevalent. Those for whom *ha-* is productive apply it to Portuguese roots as well (e.g. *ha-menus* 'causeless' = 'reduce').

9.6.3 Constructions with *halo* 'make, do'

An alternative to using the prefix *ha-* is to use a nuclear serial verb construction in which the verb *halo* 'make, do' is immediately followed by a verb or adjective indicating what is caused. The object of this sequence of verbs follows the second verb.

(9.10) *Sira halo mate ikan.*

3P make die fish

'They killed the fish (by neglecting to feed them).'

For some roots (e.g. *badak* 'short', *bokon* 'wet'), speakers alternate freely between the two constructions (e.g. *ha-badak*, *halo badak*), with no apparent difference in meaning. For others (e.g. *foer* 'dirty'), some speakers accept a derivation with *ha-*, while others accept only the use of *halo*. For yet other roots, causation is only ever expressed by *halo* (e.g. *halo kole* 'make tired').

On the other hand, there are expressions in which verbs derived by *ha-* cannot be replaced with a construction involving *halo*. For instance, *ha-tama sasán* 'cause-enter goods' = 'bring in the bride price', is never expressed as **halo tama sasán*, nor can *hasai rán* 'cause-exit blood' = 'take blood (for a blood test)' be replaced by **halo sai rán*.

A third, periphrastic, causative construction allows considerably more syntactic and semantic possibilities. In this construction *halo* 'make, do' is followed by the object noun phrase, which is in turn followed by a predicate expressing what the object referent has been caused to do or be.

(9.11) *Moras AIDS halo ema barak mate.*

sick AIDS make person many die

'The AIDS illness has caused many people to die.'

Here the causative complement can be any verbal or adjectival predicate, can have the full range of arguments (except that the subject must be the object of *halo*), can be independently negated (9.12), and allows a wide range of adverbial and other modifiers (9.13). This construction is readily used of indirect causation, and of inanimate subjects being the cause of human reactions (e.g. 'the news made us happy', 'the poison made him sick'). Here the referent of the causee noun phrase (which is object of *halo* and subject of the causative predicate) is attributed some level of control; it must at least have sufficient control for the predicate to be properly predicated of it. Thus, while one can *ha-tama osan ba banku* 'cause-enter money to bank' = 'put money in the bank', one cannot **halo osan tama ba banku* 'make money enter to bank', as money cannot by itself *tama* 'enter' banks.⁹

⁹ The distinctions between the three causative constructions are thus consistent with Givón's (1980) binding hierarchy for manipulative verbs. That is, the syntactically more tightly-bound constructions (morphological *ha-*, and to a lesser extent serialised *halo*) are also those in which the actor exercises stronger influence, and the undergoer has less independence.

- (9.12) *Situasaun né mak halo ema Timór barak liu la koalia Portugés.*
 situation this FOC make person Timor many more not talk Portuguese
 ‘It is this situation (insufficient education for the common people) that caused most Timorese people to not speak Portuguese.’
- (9.13) ... *halo ita nia maluk Senhora Catharina mós konfundi ituan.*
 make 1PI POS friend Mrs Catharina also confused a.little
 ‘(The following statistical result) caused our colleague Mrs Catharina to be somewhat perplexed.’

9.6.4 Causation using *fó* ‘give’

Causative predicates can, for some verbs, be formed by nuclear serialisation with a preceding verb *fó* ‘give’. There is a strong implication that the causation was successful. Usually, the original subject of the second verb becomes the direct object of the resulting serialised construction (e.g. *fó hán bebé* ‘give eat baby’ = ‘feed (the) baby’), although some verbs allow *ba* ‘to’ to introduce this argument, as illustrated by the examples below.

Verb 2		Serialisation	
<i>barani</i>	Vi ‘brave, dare’	<i>fó barani (ba)</i>	‘embolden’
<i>hán</i>	Vt ‘eat’	<i>fó hán</i>	‘feed’
<i>harís</i>	Vi ‘bathe’	<i>fó harís</i>	‘bathe (someone)’
<i>hatais</i>	Vt ‘wear, dress’	<i>fó hatais</i>	‘dress (someone)’
<i>hemu</i>	Vt ‘drink’	<i>fó hemu</i>	‘give drink to’
<i>sala</i>	Vi ‘err’, N ‘error’	<i>fó sala ba</i>	‘accuse’
<i>susu</i>	Vt ‘suck’, N ‘breast’	<i>fó susu (ba)</i>	‘breastfeed’

In the examples above, the object of the second verb is unspecified. For instance, the expression *fó hán* ‘give eat’ = ‘feed’ does not allow one to indicate what type of food has been given. In some expressions, however, the object (e.g. of *aluga* ‘rent’ below) or complement (e.g. of *hatene* ‘know’ below) of the second verb functions as the object or complement of the serialised construction, with the subject of the second verb becoming an addressee or recipient introduced by *ba* ‘to’ or *mai* ‘come’, as in (9.14).¹⁰

Verb 2		Serialisation	
<i>empresta</i>	Vt ‘borrow’	<i>fó empresta O ba Rec.</i>	‘lend (out) O to Recipient’
<i>hatene</i>	Vt ‘know’	<i>fó hatene O ba Addr.</i>	‘inform of O to Addressee’
<i>aluga</i>	Vt ‘hire, rent’	<i>fó aluga O ba Rec.</i>	‘rent out O to Recipient’

¹⁰ In some cases it is not clear whether the word following *fó* is a noun or a verb. When it is a noun, however, it can be modified adjectivally, and the recipient is necessarily introduced by *ba* ‘to’ or *mai* ‘come’ (e.g. *fó sala barak ba nia* ‘give error many to 3S’ = ‘accuse him of many things’).

- (9.14) *Ema f́o hatene ba sira seluk dehan “Iha nemai.”*
 person give know to 3P other say LOC here
 ‘The person informed the others, “Here (it is).”’

Yet another pattern is observed for *sai* ‘exit’, as its subject becomes the object of the sequence *f́o sai* ‘reveal (something to someone)’, and the addressee phrase specifies who the information is revealed to.

In a few instances, the addition of *f́o* to a transitive verb makes no apparent difference in meaning.

Verb		Serialisation	
<i>bandu</i>	Vt ‘prohibit’, N ‘prohibition’	<i>f́o bandu</i>	‘prohibit’
<i>kastigu</i>	Vt ‘punish’	<i>f́o kastigu</i>	‘punish’

Sequences of *f́o* and a verb have the character of a compound, in that they are standardised expressions which are lexically restricted (e.g. ‘put to sleep’ is expressed by *ha-toba* ‘cause-sleep’, not **f́o toba*), and in that the two words cannot be separated. Nevertheless native speakers feel them to be two separate words.

There is a second, looser, construction in which *f́o* ‘give’ is in core-layer serialisation with a transitive verb. Here *f́o* is followed by a direct object specifying what is given, then a recipient introduced by *ba* ‘to’ or *mai* ‘come’, and finally by a transitive verb specifying what the recipient was to do to the object received. The second verb specifies purpose, and there is a strong implication that the purpose was indeed carried out.

- (9.15) *Nia f́o bé ba labarik hemu.*
 3S give water to child drink
 ‘She gave the child water to drink (and he drank it).’

In contrast to the nuclear serialisation construction discussed above, there appear to be no lexical restrictions on the second verb in this core-layer serialisation. Another difference is that the recipient of *f́o* ‘give’ is presented as the agent of the second verb. For instance, the nuclear serialisation *f́o hán* ‘give eat’ is interpreted primarily in terms of feeding someone (e.g. an infant) who cannot feed themselves. In contrast, the core layer serialisation *f́o paun ba nia hán* ‘give bread to 3S eat’ is interpreted primarily in terms of the recipient themselves eating the proffered bread.

9.6.5 Resultative serial verbs

In addition to serialisation with *halo* ‘make, do’ and *f́o* ‘give’, there is a third set of serial verbs for which the subject of the second verb corresponds to the object of the first verb.¹¹

In these, it is the second slot that appears to be lexically restricted; this verb specifies the result of the activity described by the first verb. Verbs that can fill the second slot in nuclear-

¹¹ These serialisations are called ‘cause-effect’ serialisation by Durie (1997:331).

layer resultative serialisation include the directionals (e.g. *duni sai nia* 'chase exit 3S' = 'chase him out', *soe sae* 'throw ascend' = 'throw upwards'), and *mate* 'die, dead' (9.16).

- (9.16) *Se o bosok tán dala ida, ami tiru mate kedan o agora.*
 if 2S lie again occasion one 1PE shoot dead IMMED 2S now
 'If you lie once more, we'll shoot you dead immediately.'

Intransitive *hela* 'stay, reside' follows transitive transfer verbs (e.g. of giving, putting, leaving, or throwing), to indicate that the object remains in the new location, at least temporarily (9.17).

- (9.17) *Lori hela Senhora ba Delta.*
 take stay Mrs To Delta
 (Instruction to a taxi driver:) 'Take this lady to Delta.'

Alternatively, an intransitive verb can follow the object of the initial verb, in core-layer serialisation (e.g. *duni sira sai* 'chase 3P exit' = 'chase them out'). In this case, the second predicate can take adverbial modifiers or complements (e.g. *bobar nia sai mate* 'encircle 3S become dead' = '(the python) wound around him until/and he died').

10. Complex sentences

10.1 Introduction

This chapter deals with structures beyond the clause. The first section deals with topic noun phrases. This is followed by discussion of the various means of linking clauses, including parataxis, coordinators, complementation, conjunctions, other connectors, and the embedding of relative clauses within a noun phrase.

10.2 Topics

A noun phrase can be topicalised by placing it before a clause. There is nothing in the sentence to indicate the semantic role of the noun phrase within the following clause or clauses. Although some topics are marked by final *né* ‘this’, as in (10.2), many are recognised as topics only by their position (10.1).

The topic is usually coreferential with a noun phrase within the following clause, resulting in what is known as a left-dislocation construction. The most common is coreference with the subject, which is double-underlined in (10.1). Alternatively, the topic can be coreferential with the object, or with the possessor of the subject.

- (10.1) *La oras, primu ida naran Mateus, nia telefoni fali mai.*
not time cousin.MASC one name Matthew 3S telephone back come
‘In no time, a cousin named Matthew—he rang us back.’

It is however also possible for the topic to not be coreferential with any noun phrase within the clause. In this case, topics “limit the applicability of the main predication to a certain restricted domain”, and set a “spatial, temporal or individual framework within which the main predication holds” (as proposed by Chafe (1976:50) for ‘Chinese-style topics’). Other means of specifying such initial topics are to introduce them by *ba* ‘as for’ (section 6.3) or *kona-ba* ‘concerning’ (section 6.7).

- (10.2) *Ami ema rua-nulu né, tau uluk mak hau nia naran.*
IPE person two-tens this put first FOC 1S POS name
‘(Of) the twenty of us, it was my name that was registered first.’

10.3 Parataxis

In formal speech and in writing it is usual to explicitly mark the relationships between clauses by conjunctions and other connective words. In everyday speech, however, two clauses can be simply juxtaposed, with non-final intonation on the initial clause, and without requiring any explicit indication of the semantic relationship between the two clauses. The most common relation is for the first clause to present a condition or time frame for the second ('if', 'when', 'while'). For explicit means of marking such relations see sections 10.6.2 and 10.6.5.

- (10.3) *Hau nia oan tanis, hau laran moras la-halimar.*
 1S POS child cry 1S inside sick not-play
 '(When) my child cries, I am truly upset.'

A second relation shown by juxtaposition is coordination. This occurs for clauses and verb phrases, but is rare for noun phrases. It is illustrated by example (10.10) in section 10.4.2, along with explicit means of indicating coordination.

A third relation between juxtaposed clauses or verb phrases is that the second has a similar meaning to the first, as do *respeita ema* 'respect people' and *konsidera ema* 'consider people' in the following example.

- (10.4) ... *sira bele respeita ema, konsidera ema, hanesan mós sira nia án.*
 3P can respect person consider person like also 3P POS self
 '(We instruct the children while they are young, so that later) they will respect others—show consideration to others—as they do to themselves.'

The following example shows several relationships at once, namely concession (marked by *mós* 'also', which is discussed in section 10.6.6), condition, and repetition.

- (10.5) *Nia bele bót bá mós, nia halo sala ruma, ita tenki hanorin nia,*
 3S can big go also 3S do wrong some 1PI must teach 3S

hanorin di-diak.
 teach RDP-good.
 'Even (when) she gets older, (if) she does something-or-other wrong, we must instruct her—instruct (her) well (i.e. not just beat her).'

10.4 Phrasal and clausal coordination

10.4.1 Introduction

There are three coordinators used to mark conjunction ('and'), and two for disjunction ('or'). These differ partly in the social contexts in which they are used, and partly in their syntactic function. In addition, certain constituents can be coordinated simply by juxtaposition.

10.4.2 Conjunction

For conjunction, *ho* (also a preposition meaning ‘with’, discussed in 6.5) primarily coordinates noun phrases (10.6), while *i* (often spelled ‘e’ as per its Portuguese source *e* ‘and’) primarily coordinates clauses and sentences (10.7). The coordinator *i* also occurs in Portuguese numerals to introduce the final digit (e.g. *setenta i três* ‘seventy and three’ = ‘seventy-three’). For coordination of predicates, *ho* is preferred for coordination of adjectives (e.g. *mean ho mutin* ‘red and white’) and other stative predicates, while *i* is used if the second predicate refers to an activity that occurs after the first (‘and then’) (10.8).

The coordinators *ho* and *i* are both used in everyday speech, and many authors also use them in writing. In contrast, Tetun Terik *no* is widely used (often to the total exclusion of the other coordinators) in church and in formal settings, but is rare in everyday speech in Tetun Dili. *No* coordinates all types of constituents (e.g. *Marta no Maria* ‘Martha and Mary’).¹

(10.6) *Dala ruma ami bá naok aifarina ho batar iha tós.*
 occasion some IPE go steal cassava and maize LOC garden
 ‘Sometimes we went and surreptitiously took cassava and corn from the gardens.’

(10.7) *I ema husu, “Sé mak hakarak hili autonomia,*
 and person ask who FOC want choose autonomy

I sé mak hakarak hili ukun rasik an?”
 and who FOC want choose rule own self
 ‘And people asked, “Who wants to choose autonomy, and who wants to choose independence?”’

(10.8) *Sira kaer nia i baku nia.*
 3P seize 3S and beat 3S
 ‘They seized him and beat him.’

For coordination of more than two items, the coordinator is typically placed only before the final term (10.9). A less common alternative is to place it between each of the elements in the list.

(10.9) *Dadér, loraik ho kalan, ami bele hanorin nia.*
 morning afternoon and night IPE can teach 3S
 ‘Morning, afternoon and evening, we can teach her (how to eat properly).’

Clauses and verb phrases can also be coordinated simply by juxtaposing them, particularly in speech (10.10). Such coordination by juxtaposition is rare for noun phrases. However two commonly-coordinated terms from a single semantic set may be juxtaposed within a compound (e.g. *inan-aman* ‘mother-father’ = ‘parents’; see section 4.4.3).

¹ The sequences *ho mós* and *no mós* ‘and also’ are felt by some speakers to be single words, and are written accordingly. In contrast *i mós*, combining the Portuguese coordinator with a Tetun adverb, is always considered to be two words.

- (10.10) ... *para atu selu eskola, sosa kadernu, sosa buat ruma.*
 so.that IRR pay school buy notebook buy thing some
 ‘(I worked after school to earn money) to pay for school, (and) buy notebooks, (and) buy various things.’

‘Both A and B’ can be expressed in informal speech (but not in formal speech or writing) by placing *mós* ‘also’ in two parallel clauses, immediately following the constituent that is varied. This construction appears to be more common amongst speakers whose origins are from the south coast of Timor.

- (10.11) *Ita koa halo kikoan mós diak, halo bo-bót mós diak.*
 IPI cut make tiny also good make RDP-large also good
 ‘Whether you cut (the meat for *vifi*) into small pieces or large pieces—both are fine.’

10.4.3 Disjunction

The two coordinators for disjunction (‘or’) are *ka* and *ou*. The native Tetun term *ka* is by far the more common in speaking, and is also used in the church register. The Portuguese loan *ou* (variously pronounced /ou/, /o/ or /u/) is more common than *ka* in writing (other than church documents); when used in informal speaking it reflects heavy Portuguese influence.² Both can coordinate a wide range of constituents, including noun phrases, adjectives, predicates and clauses.

Ou is only used for closed lists of options, with the *ou* preceding either each non-initial option or only the last, and with a possible pause before the *ou* (10.12). In contrast, *ka* can be used for closed or open lists, and the pause (if any) follows the *ka*. For a closed disjunction, *ka* has the same position as *ou* (e.g. *Tetun ka Portugés* ‘Tetun or Portuguese’). However in an open disjunction, it follows the final option as well, as in (10.13). Yes–no questions use *ka lae* ‘or not’ or simply a final *ka* (see 7.6.2).

- (10.12) *Ema Japaun sira uza Inglés ou sira lori rasik sira nian intérprete*
 person Japan PL use English or 3P bring own 3P POS interpreter

ou tradutór.

or translator

‘(For diplomatic relations) Japanese people use English or they bring their own interpreter or translator.’

- (10.13) ... *hanesan basa nia ka, baku ho sintu ka.*
 like slap 3S or beat with belt or
 ‘(If she keeps misbehaving, we can use some force) such as slapping her, or beating her with a belt (or whatever).’

² Some people whose Tetun is strongly influenced by Portuguese also use the construction *ou A ou B* ‘either A or B’.

10.5 Complement clauses

10.5.1 Introduction

This section deals with constructions in which verb phrases, clauses or sentences function as arguments. We will for convenience follow Dixon's (1991) semantic classification of such verbs, while using the term 'complement-taking predicate' (Noonan 1985:43) to distinguish the main verb from the verb in the complement.

There are two types of complement clause in Tetun, which we will label by Noonan's (1985:49, 73) terms 'sentential' (or 'sentence-like') and 'reduced'.³

10.5.2 Sentential complements

Sentential complements have the full range of possibilities of a clause, including independent mood (e.g. interrogative, declarative), negation, and subject. The time reference may be independent of that of the complement-taking predicate (e.g. 'say that...'), or dependent on it (e.g. 'see someone come', 'hear her speak'), depending on semantics. Sentential complements are introduced by *katak* 'that', *para* 'so that', or have no complementiser.

The complementiser *katak* 'that' introduces sentential complements for a wide range of verbs, including verbs of speaking (e.g. *dehan* 'say' (10.16), *hatete* 'tell, say', *hanorin* 'teach'), thinking and knowing (e.g. *hanoin* 'think', *hatene* 'know', *fiar* 'believe'), and attention (e.g. *haré* 'see' (10.14), *rona* 'hear', *senti* 'feel', *hatudu* 'show' (10.15)). Some of the verbs of thinking and knowing can alternatively take reduced complements; in this case the sentential complement can generally be translated by a complement introduced by 'that', and the reduced complement by an infinitival clause (e.g. *hanoin katak...* 'think that...', *hanoin (atu) bá* 'think IRR go' = 'remember to go, think to go'). For verbs of speaking, *katak* usually introduces indirect quotes (10.16), but can for some speakers introduce direct quotes as well (10.22).

(10.14) *Ami haré katak ajuda umanitariu ba hahán la kleur tán hotu ona.*
 IPE see that help humanitarian for food not long more finish ANT
 'We see that the humanitarian food aid will soon be finished.'

(10.15) *Sira hatudu mai ita katak feto mós bele sai asuain.*
 3P show come IPI that woman also can become hero
 'They showed us that women too can become national heroes.'

(10.16) *Sira dehan katak ami tenki koalia Portugés.*
 3P say that IPE must speak Portuguese
 'They said that we had to speak Portuguese.'

Katak is also a verb meaning 'signify, mean' (10.17).⁴

³ There is no evidence for nominalisation as a complementation strategy in Tetun.

⁴ In the Fehan dialect of Tetun Terik *katak* is a verb meaning 'say', but in Dili it has been completely grammaticised as a complementiser. In the church register *katak* is also used to mean 'that is, namely' (e.g.

- (10.17) *Né katak sira bele hetan fatin rezervadu iha CNRT.*
 this signify 3P can get place reserved LOC CNRT
 ‘This means that they can get a reserved place in the CNRT (the National Council of Timorese Resistance).’

The purposive conjunction *para* ‘so that’ (see section 10.6.4) optionally introduces sentential complements of some verbs, including verbs of wanting.

- (10.18) *Hau hakarak para nia diak.*
 1S want so.that 3S good
 ‘I want her (my daughter) to be good.’

Sentential complements without complementisers occur for direct quotes (see below), and for verbs of thinking and knowing. (In the following examples, square brackets surround the complements.)

- (10.19) *Ita la hatene lós [ita koalia ho sé mak ida né].*
 1PI not know true 1PI talk with who FOC one this
 ‘We didn’t know for sure who it was we were talking with (as it was so dark).’

The complement is the final element in the matrix clause. It thus follows addressee noun phrases and prepositional phrases (10.15). In this it is unlike noun phrase objects, which precede prepositional phrases.

As Noonan (1985:130) points out, attention verbs, when these are used as verbs of immediate perception (e.g. ‘see X’ as opposed to ‘see that X’), take complements with determined time reference, and so have limited aspectual possibilities. This follows naturally from the fact that one can only perceive what is happening at that particular time (e.g. *haré nia mai* ‘see 3S come’ = ‘saw him come’).

Direct quotes, which are of course less integrated into the clause than are indirect quotes, are often introduced directly by a verb of speaking (10.20). Alternatives are to introduce the quote by *hanesan né* ‘like this’, by the verb *dehan* ‘say’ (10.21), or by *dehan katak* ‘say that’ (10.22). While in speech the verb of speaking always precedes the quote (10.20), in journalistic writing it may be placed (without a complementiser) after the quote instead (10.23).

- (10.20) *Nia hatete, “Hei! O keta kaer hau!”*
 3S tell hey 2S PROH hold 1S
 ‘He said, “Hey! Don’t hold me!”’
- (10.21) *Liurai né bá hatete ba malae dehan “Ó,...”*
 ruler this go speak to foreigner say oh
 ‘The ruler went and spoke to the foreigners, saying “Oh, ...”’

Páskua, katak judeu sira nia festa ‘Passover mean Jew 3P POS festival’ = ‘Passover, that is, the festival of the Jews’).

(10.22) *Nia husu ba ami dehan katak*, “Nusá? Imi sai ka lae?”
 3S ask go 1PE say that “why 2P exit or not
 ‘He asked us, “What’s up? Are you leaving (Dili) or not?”’

(10.23) “*Bá haré tok*” *katuas haruka*.
 Go see try old.man order
 “‘Go and have a look,’ the old man ordered.’

10.5.3 Reduced complements

Reduced complements consist of a verb phrase, which cannot have an independently specified subject, but can be independently negated.

Reduced complements have a time reference which is restricted relative to that of the complement-taking predicate, in that it must be future or co-temporal (not past). As such the most common tense–aspect marker in reduced complements is irrealis *atu* (discussed in 8.2.4.2), although for most complement-taking verbs it is more common to have no tense–aspect marking in the complement at all.

Reduced complements (surrounded by square brackets in the examples below) are found for a wide range of verbs, including verbs of thinking and knowing (e.g. *hatene* ‘know (how to / whether to)’), wanting and liking (e.g. *hakarak* ‘want (to)’, *lakohi* ‘not want, refuse’, *tauk* ‘afraid (to)’, *gosta* ‘like’), trying (e.g. *buka* ‘seek (to)’, *esforsa-an* ‘strive (to)’, *la biban* ‘not manage (to)’), beginning (e.g. *komesa* ‘start’, *prepara* ‘prepare’, *kontinua* ‘continue’), and *la serve* ‘not worthy’.

(10.24) *Hau hakarak [(atu) bá Bidau]*.
 1S want IRR go Bidau.
 ‘I’d like to go to Bidau.’

(10.25) *Ami komesa [koalia kona-ba oinsá atu dezenvolve Timór]*.
 1PE start talk about how IRR develop Timor
 ‘We started talking about how to develop Timor.’

(10.26) *Sira promete [atu fó osan]*.
 3P promise IRR give money
 ‘They promised to give money.’

Although *komesa* ‘start’ takes a following complement (10.25), speakers disagree on whether the Portuguese loan *para* ‘stop’ can take a complement, while *hotu* ‘be finished, finish’ instead follows the verb stating what is finished (see 8.2.2.4). The verb *koko* ‘try’ is for some speakers restricted to post-verbal position, as it is in Tetun Terik (e.g. *hemu koko tua* ‘drink try wine’ = ‘have a taste of the wine’). In the church register, however, it takes following complements as would its equivalent in Portuguese (e.g. *koko atu haré nia* ‘try IRR see 3S’ = ‘try to see him’).

Verbs of ordering (e.g. *haruka* ‘order, send’ (10.27)) and of causing (e.g. *husik* ‘allow’, *ajuda* ‘help’ (10.28)) take an object noun phrase followed by a reduced complement, with the

object being interpreted as the semantic subject of the complement verb phrase. Irrealis *atu* is possible but uncommon in such complements. Verbs of causing also occur in this construction. The most general verb of causation, *halo* 'make, do', is discussed in 9.6.3, where this construction is compared with other constructions expressing causation.

(10.27) *Sira haruka hau [tama iha SMP].*
 3P order 1S enter LOC junior.high.school
 'They ordered/sent me to start junior high school.'

(10.28) *Ita ajuda nia [lao].*
 1PI help 3S walk
 'We help her (the toddler) walk.'

Reduced complements of many verbs can optionally be introduced by the purposive conjunction *para* 'so.that' (10.29) (see section 10.6.4), optionally in combination with irrealis *atu* (10.30). These include verbs of ordering and causing (e.g. *haruka* 'order', *obriga* 'compel', *ajuda* 'help'), and wanting (e.g. *hakarak* 'want'). Unlike *atu*, *para* is not used in clauses expressing non-desire; that is, it does not occur in complements of predicates such as *lakohi* 'not want, refuse', or *la hakarak* 'not want'. Although *para* is common in speaking, it is used by only some writers, and not at all in church literature.

(10.29) *Sira obriga nia [para hatudu dalan].*
 3P compel 3S so.that show way
 'They compelled him to show the way.'

(10.30) *Agora imi prepara [para atu bá Dili].*
 now 2P prepare so.that IRR go Dili
 'Now you (must) prepare to go to Dili.'

In informal speech it is possible (but not common) for a complement of *hatene* 'know' to precede the complement-taking verb. The complement must be short, generally consisting of a single verb, a verb plus adverb, or a transitive verb plus single-word object. This construction occurs primarily with negative clauses (10.31), but is also found for contrastive ones. It usually indicates that the subject referent is unable to do something (or at least to do something well) as a result of not having learned it. Note that 'complements' of *bele* (section 8.3.2) and objects of transitive verbs (section 7.1.4) can be fronted under similar conditions.

(10.31) *Sira [hakerek] la hatene.*
 3P write not know
 'They are illiterate.'

10.6 Adverbial clauses

10.6.1 Introduction

An adverbial clause is a clause which functions to modify another clause. Adverbial clauses encoding time, condition, and concession tend to precede the main clause, while those specifying reason or purpose tend to follow the main clause.

10.6.2 Temporal

Temporal subordinate clauses usually precede the main clause (10.32, 10.35), but can also follow it. The exception is clauses introduced by *tó* ‘until’, which typically follow the main clause (10.33). Most temporal subordinate clauses are introduced by a conjunction. For discussion of temporal clauses without a conjunction see section 10.3.

(10.32) *Molok atu hahú koalía, ita reza lai.*
 before IRR start speak IPI pray FIRST
 ‘Before (we) start speaking, we’ll pray first.’

(10.33) *Ami hela iha Ailoklaran nuné tó situa^sun normál.*
 IPE stay LOC Ailoklaran like.this until situation normal
 ‘We stayed in Ailoklaran like this until the situation was normal (again).’

Conjunctions meaning ‘before’, ‘during’, ‘after’, ‘since’ and ‘until’ are all also classed as prepositions, since they can also introduce noun phrases; these are discussed in 6.6.

Simultaneity is usually expressed by *bain-hira* ‘when, whenever’ (lit. ‘day-how.many’) or *kuandu* ‘when, whenever, if’ (from Portuguese *quando* ‘when’).⁵ These overlap significantly, and both can be used for past, present or general situations. Nevertheless, there are differences. The term *bain-hira* cannot readily be used if there is doubt as to whether the event described in the subordinate clause actually happened or will happen; it can thus be translated ‘when, whenever’, but not ‘if’. In contrast, *kuandu* has an element of conditionality as well as time. Although it can introduce a subordinate clause referring to a single event that definitely occurred (10.35a), it is more often used of future or generic situations, being translatable according to context as ‘when’, ‘whenever’ or ‘if’ (10.36). It thus overlaps with *se* ‘if’ (see section 10.6.5), although it tends not to be used for pure conditionals, with no interest in time (e.g. one says *se bele* ‘if possible’, not **kuandu bele*). The church register uses *wain-hira* (the Tetun Terik pronunciation of *bain-hira*), but not *kuandu*.

⁵ Note that *bain-hira* is also an interrogative adverb, used to question future time. In this capacity it contrasts with *hori-bain-hira* (lit. ‘at.(past.time)-day-how.many’), which is used to question past time (see section 7.6.1). Some people allow *hori-bain-hira* to be used as a conjunction meaning ‘when (in the past)’, but this is not usual usage. Unlike its Portuguese source, most speakers do not use *kuandu* as an interrogative adverb in Tetun.

- (10.34) *Bain-hira hau bá Filipina, hau iha problema liu-liu kona-ba Inglés.*
 when 1S go Philippines 1S have problem RDP-more about English

‘When I went to the Philippines, I had problems, especially with regard to English.’

- (10.35) a. *Kuandu nia mai, hau toba hela.*
 when 3S come 1S sleep CONT
- b. *Bain-hira nia mai, hau toba hela.*
 When 3S come 1S sleep CONT

‘When she came, I was sleeping.’

- (10.36) *Kuandu la rega, dút namlaik.*
 when not sprinkle grass Wilt
 ‘If/when grass isn’t watered, it wilts.’

The word *momentu* is primarily a noun meaning ‘moment, short period of time’, just as it is in Portuguese. In this capacity it is often used to specify a setting, either with or without a preceding preposition *iha* (e.g. *(iha) momentu né* ‘(LOC) moment this’ = ‘at this/that time’). The term has also been extended to function as a conjunction meaning ‘while, at the moment when’ (10.37), in which capacity it is mainly used by more educated speakers. The extension to conjunction can be interpreted as a reanalysis of the noun *momentu* when it is modified by a relative clause which has no relativiser.

- (10.37) *Momentu sira hán hela, iha feto ida mai hasoru Jesus.*
 moment/while 3P eat CONT EXIST woman one come meet Jesus
 ‘Just as they were eating, a woman came to meet Jesus’.

10.6.3 Reason

Reason clauses are, in speaking, usually introduced by *tanba* ‘because’ (from *tán-ba* ‘because-to/for’). In the vast majority of cases the reason follows the main clause (10.38). When it precedes the main clause, it is frequently reduced to an anaphoric *tanba né* ‘because.of this’ (10.39). Other conjunctions with the same meaning are *tán* (used along with *tanba* in the church register, but uncommon in other writing), *purké* (from Portuguese *porquê*, rare, used by some Portuguese-influenced speakers), and *basá* (from *ba-sá* ‘for-what’, mainly restricted to the church register).

(10.38) *Nia estuda mais la aprende buat ida tanba nia ulun tós.*
 3S study but not learn thing one because 3S head tough
 'He studies but doesn't learn anything because he is a slow learner.'

(10.39) *Tanba né mak ami tauk.*
 because this FOC 1PE afraid
 'It is because of this that we are afraid.'

Another reason conjunction is *komu* 'as, since, because' (from Portuguese *como* 'as, since'), which is primarily used by Portuguese-influenced speakers and in formal non-church situations. Clauses introduced by *komu* usually precede the main clause.

(10.40) ... *Komu ami tauk, ami bá iha Polda.*
 because 1PE afraid 1PE go LOC police.HQ
 '(People were shooting everywhere.) Since we were afraid, we went to the provincial police headquarters.'

10.6.4 Purpose

A purposive clause expresses a motivating event which is unrealised at the time of the main clause event. Tetun purposive clauses usually follow the main clause.

There are three means of introducing purposive clauses, namely *para*, *atu* and *hodi*. The combinations *para atu*, *para hodi* and (less commonly) *atu hodi* also occur.

The conjunction *para* (from Portuguese *para que* 'so that, in order that') introduces purposive clauses (10.41) or verb phrases, with no implication that the purpose is achieved. It also optionally introduces complements of a range of verbs, including verbs of ordering, causing and wanting (see section 10.5). Although *para* is common in speaking, it is used by only some writers, and not at all in church literature.

(10.41) *Milísia sira sai bá hamrík iha lurón né nakonu, para kareta la bele liu.*
 militia PL exit go stand LOC road this full so.that vehicle not can pass
 'The militia stood all over the road, so that the vehicles couldn't pass.'

The irrealis marker *atu* is another common means of introducing purposive verb phrases. In writing, it can also introduce purposive clauses, although this is far less common. As with *para*, the purpose need not be achieved. In addition, *atu* optionally introduces verb phrase complements of a wide range of verbs (section 10.5.3) whose complements denote as-yet-unrealised situations. It is also common in subordinate clauses introduced by the conjunctions *antes* and *molok*, both meaning 'before'. This use of *atu* in purposive and complement clauses occurs in all registers. For more details about *atu*, see 8.2.4.2.

(10.42) *Sira sai mai atu halo funu hasoru marinheiru sira né...*
 3P exit come IRR make war meet sailor PL this
 'They came out to fight against the sailors. (But they didn't fight, as it was a priest who was first to come out of the boat and face them.)'

The third means of introducing purposive verb phrases (or for some people, clauses), is with *hodi* (also a verb meaning ‘carry, bring’).⁶ In this case, there is a strong implication that the purpose specified by the clause will come about; *hodi* is thus often translatable by both ‘in order to’ and ‘and’. This construction is primarily used in written and formal situations.

- (10.43) *Nia halai hodi subar hosi milísia.*
 3S run PURPOSE hide from militia
 ‘He ran away to hide from the militia (and did in fact hide).’

10.6.5 Condition

Conditional clauses usually precede the main clause. They are (except in the church register) mostly introduced by the Portuguese loan *se* ‘if’ (10.44). About 15% of conditional clauses with *se* contain the adverb *karik* ‘perhaps’ as well (10.45). While *se* on its own can be used for conditions that are known to be either true or false, *karik* requires that the truth of the condition not be known.

- (10.44) *Se ita la fán sasán, osan la iha.*
 if 1PI not sell goods money not EXIST
 ‘If we don’t sell goods, (we) don’t have money.’

- (10.45) *Se lori karik ami bá iha Polres, ema bele estraga hotu ami.*
 if bring perhaps IPE go LOC police.station person can destroy all 1PE
 ‘If (they) would bring us to the district police station, people could hurt/rape all of us.’

An alternative is to simply juxtapose the initial conditional clause to the main clause, without a conditional conjunction (see examples 10.3 and 10.5 in section 10.3). The church register, which strongly avoids the Portuguese loan *se* ‘if’, consistently uses such juxtaposition, additionally marking the conditional clause by the adverb *karik* ‘perhaps’.

10.6.6 Concession

Concession (‘although, even if’) can be marked by the auxiliary *bele* ‘can’, in either its usual pre-predicate position (10.46) or (at least in informal speech) in the pre-subject position. In informal speech, clause-final *mós* (an adverb meaning ‘also’) in the initial clause marks the relation ‘even if, even though’ (10.47).

⁶ As a means of joining clauses, *hodi* is far more common in Tetun Terik than in Tetun Dili, and is used there for a wider range of semantic relationships. In the church register, too, it is used more widely, including as coordinator of verb phrases (e.g. *rona hodi la halo tuir* ‘hear and not do follow’ = ‘hear and not obey’), and as a common means of introducing the verb *dehan* ‘say’ before direct quotes (e.g. *Jesus hatán hodi dehan “...”* ‘Jesus answer and say’ = ‘Jesus replied, “...”’). For discussion of its grammaticisation from verb to verb phrase coordinator in Tetun Terik, see Van Klinken (2000).

(10.46) *Ita bele kiak, maibé lalika tristi.*
 2S.HON can poor but need.not sad
 ‘Even if you are poor (which you may or may not be), there is no need to be sad.’

(10.47) *Emá husu ajuda mós nia la hatene.*
 person ask help also 3S not know
 ‘Even if people ask (him) for help, he doesn’t pay heed.’

The conjunction *maski* ‘although, even though’ is primarily used in church and written registers (10.48); only a minority of people use it in speaking. The Indonesian loan conjunction *biar* ‘although’ is considered by some to be accepted in Tetun, while others consider it to be Indonesian only. The Portuguese loans *mezmu*, *mezmuké* and *embora*, all glossed ‘although’, are occasionally used by some strongly Portuguese-influenced speakers and writers, particularly in formal situations.

(10.48) *Ita tenki dezenvolve ona Tetun, maski la sai lian ofisiál.*
 1PI must develop ANT Tetun though not become language official
 ‘(We) must develop Tetun even if it doesn’t become the official language (and much more if it does become the official language).’

10.7 Other means of connecting clauses

There are a range of connectors which occur at the beginning of a clause or larger unit of speech, and connect it semantically to what has gone before. These differ from the conjunctions discussed in section 10.6, in that the clauses introduced by these connectors do not seem to be subordinate to the preceding clause, and in that the order of the two clauses is not interchangeable.

Contrast may be indicated by *maibé* ‘but’ (used in all registers) or *mais* ‘but’ (based on Portuguese *mas*, and usually used in speaking only, particularly by people who are more influenced by Portuguese).

(10.49) *Hau husu, maibé nia la fó.*
 1S ask but 3S not give
 ‘I asked (for money), but he didn’t give any.’

Sequence is frequently indicated by the Portuguese loan *depois* ‘afterwards, then, after that’. The church register instead uses *hafoin*. The expression *tuir mai* ‘and next’ (lit. ‘follow come’) introduces the next item in a series, such as the next speaker at a forum, or the next player in a game.

(10.50) *Hau hán meudia nian, depois deskansa.*
 1S eat midday POS then rest
 ‘I ate the midday (meal), then took a nap.’

Consequence can be introduced by *entaun* ‘then, so, as a result’ (from Portuguese *então*), *né duni* ‘therefore’ (lit. ‘this indeed’), or *nuné* ‘so, therefore’ (also ‘like this’).

Mak (‘FOCUS’; see 7.8) can follow a clause or a connector such as *depois* ‘afterwards, then, after that’ to indicate that only after the situation described in the first clause has come about, can/did that described in the second clause occur. The same meaning can be expressed by *foin* ‘very recently, only just, JUST’ in the second clause, with *foin* either preceding or following the subject. It is common for *mak* and *foin* to co-occur (10.51). The first clause is optionally introduced by a conjunction meaning ‘if’ or ‘when’.

- (10.51) *Ami hein tó tuku nén mak foin sai hosi Gleno, tún mai Dili.*
 IPE wait until o'clock six FOC JUST exit from Gleno descend come Dili.
 ‘We waited until six o'clock before departing from Gleno, and coming down to Dili.’

Tail–head linkage (repeating part of the end of one sentence as the beginning of the next) is a very common feature in Tetun Terik. In Dili it is used repeatedly by some speakers when telling stories; others do not use it, and some strongly disapprove of it. It is not used at all in writing. In the following example, the repeated words are underlined. Note that there is no special tense–aspect marking on the repetition (in contrast to Tetun Terik, which marks the repeated segment with various indicators of past time (van Klinken 1999a:305)).

- (10.52) ... *servisu mós para. Servisu para, ami ida-idak halai ba foho.*
 work also stop work stop IPE each run to mountain
 ‘(The civil war started,) and work stopped. Work stopped, and we each ran away to the mountains/country.

Halai ba foho, hau bá Suai.
 run to mountain IS go Suai
 When we ran away to the country, I went to Suai.’

10.8 Relative clauses

Relativisation on subjects (10.53), objects (10.54) and time phrases uses the ‘gap’ strategy, whereby the relativised constituent is simply omitted within the relative clause. Relativisation on location or other constituents is rare, and there is little agreement amongst speakers as to how this should be done. Apart from the compulsory omission of the relativised constituent, and an inability to topicalise constituents within the relative clause, there are no apparent restrictions on the internal structure of relative clauses. For instance, they allow the full range of tense–aspect options, arguments, complements, and peripheral constituents (10.55).

- (10.53) *oan nebé sei kík*
 child REL still small
 ‘child who is still small’ or: ‘children who are still small’

(10.54) *buat nebé mak nia hakarak*
 thing REL FOC 3S want
 ‘the thing(s) which she wants’

(10.55) *buat ida nebé hatudu mai ita katak...*
 thing one REL show come IPI that
 ‘something which shows us that (the majority of the respondents over the age of thirty-five support Portuguese as the official language)’

There is considerable variation amongst speakers and between registers as to the choice of relativiser, the extent to which the relativiser is felt to be omissible, and the commonness or acceptability of the determiner *né* ‘this’ following the relative clause.

The most common relativiser in speaking is *nebé* (also an interrogative ‘where, which’). Clauses introduced by *nebé* rarely have a following *né*, even though this is possible. The shorter *bé* is common in church and formal writing; when used in speaking it commonly pairs with *né* after the relative clause (10.56). Some speakers however do not accept *bé* as a relativiser at all. Portuguese *ke* is used as a relativiser by some people in speaking, but is seldom used in writing. It also occurs in sequence with *nebé*, apparently as a complex relativiser (10.57).

(10.56) *livru bé nia sosa né*
 book REL 3S buy this
 ‘the book which she bought’

(10.57) *buat nebé ke ita bandu*
 thing REL REL IPI forbid
 ‘the things which we forbid’

In about a quarter of relative clauses introduced by *nebé*, the relativiser is followed by *mak*, which is in other contexts a focus marker; however, *mak* in this context presents the relativised constituent as known information (10.54). This focus marker is also found in relative clauses with no relativiser (10.59), but appears not to follow the relativisers *bé* or *ke*. For discussion of *mak*, and further examples, see section 7.8. The function of *mak* in relative clauses requires further research. It is also still unclear to what extent this function is related to the general Tetun Dili use of *mak* as a focus marker, and to what extent it is a relic of *mak*’s function as a relativiser in Tetun Terik.

In Tetun Dili, *mak* appears to still be used as a relativiser only in fixed formulae, particularly in *mak ami respeita* ‘REL 1PE respect’ = ‘whom we respect’. This expression commonly post-modifies names or titles of people in situations in which they are addressed very formally, including the formal listing of guests of honour found at the beginning of speeches, and some written correspondence.

Some Tetun Dili speakers feel that a relativiser is required in most contexts, while others consider it to be readily omissible. All, however, prefer to omit a relativiser when the noun phrase is the argument of a preceding *iha* ‘EXIST, there is’ (10.58, 10.59).

(10.58) *Iha ema balu temi ita.*

EXIST person some mention 2S.HON

‘There are people mentioning you.’ (This is what people say when one sneezes.)

(10.59) *Iha kolega barak mak dehan hanesan né.*

EXIST colleague many FOC say like this

‘There are many workmates who say this.’

Relativisers are also readily omissible when the modifier is a ‘body-good’ expression (e.g. *ema ain át* ‘person leg bad’ = ‘crippled person’) or introduced by *naran* ‘name’ (e.g. *ema ida naran Carlos* ‘person one name Carlos’ = ‘a person named Carlos’). See sections 7.2 and 7.3 for discussion of these constructions. Adjective phrases following *ida* ‘one’ (which can by virtue of their position and semantics be analysed as relative clauses) similarly lack a relativiser (see section 5.3.4).

Appendix: Texts

In all the texts presented here, editing has been restricted to removing the minimal number of hesitations and false starts.

Text 1: History behind setting up an NGO

A man in his mid-twenties explains to one of the authors the history that led to him and some friends establishing an NGO (non-government organisation) early in 2000. The aim of the NGO was to improve the economic conditions of poor people in East Timor. The speaker did all his schooling in Dili, and completed a university degree in Java, Indonesia. Although he does not speak Portuguese, he consistently uses Portuguese loans even where some other young people would borrow from Indonesian.

- (T1.1) *Tinan tolu, tinan hát liu bá, ami iha Indonesia,*
year three year four pass go IPE LOC Indonesia
Three or four years ago, (when) we were in Indonesia (attending university),
- (T1.2) *ami komesa koalia kona-ba oinsá atu dezenvolve Timór.*
IPE start speak about how IRR develop Timor
we started talking about how to develop Timor.
- (T1.3) *Iha momentu nebá, ami fiar katak loron ruma Timór sei ukun-an.*
LOC moment there IPE believe that day some Timor FUT rule-self
At that time, we believed that some day Timor would be independent.
- (T1.4) *I se Timór ukun-an, ita tenki dezenvolve ona Timór.*
and if Timor rule-self IPI must develop ANT Timor
And if Timor were to be independent, we would already have to start developing Timor.
- (T1.5) *Para Timór nia dezvoltimentu bele diak ba oin,*
so.that Timor POS development can good to face
In order that the development of Timor would progress well,

- (T1.6) *ita tenki hanoin ona atu dezenvolve hosi ema kík¹ sira:*
 IPI must think ANT IRR develop from person small PL
 we should already be considering developing starting from the little people:
- (T1.7) *ema sira nebé forsa la iha iha ekonomia,*
 person PL REL strength not have LOC economy
 the people who don't have power in the economy,
- (T1.8) *ema sira nebé la iha matenek atu muda sira nia ekonomia;*
 person PL REL not have cleverness IRR move 3P POS economy
 the people who don't have the intelligence to change their economic status.
- (T1.9) *Ema sira né tenki ajuda²*
 person PL this must help
 These people need to be helped
- (T1.10) *para sira bele ha-sae sira nia ekonomia.*
 so.that 3P can cause-ascend 3P POS economy
 so that they can raise their economic status.
- (T1.11) *Entaun, ami komesa hanoin ida né ona.*
 so IPE start think one this ANT
 So, we started to think about this.
- (T1.12) *I, momentu nebé ami seidauk hatene,*
 and moment there IPE not.yet know
 And, at that time, we didn't yet know
- (T1.13) *oinsá mak atu bele ajuda ema kík sira né.*
 how FOC IRR can help person small PL this
 how it was that we could help these little people.
- (T1.14) *Hau eskola hotu tiha, mais, antes hau eskola hotu,*
 1S school finish PRF but before 1S school finish
 I have finished my schooling (at university), but, before I finished it,
- (T1.15) *dala ruma tinan-tinan ou fulan nén-nén, hau sempre mai iha Timór,*
 occasion some RDP-year or month RDP-six 1S always come LOC Timor
 on occasion, every year or every six months, I always came to Timor,
- (T1.16) *atu haré tuir kona-ba kondisaun Timór halo nusá,*
 IRR see follow about condition Timor do how
 to check out how the situation in Timor was,

¹ In Tetun, the *ema kík* 'little people' stand in contrast to the *ema bót* (or *em-bót*) 'person big' = 'VIPs, leaders, important people'.

² Object-verb clauses such as this sometimes function much like passives (sections 7.1.1 and 7.1.3).

- (T1.17) *saida mak ami bele halo, para atu ajuda ema kík sira.*
 what FOC 1PE can do so.that IRR help person small PL
 (and) what we could do in order to help the little people.
- (T1.18) *I hau eskola hotu tiha, hau mai iha né,*
 and 1S school finish PRF 1S come LOC this
 And (when) I finished my schooling, I came here,
- (T1.19) *hau komesa koalia ho kolega sira:*
 1S start speak with friend PL
 (and) I started talking with friends:
- (T1.20) *balu iha universidadi, balu servisu iha fatin seluk,*
 some LOC university some work LOC place other
 some at university, some working in other places,
- (T1.21) *i komesa hanoin kona-ba ida né.*
 and start think about one this
 and (we) started to think about this.
- (T1.22) *I, momentu neba ami halo ona konseitu ida,*
 and moment there 1PE make ANT draft.paper one
 And, at that time we made a draft plan,
- (T1.23) *para atu formula ami nia ideia né.*
 so.that IRR formulate 1PE POS idea this
 in order to formulate our ideas.

Text 2: Why people create disturbances

A woman in her forties explains to one of the authors why people would accept money to cause a serious disturbance at a well-attended soccer match in Dili. The speaker did her schooling during the Portuguese era. She subsequently became a school teacher, initially teaching in Portuguese and then, after the Indonesian takeover, in Indonesian.

In the portion of the text presented below, 25% of the word tokens are Portuguese loans; if one counts only the unique words (e.g. counting *ema* only once), the proportion of Portuguese loans rises to 36%. Such a high proportion of loans is not unusual for Portuguese-educated people when they are speaking of complex matters.³

- (T2.1) *Tanba sa? Sira ema át – mau fundu.*
 because what 3P person bad wicked core
 Why? They are evil—wicked to the core.

³ Note that the loans cannot be attributed to the fact that the addressee was a foreigner, as the author in question can speak Tetun and Indonesian, but not Portuguese. In such cases it is not unusual for speakers to increase the percentage of Indonesian loans to facilitate communication.

- (T2.2) *Sira nia laran át.*
3P POS inside bad
They are wicked.
- (T2.3) *Sira la gosta haré ema diak, ema kontenti, ema pás.*
3P not like see person good person happy person peace
They don't like seeing good people, contented people, people at peace.
- (T2.4) *Sira la gosta, tanba sira né hanesan⁴ halo buat át barak ona,*
3P not like because 3P this like do thing bad many ANT
They don't like (it) because they have done many bad things,
- (T2.5) *ke sira iha tendénsia ba buat át deit.*
REL 3P have tendency to thing bad just
(such that) they have a tendency only toward bad things.
- (T2.6) *Né duni sira nia konsiénsia né, sira nia laran;...⁵*
therefore 3P POS conscience this 3P POS inside
So their conscience, their heart,...
- (T2.7) *Né la iha pás, la iha trankuúlu⁶.*
this not have peace not have tranquillity
It doesn't have peace, doesn't have tranquillity.
- (T2.8) *Né duni, sira komu halo ona maldadi barak,*
therefore 3P because do ANT cruel.thing many
So, because they have done many cruel things,
- (T2.9) *sira iha tendénsia ba buat át deit.*
3P have tendency to thing bad just
they have a tendency only to evil.
- (T2.10) *Sira kuandu haré ema kontenti, ema barak saida⁷,*
3P when see person happy person many whatever
They, when they see people happy, or many people whatever,

⁴ *Hanesan* is frequently used in colloquial speech to introduce an example, or to prevaricate; much as its gloss 'like' is used in colloquial English.

⁵ At this point, having already stated the concept of 'conscience' in Portuguese and paraphrased it in Tetun, the speaker further reiterated it in Indonesian and again in Portuguese (*hati nurani ka, konsiénsia*), presumably on the assumption that the addressee would not understand *konsiénsia* 'conscience'. Such rephrasing in Indonesian is at times required when Portuguese-educated people communicate with those educated in Indonesian.

⁶ This adjectival loan word is used as a noun, in the same way as native Tetun adjectives can be nominalised by zero derivation (section 4.2.3).

⁷ *Saida* 'what, whatever', can in colloquial speech be used like this, as a predicate or noun phrase modifier meaning 'do/be whatever'.

- (T2.11) *mai*⁸ *kedan buat át iha sira nia ulun dehan*, “*tenki soe granada*,
 come IMMED thing bad LOC 3P POS head say must throw grenade
 there immediately comes something evil into their heads saying, “(you) must throw a
 grenade,
- (T2.12) *tenki halo át, tenki baku, tenki soku malu, tenki oho malu*”,
 must do bad must beat must punch RECI must kill RECI
 must do evil, must beat (one another), must punch one another, must kill one
 another”,
- (T2.13) *para hodi halo dezorden*.
 so.that so.that make disorder
 in order to create disorder.
- (T2.14) *Sira gosta haré dezorden*.
 3P like see disorder
 They love seeing disorder.

Text 3: Our escape from the militia

A woman in her forties tells one of the authors about her escape from a militia group. She was working for UNAMET, the United Nations body which oversaw the popular consultation in August 1999, at which the East Timorese voted overwhelmingly for independence. The events occurred after the close of voting. The staff had already been held up for many hours, being prevented from leaving the town to go to Dili.

- (T3.1) ..., *tó fali tréz oras*,
 until again three o'clock
 (We were inside the building), until three o'clock,
- (T3.2) *mak malae*⁹ *UNAMET sira haruka ami mai fali sae kareta*;
 FOC foreigner UNAMET PL order IPE come again ascend vehicle
 then the UNAMET foreigners told us to again come and get into the vehicles.
- (T3.3) *Maibé, ami la sae ona kareta míkrolet*;
 but IPE not ascend ANT vehicle mini-bus
 But, we didn't get into the mini-bus (as we had before, when the militia wouldn't let
 us leave);

⁸ This is an unusual instance of *mai* ‘come’ preceding its subject. For details of verb–subject clauses, see section 7.1.6.

⁹ The term *malae* ‘foreigner’ is a polite means of referring to foreigners.

- (T3.4) *ami sae hotu UNAMET ninia¹⁰ kareta.*
IPE ascend all UNAMET 3S.POS vehicle
we all got into UNAMET's cars.
- (T3.5) *Ami sae hotu UNAMET nia kareta,¹¹ ami sae komesa tuku tolu,*
IPE ascend all UNAMET POS vehicle IPE ascend start o'clock three
We all got into UNAMET's vehicles, we got in at 3 o'clock,
- (T3.6) *ami túr hela iha kareta laran nuné nuné nuné,¹²*
IPE sit CONT LOC vehicle inside like.this like.this like.this
we were sitting in the vehicle like this (on and on),
- (T3.7) *ema la fó ami sai.*
person not give IPE exit
people wouldn't let us leave.
- (T3.8) *Ami hein nuné tó tuku nén, mak ami foin sai hosi N—.*
IPE wait like.this until o'clock six FOC IPE JUST exit from (town)
We waited like this until six o'clock, (and) only then left N— (the town),
- (T3.9) *tún mai¹³ Dili.*
descend come Dili
(and) came down to Dili.
- (T3.10) *Mai iha dalan, malae hateten ba xofér sira nebé mak lori kareta,*
come LOC path foreigner tell to driver PL REL FOC drive vehicle
Coming along the road, the foreigner(s) said to the drivers who drove the cars,
- (T3.11) *dehan katak "Iha dalan, la bele para.*
say that LOC path not can stop
"On the way, don't stop.
- (T3.12) *Bele halo nusá mós, la bele para.*
can do how also not can stop
No matter what, don't stop.
- (T3.13) *Se para, entaun sira bele tiru. Tanba sá?*
if stop so 3P can shoot because what
If (we) stop, then they can shoot. Why?

¹⁰ The full possessive marker *ninia* is in the next clause replaced by the more common variant *nia*.

¹¹ This is an example of tail-head linking, where part of the end of one sentence is repeated at the beginning of the next sentence (see section 10.7). It is ubiquitous in the story-telling of some people, but strongly avoided in writing, and some speakers claim to not even recognise the phenomenon.

¹² Repetition like this is used by some good story-tellers, but is strictly avoided in writing.

¹³ It is customary to specify motion along two axes; whether it is upwards or downwards (even if the altitude variation is very slight), and whether it is towards or away from the speaker.

- (T3.14) *Kareta UNAMET né né,¹⁴ ita lori (be) estáf¹⁵ lokál mak barak.*
 vehicle UNAMET this this IPI carry HES staff local FOC many
 (In) these UNAMET cars, we are carrying mostly local staff.” (Note that local staff
 were the primary target of militia attack.)
- (T3.15) *Né duni, mai, xofér mós halo tuir, xofér la para.*
 therefore come driver also do follow driver not stop
 So, (as we) came, the drivers acted accordingly, and didn’t stop.
- (T3.16) *Ami mai tó iha P—*
 IPE come arrive LOC (place)
 We came to P—
- (T3.17) *(kruzamentu atu bá A—, bá B—, mai nemai);*
 intersection IRR go (town) go (town), come here
 (the intersection to go to A—, to B—, and to here (i.e. Dili));
- (T3.18) *ami tó iha né, ema tau ai mesa bo-bót bo-bót tiha deit,*
 IPE arrive LOC this person put wood solely RDP-big RDP-big PRF just
 we arrived there, (and) people had put out big logs,
- (T3.19) *satan netik tiha lurón klaran né, para kareta la bele halai lai-lais;*
 block hinder PRF road middle this so.that vehicle not can run RDP-fast
 blocking off the middle of the road, so that cars couldn’t drive fast;
- (T3.20) *kareta tenki lao neineik para sira bele halo buat ruma ba kareta.*
 vehicle must walk slowly so.that 3P can do thing some to vehicle
 the cars had to travel slowly so that they (the people who put up the barricades) could
 do something or other to the cars.
- (T3.21) *Maibé xofér sira mós barani tebe-tebes.*
 but driver PL also brave RDP-true
 But the drivers were truly brave.
- (T3.22) *I sira sés kareta lai-lais; nuné kareta halai nafatin.*
 and 3P move.aside vehicle RDP-quick like.this vehicle run continue
 And they quickly turned the cars aside; in this way the cars kept driving.

¹⁴ The first *né* is anaphoric, while the second marks the noun phrase as a topic (see section 5.3.2).

¹⁵ This appears to be an attempt to make the Indonesian term *staf* ‘staff’ sound like Portuguese. Since the events recounted took place during Indonesian rule, the usual means of referring to local staff employed by the United Nations was the Indonesian phrase *staf lokal* (with stress on the /o/ instead of the /a/ as in Portuguese).

Text 4: A press report

The following written text from the newspaper *A Voz de/Suara Timor Lorosae* (July 4, 2001) illustrates the rapidly developing register used in the press. It is presented in its original spelling, with our spelling underneath where different.

In the portion of the text presented below, 50% of the word tokens are Portuguese loans. Most but not all of the grammatical terms are native Tetun words. However of the nouns, verbs, adjectives and adverbs, only two (i.e. 10%) are not borrowed, namely *loron* 'day' and *hahú* 'start'. The latter is a Tetun Terik verb which is commonly used in the church register; in everyday speech, its Portuguese synonym *komesa* is far more common.

- (T4.1) *Porta-voz UNTAET nian informa liu tan katak*
porta-vós UNTAET POS inform pass as.well that
 spokesperson UNTAET POS inform pass as.well that
 The UNTAET (United Nations Transitional Authority in East Timor) spokesperson further informed that
- (T4.2) *"iha 15 de Julho loron hahu campanha eleitoral nian,*
Julhu hahú kampanha eleitoral
 LOC 15 of July day start campaign electoral POS
 "On 15th July, the day on which the election campaign begins,
- (T4.3) *membro gabinete sira nebe activo politicamente*
membru nebé aktivu politikamenti
 member cabinet PL REL active politically
 those cabinet members who are active politically
- (T4.4) *sei atu suspende sira nia fungsaun governamental [sic]*
funsaun governamental
 FUT IRR suspend 3P POS function governmental
 would have to suspend their government functions,
- (T4.5) *i administrador transitorio Sergio Vieira de Mello maka*
administradór transitóriu
 and administrator transitional Sergio Vieira de Mello FOC
 and the transitional administrator, Sergio Vieira de Mello,
- (T4.6) *sei dirigi elemento-s gabinete nian*
diriji elementu-s
 FUTdirect element-PL cabinet POS
 would henceforth be in charge of those cabinet ministers
- (T4.7) *nebe laos husi membro partido nian ida".*
nebé laós hosi membru partidu
 REL indeed.not from member party POS one
 who are not in fact members of a party.

Text 5: A gospel portion

The following portion from the published translation of Luke 9:1–3 illustrates the register used in the Catholic Church. It is presented in the original spelling, which differs from ours mainly in its representation of long vowels, and in maintaining final ‘e’ and ‘o’ (rather than ‘i’ and ‘u’) on Portuguese loans. Our spelling follows where there are differences.

The church register tends to avoid Portuguese loans where possible. This is amply illustrated by the fact that this text has only two loans, *reinu* ‘kingdom’ and *diabu* ‘devil, evil spirit’. The press passage above, which has a similar overall word count, has ten times as many loans. Tetun Terik influence on the church register is reflected in such things as the use of *no* and *hodi* as coordinators (instead of *ho* and *i*), and the use of *kbít* ‘power’ (instead of Portuguese loans such as *forsa* or *podér*).

- (T5.1) *Jesus haruka sira nain sanulu resin rua,*
 Jesus send PL CLS:human ten extra two
 Jesus sent the twelve of them
- (T5.2) *hodi fo kbiit ba sira atu duni sai diabo sira*
fó kbít diabu
 and give power to 3P IRR chase exit devil PL
 and gave them power to cast out the evil spirits
- (T5.3) *no ha-lakon moras sira.*
 and cause-disappear sickness PL
 and heal the diseases.
- (T5.4) *Nia haruka sira atu haklaken Maromak nia reino,¹⁶*
reinu
 3S send/order 3P IRR preach God POS kingdom
 He sent/ordered them to preach God’s kingdom
- (T5.5) *No mos atu fo isin diak ba ema moras sira.*
mós fó
 and also IRR give body good to person sick PL
 and also to give health to the sick people.
- (T5.6) *Nia dehan ba sira: "Keta lori buat ida iha dalan..."*
 3S say to 3P PROH carry thing one LOC road
 He said to them, “Don’t carry anything on the way...”

¹⁶ In the church register, *reinu* means ‘kingdom’, as per its Portuguese source *reino*. In everyday usage, it refers instead to a ruler’s subjects, that is, to the citizens of the kingdom.

References

- Bowden, John, 1992, *Behind the preposition: grammaticalisation of locatives in Oceanic languages*. Canberra: Pacific Linguistics, B-107.
- Bybee, Joan L., Revere Perkins and William Pagliuca, 1994, *The evolution of grammar: tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.
- Chafe, Wallace L., 1976, Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In Charles N. Li, ed. *Subject and topic*, 25–55. New York: Academic Press.
- Clark, Eve V., 1978, Locationals: existential, locative, and possessive constructions. In Joseph H. Greenberg, ed. *Universals of human language, volume 4: Syntax*, 85–126. Stanford: Stanford University Press.
- Costa, Luís, 2000, *Dicionário de Tétum-Português*. Lisbon: Edições Colibri.
- Crowley, Terry, 1987, Serial verbs in Paamese. *Studies in Language* 11:35–84.
- Dixon, R.M.W., 1991, *A new approach to English grammar, on semantic principles*. Oxford: Clarendon Press.
- Durie, Mark, 1997, Grammatical structures in verb serialization. In Alex Alsina, Joan Bresnan and Peter Sells, eds *Complex Predicates*, 289–354. Stanford: CSLI Publications.
- Early, Robert, 1993, Nuclear layer serialization in Lewo. *Oceanic Linguistics* 32:65–93.
- Eccles, Lance, 1998, Perfective and inchoative verbal markers in Tetum. *Studies in Languages and Cultures of East Timor* 1:39–51.
- 1999, The verb markers of Tetum-Praça and their grammaticalization. *Studies in Languages and Cultures of East Timor* 2:22–37.
- Foley, William A. and Mike Olson, 1985, Clausehood and verb serialization. In Johanna Nichols and Anthony C. Woodbury, eds *Grammar inside and outside the clause. Some approaches to theory from the field*, 17–60. Cambridge: Cambridge University Press.
- Foley, William A. and Robert D. Van Valin, Jr, 1984, *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Fox, James J., 1997, The historical position of Tetun among the languages of the Timor area. *Tetun language conference*, Darwin.
- Givón, Talmy, 1978, Definiteness and referentiality. In Joseph Greenberg, ed. *Universals of human language, volume 4: Syntax*, 291–330. Stanford: Stanford University Press.

- 1980, The binding hierarchy and the typology of complements. *Studies in Language* 4:333–377.
- Grimes, Barbara F., ed. 2000, *Ethnologue: languages of the world* (14th edition). Dallas, Texas: Summer Institute of Linguistics.
- Grimes, Charles E., 1991, The Buru language of eastern Indonesia. PhD, Department of Linguistics, Research School of Pacific Studies, The Australian National University.
- Grimes, Charles E., Tom Therik, Barbara Dix Grimes and Max Jacob, 1997, *A guide to the people and languages of Nusa Tenggara*. Kupang: Artha Wacana Press.
- Hajek, John, 2000, Language planning and the sociolinguistic environment in East Timor: Colonial practice and changing language ecologies. *Current Issues in Language Planning* 1:400-414.
- in press, Language maintenance and survival in East Timor: All change now? Winners and losers. In D. Bradley and M. Bradley, eds *Language endangerment and language maintenance: an active approach*. London: Curzon Press.
- Hajek, John and Alexandre Vital Tilman, 2001, *East Timor Phrasebook*. Melbourne: Lonely Planet Publications.
- Hawkins, John A., 1978, *Definiteness and indefiniteness: a study in reference and grammaticality prediction*. London: Croom.
- Hengeveld, Kees, 1992, *Non-verbal predication: theory, typology, diachrony*. Berlin: Mouton de Gruyter.
- Himmelman, Nikolaus P., 1996, Demonstratives in narrative discourse: a taxonomy of universal uses. In Barbara Fox, ed. *Studies in anaphora*, 203-252. Amsterdam: Benjamins.
- Hull, Geoffrey, 1994, A national language for East Timor. In István Fodor and Claude Hagège, eds *Language reform: History and future. La Réforme des Langues: Histoire et Avenir. Sprachreform: Geschichte und Zukunft*, volume 6, 347-366. Hamburg: Helmut Buske Verlag.
- 1996a, A language policy for East Timor: background and principles. *It's time to lead the way: writings from a conference on East Timor and its people*, 38–59. Melbourne: East Timor Relief Association.
- 1996b, *Mai kolia Tetun: a beginner's course in Tetum-Praça: the lingua franca of East Timor* (second edition). North Sydney: Australian Catholic Relief and the Australian Catholic Social Justice Council.
- 1998, The languages of Timor 1772–1997: a literature review. *Studies in languages and cultures of East Timor* 1:1–38.
- 1999, *Standard Tetum-English dictionary*. Sydney: Allen & Unwin in association with the University of Western Sydney Macarthur.
- Hull, Geoffrey and Lance Eccles, 2001. *Tetum reference grammar*. Winston Hills: Sebastião Aparício da Silva Project in conjunction with Instituto Nacional de Linguística, Universidade Nacional de Timor Lorasa'e.

- Martins, Fransisco Miguel, 1998, *Voice in Tetum-Dili language*. MA, Gadjah Mada University.
- Morris, Cliff, 1984a, *Ai knananuk ho ai knanoik nousi Rai Timur: rai nousi lafaek dukur. Verse and legends from Timor the land of the sleeping crocodile, book 1*. Frankston, Victoria: H.C. Morris.
- 1984b, *Tetun - English dictionary*. Canberra: Pacific Linguistics, C-83.
- Mühlhäusler, Peter, 1986, *Pidgin and creole linguistics*. Oxford: Basil Blackwell.
- Noonan, Michael, 1985, Complementation. In Timothy Shopen, ed. *Language typology and syntactic description, volume II: Complex constructions*, 42–140. Cambridge: Cambridge University Press.
- Prince, Ellen F., 1981, Toward a taxonomy of given-new information. In Peter Cole, ed. *Radical Pragmatics*, 223–255. New York: Academic Press.
- Schachter, Paul, 1974, A non-transformational account of serial verbs. *Studies in African Linguistics Supplement* 5:253–270.
- Sneddon, James Neil, 1996, *Indonesian reference grammar*. St Leonards: Allen & Unwin.
- Thomason, Sarah G., 1997, A typology of contact languages. In Arthur K. Spears and Donald Winford, eds *The structure and status of pidgins and creoles*, 71–88. Amsterdam/Philadelphia: John Benjamins.
- Thomaz, Luis Filipe F.R., 1981, The formation of Tetun-Praça, vehicular language of East Timor. In Nigel Phillips and Anwar Khaidir, eds *Papers on Indonesian languages and literatures*, 54–83. Paris: Cahier d'Archipel 13.
- van Engelenhoven, Aone and Catharina Williams-van Klinken, in preparation, Sketches from the Timor region: Tetun and Leti. In Sander Adelaar and Nikolaus Himmelmann, eds *The Austronesian languages of Asia and Madagascar*. Richmond: Curzon Press.
- van Klinken, Catharina, 1999a, *A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor*. Canberra: Pacific Linguistics, C-155.
- 1999b, Writing Tetum-Prasa: a communication challenge. *Studies in Languages and Cultures of East Timor* 2:14–21.
- 2000, From verb to coordinator in Tetun. *Oceanic Linguistics* 39:350–363.
- Van Valin, Robert D., Jr, 1993, A synopsis of role and reference grammar. In Robert D. Van Valin, Jr, ed. *Advances in role and reference grammar*, 1–164. Amsterdam: John Benjamins.
- Williams-van Klinken, Catharina, in preparation, High registers of Tetun Dili: Portuguese press and purist priests. *Proceedings of the 2001 Conference of the Australian Linguistics Society*.

Index

- accompaniment, 45
addressee, 43
 questioned, 65
adjective, 14–15. *See* body-good
 expression, quantifying adjective
 as ‘possessor’, 35
 as root for derivation, 19–21, 22, 97–98
 derived, 20, 24
 in reciprocal expression, 62
adjective phrase, 36–37, 117
 function, 25, 30, 57
adverb, 15, 73, 91–92. *See* interrogative
 adverb
 derived, 21, 22, 24
adverbial clause, 109–14
affix. *See* prefix, suffix
alienability, 34
alternative interrogative, 68
án, -an ‘REFLEXIVE’, 63–64
anaphora, 28, 53
anterior aspect, 75–76
antes ‘before’, 46
aspect. *See* tense–aspect
associative relation, 35
até ‘until’, 48
atu ‘IRREALIS’, 83–84, 109, 112
atu-mate ‘very’, 90
auxiliary, 15, 68, 81, 84–87
bá
 ‘go’, 43, 94
 command, 69
 deictic, 43, 94
ba ‘to, for’, 44
bain-hira
 ‘when, whenever’, 110–11
 ‘when?’, 66
balu ‘some’, 31
basá ‘because’, 111
bé relative clause marker, 116
bele ‘can’, 84–86
 in concessive clause, 113
beneficiary, 43
 questioned, 65
besik ‘close to, near’, 48
body-good expression, 57–60, 117
calque. *See* Portuguese calque
causative, 97–100
 periphrastic, 98–99
 prefix *ha-*, 97–98
 serial verb, 98–99
church register, 4, 126
 differences, 32, 33, 38, 42, 46, 47, 49,
 58, 65, 69, 82, 86, 104, 105, 106,
 108, 109, 110, 111, 112, 113, 114,
 116
classifier. *See* numeral classifier
clause, 51–61
 adverbial, 109–14
 complement, 106–9
 verb-initial, 55–57
cleft construction, 69
clitic, 36, 63
command, 69
comparative, 90
complement clause, 40, 83, 106–9
 fronted, 85, 109
complementiser, 106–7
completive aspect, 77
compound, 21, 22–24, 99–100
 phonological reduction, 12
concessive clause, 113–14
conditional clause, 103, 113
conjunction, 16
 coordinating, 103–5

- subordinating, 45–48
- connective adverb, 114–15
- consequence ('so'), 115
- consonant, 8–11
 - cluster, 13
- continuous aspect, 79–80
- contrast ('but'), 114
- coordination, 103–5
- copula, 57
- core-layer serial verb, 93
- daudauk* 'PROGRESSIVE', 80–81
- de* 'of', 50
- definiteness, 27–29, 30, 53, 55
- deictic particle, 43, 94
- deixis. *See* deictic particle, demonstrative,
 - personal pronoun
- demais* 'too, excessive', 90
- demonstrative, 27–29
- deontic modal auxiliary, 84–87
- depois*
 - 'after', 46
 - 'afterwards, then', 114
- determiner, 15
- detransitivising prefix, 19–20
- dezde* 'since (time)', 47
- direction verb, 94
- disjunction, 105
- ditransitive verb, 51
- dór* 'AGENT', 20
- duranti* 'for, during', 46
- ellipsis of subject and object, 53
- embora* 'although', 114
- entaun* 'so, as a result', 115
- entre* 'between', 49
- equative clause, 60
- exclamation, 16
- existential clause, 55–56, 116
- floating quantifier, 15, 31
- fó* 'give', 51
 - in serial verb construction, 99–100
- focus, 69–71
- foin* 'JUST, very recently', 77–78, 115
- future, 82–83, 85
- gender, 20, 23, 26, 37
 - of *obrigadu/a*, 16
- generic, 27
- genitive marking, 36
- goal (location), 43, 45
- grammaticalisation, 49, 77, 95–96
- greeting, 16
- ha-* verb-deriving prefix, 97–98
- hafoin* 'afterwards, then', 114
- hahú* 'start, from', 47
- halo* 'do, make', 18, 98–99
- halo nusá* 'how?', 66
- halo oinsá* 'how?', 66
- hanesan* 'like', 49
- hela* 'CONTINUOUS', 79–80
- hesitation marker, 16
- hira* 'how many?, several', 32, 65
- hirak* 'certain', 32
- ho*
 - 'and', 103–5
 - 'with', 45
- hodi*
 - 'bring', 95–96
 - 'so that, and', 113
- hori* 'at (past time)', 47
- hori-bain-hira* 'when (past)?', 47, 66
- hosi* 'from', 44–45
- hotu*
 - 'all, also', 31
 - 'be finished', 77
- hotu-hotu* 'all', 31
- i* 'and', 103–5
- ida*
 - 'one, a', 30
 - with negator, 89
- ida-idak* 'each', 32
- iha*
 - 'EXIST, have', 55–56
 - 'LOC, in, on, at', 41–43
- imperative, 69
- inalienability, 34
- incorporation, 55. *See* complement clause
 - fronted
- indefiniteness, 30, 32, 89
- Indonesian. *See* Malay
- information interrogative, 64–67
- instrument, 45, 95–96
- instrument serial verb, 95–96
- intensifier, 89–91
- interjection, 16
- interrogative, 64–68

- interrogative adverb, 66
interrogative pronoun, 29, 64–65
irrealis, 83–84
ka ‘or’, 105
karik ‘perhaps’, 113
katak ‘that’ (complementiser), 106–7
ke relative clause marker, 116
kedas ‘IMMEDIATE’, 79
keta ‘PROHIBITIVE’, 86
kin relation, 34, 36
komesa ‘start, from’, 47
komu ‘since, because’, 112
kona-ba ‘about (topic)’, 49
konformi ‘according to’, 50
kuandu ‘when, if’, 110–11
la ‘not’, 87
ladún ‘not very’, 88
lae ‘no’, 88
la-halimar ‘very’, 90
lai ‘FIRST’, 78
lalikan ‘need not’, 86
laós ‘not’, 87–88
left-dislocation, 102
liu
 ‘more, most’, 90–91
 ‘pass, after’, 47
loan. *See* Portuguese loan, Malay loan
location. *See* goal, source
 peripheral, 72
location noun, 34, 41–43
lori ‘bring’, 95–96
lós ‘very’, 90
mai ‘come, to, for’, deictic, 43–44, 94
maibé ‘but’, 114
mais ‘but’, 114
mak
 ‘FOCUS’, 60, 65–66, 67, 69–71, 78, 90,
 115, 116
 relative clause marker, 116
Malay/Indonesian, 1, 2, 3, 37–39
 influence on phonology, 8, 10, 11, 13
 loan, 11, 41, 114
malu ‘RECIPROCAL’, 61–63
manner, 45, 91–92
maski ‘although’, 114
mezmu ‘although’, 114
mezmuké ‘although’, 114
molok ‘before’, 46
momentu ‘moment, when’, 111
morphology, 16–24, 97–98
mós ‘also’, 105
 in concessive clause, 113
motion–action serial verb, 94
motion–direction serial verb, 94–95
-n genitive, 36
nain
 ‘agent’, 21
 classifier for humans, 39
nak- detransitivising prefix, 19–20
nam- detransitivising prefix, 19–20
naran ‘name’, 34, 61
né ‘this’, 27–29
né duni ‘therefore’, 115
nebá ‘there, that’, 27–28
nebé
 ‘where, which?’, 65
 relative clause marker, 116
negation, 59, 87–89
no ‘and’, 104
nominalisation, 19–21, 34
noun, 14. *See* location noun
 as root for derivation, 18–19, 21
 complement of, 40
 derived, 19–21, 23–24
noun phrase, 25–26
 coordination, 104
 disjunction, 105
 function, 36, 60, 71
 without determiner, 27
nuclear-layer serial verb, 93
nudar ‘as, in the capacity of’, 49
number. *See* plural marking
 agreement, 37
numeral, 15, 30, 37–39
 compound *ida-rua*, 32
 hira ‘how many, several’, 32, 65
 ida ‘one, a’, 30
 reduplicated, 21–22
numeral classifier, 15, 39–40
numeral phrase, 39
 function, 25, 60
nuné ‘so, therefore’, 115
nunka ‘never’, 88
nusá ‘why?, how is it?’, 66

- object, 53
 - following long recipient, 52
 - fronting, 53–54
 - incorporation, 55
 - questioned, 65
- oblique argument, 51
- oinsá ‘how?’, 66
- ona* ‘ANTERIOR’, 75–76
- ordinal number, 39
- orthography, 6–7
- ou* ‘or’, 105
- para* ‘so that’, 107, 109, 112
- para* ‘so, very’, 90
- parataxis, 103
- parts of speech. *See* word classes
- part–whole relations, 34
- passive (none), 52
- passive-like constructions, 52
- perfect aspect, 76
- perfective aspect, 74–75
- periphery of clause, 71–72
- personal pronoun, 26–27
- phonology, 7–13
- plural marking in NP, 30–31
- polar interrogative, 67–68
- Portuguese, 1–4, 5–6, 18–19, 40
 - calque, 52, 57, 58, 64, 108
 - in morphology, 20, 36, 98
 - influence on phonology, 7–13
 - loan, 30–31, 53, 57, 59, 73, 87, 88, 90, 116, 120, 125. *See* gender
- possession, 33–36
 - in compounds, 23
 - location noun, 41–43
- possessive clause, 55–56
- predicate, 51–52, 57–61
- prefix, 19–20, 97–98
- preposition, 16, 40–50, 95
- prepositional phrase, 41
 - function, 25, 60, 71–72, 92
- pro-clause, 16
- progressive aspect, 80–81
- prohibition, 69
- pronoun, 14, 26–32
- purké* ‘because’, 111
- purpose clause, 112–13
- quantifier, 31
- quantifying adjective, 15
 - reduplicated, 21
- question, 64–68
- quote, 107
- reason clause, 111–12
- recipient, 43, 51, 61, 100
 - questioned, 65
- reciprocal, 61–63
- reduced complement, 108–9
- reduplication, 12, 21–22
- reflexive, 63–64
- register, 3–4
- relative clause, 25, 115–117
- result serial verb, 100–101
- ruma* ‘some or other’, 32
- s ‘PLURAL’, 30–31
- sá* ‘what?’, 29, 65
- saida* ‘what?’, 65
- se* ‘if’, 113
- sé* ‘who?’, 64
- sei* ‘FUTURE, still’, 82–83
- seidauk* ‘not yet’, 68, 81, 88
- seluk* ‘other, another’, 32
- sentential complement, 106–8
- serial verb construction, 92–101
- sira* ‘PLURAL’, 30–31
- source (location), 44–45
- spelling, 6–7
- stress, 12
- subject, 52–53
 - of ‘body-good’ expression, 58
 - questioned, 65
- subject marking (none), 51
- subordinate clause. *See* adverbial clause, complement clause
- suffix, 20
- superlative, 90
- syllable, 13
- tag, 16
- tail–head linkage, 115
- tán* ‘because’, 50, 111
- tanba* ‘because’, 50, 111
- tanba sá* ‘why?’, 66
- tansá* ‘why?’, 66
- tebes* ‘truly’, 90
- tebe-tebes* ‘truly’, 90
- temporal clause, 103, 110–11

- tén* 'AGENT', 20
- tenki* 'must', 87
- tenki sér* 'must', 87
- tense–aspect, 73–74
- tense–aspect markers, 73–84
- Tetun Terik, 1, 2, 4
 - differences, 5, 6, 11, 17, 18, 19, 27, 29, 33, 36, 38, 43, 45, 47, 48, 51, 53, 57, 58, 63, 65, 66, 69, 75, 86, 89, 95, 96, 98, 104, 106, 108, 110, 113, 115, 116
- tiha* 'PERFECTIVE', 74–75
- tiha ona* 'PERFECT', 76
- title, 26
- tó* 'until, reach', 45, 47–48
- topic, 102
- topicalisation, 53–54
- tuir* 'according to', 50
- tuir mai* 'and next', 114
- variation, 3
- verb, 14–15, 51
 - as root for derivation, 19–21, 24, 97–98
 - derived, 17–20
- verb phrase (V+O), 45
 - as 'possessor', 35
- voice (none), 52
- vowel, 11–12
 - length, 12
 - nasalisation, 12
 - sequence, 12
- word classes, 13–16
- zero derivation, 18–19