

**The Kelabit Language,
Austronesian Voice and
Syntactic Typology**

Charlotte Hemmings

Thesis submitted for the Degree of PhD

2016

Department of Linguistics
SOAS, University of London

Declaration

I have read and understood regulation 17.9 of the Regulations for students of SOAS, University of London, concerning plagiarism. I undertake that all the material presented for examination is my own work and has not been written for me, in whole or in part, by any other person. I also undertake that any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in this work which I present for examination.

Signed:.....

Date:.....

Abstract

Western Austronesian languages are typically defined in contrast to Oceanic languages as possessing a system of ‘symmetrical’ voice alternations (Himmelmann 2005a). These are alternations in the mapping of predicate arguments to grammatical functions but, unlike passives and antipassives, do not involve syntactic detransitivisation. Instead, symmetrical voice systems appear to involve multiple transitive clause-types that are equally morphologically marked and equally syntactically transitive. This has prompted two major debates about Western Austronesian syntax, namely whether or not Western Austronesian languages have a grammatical subject, and the nature of alignment in the languages.

Western-Austronesian languages are typically subdivided into Philippine-type languages and Indonesian-type languages on the basis of structural properties. Philippine-type languages are considered more conservative and Indonesian-type languages more innovative. The Apad Uat subgroup of Northern Sarawak, which includes Kelabit, is said to be split between Philippine-type and Indonesian-type languages. Consequently, it presents a unique opportunity to enter into the theoretical debates and also to question whether the existing typology can capture the full extent of variation within Western Austronesian.

Using naturalistic and elicited materials gathered over six and a half months of linguistic fieldwork, this thesis presents an analysis of Kelabit grammar alongside

three case studies of syntactic phenomena known to differ in Philippine-type and Indonesian-type languages: voice systems; pronominal systems and word order. In each instance, the patterns in Kelabit are neither proto-typically Philippine-type, nor proto-typically Indonesian-type and hence constitute a type of their own. Moreover, they provide support for theories of alignment shift and other syntactic changes that begin with the reanalysis of the actor voice construction. Thus, it becomes apparent that the existing two-way typology is insufficient to model syntactic variation in Western Austronesian and that a more fine-grained approach is needed in order to better understand the synchronic and diachronic landscape.

Acknowledgements

I am very grateful to the Wolfson Foundation for their generous financial support of the project. I am also grateful for grants from the Linguistics Society of America and from SOAS, University of London, which have allowed me to attend conferences and summer schools and receive feedback on my work.

Most of all, I would like to thank the Kelabit community for sharing their language and their homes with me over the last few years. In particular, I would like to thank Florance Lapu Apu @ Ngabi Tepun for kindly assisting me in documenting the Kelabit language and for tireless hours of teaching me Kelabit, answering questions about the language and transcribing and translating recorded texts.¹ I would also like to thank Lucy Bulan @ Sineh Raben and Jeffrey Malang @ Tadun Bala for their assistance in Pa' Umor and Pa' Dalih and the many people who sat and explained Kelabit grammar to me, including Lucy Bulan @ Sineh Raben Bala, David Labang @ Raben Bala, David Lian Labang, Dr Poline Bala, Julia Raja @ Sineh Raja Tuan, Lynette Smith, Datin Garnette Ridu @ Sineh La'ah Aren, Gerawat Nulun @ Udan Turun, Mary Lian, Jeffrey Malang @ Tadun Bala, Connie Aping Trang, Lucie Palan Trang, John Terawe, Yahya Talla and Kelvin Egay. In addition, I would like to thank everyone who gave up countless hours to record stories, conversations and other texts, including Florance Lapu Apu @ Ngabi Tepun, Lucy Bulan @ Sineh Raben Bala,

¹ Kelabit names are introduced by an @ sign.

David Labang @ Raben Bala, Gerawat Nulun @ Udan Turun, Connie Aping Trang, Lucie Palan Trang, Sineh Pasang, Sineh Maren Lugun @ Dooq Ileh, Sineh Tingang @ Liat Ayu, Sineh Ngelawan Paren, Henrik Iboh @ Luun Anid, Dara Tigan, Jeffrey Malang @ Tadun Bala, Melamud Tepun, Sineh Bela'an Paran, Upai Raja @ Pu'un Maran, Paran Ating, Bela'an Paran, Riboh Ayu, Gerawat Riboh @ Paran Bala, Zac Lawai Labang, Jeanette Ulun @ Ratu Aren, Sineh Maran Lugun, Muda Tigan, Sineh Ngalap Ulun, Alaw Tungen @ Kera'i' Aren, Stanley Isaac Iboh @ Paren Kera'i', Lian Terawe, Muring Apu, Frazier Kaya, Daud Iboh @ Balang Gerau', David Lian Tadun @ Ribuh Tepun and Bulan Ipang @ Pu'un Maran. Finally, I would like to thank Penghulu Robertsen Bala @ Balang Masiu and the Rurum Kelabit for supporting my research in the Kelabit Highlands. I made so many wonderful friends during my time in the Kelabit Highlands and am very grateful to all of you. *Mulaq-mulaq terima kasih ngen muyuh ngabi!*

There are numerous others who helped me during my time in Malaysia. I would like to express my particular gratitude to Datuk Robert Ridu @ La'ah Aren, Datin Garnette Ridu @ Sineh La'ah Aren, Dr Poline Bala, Marilyn Raja, Lilla Raja, Lucy Bulan @ Sineh Raben Bala, David Labang @ Raben Bala, Dr Philip Raja, Datin Sri Mariam Balan, Dato Gerawat Gala, Datin Esther Balan, Datu Robert Lian @ Balang Ngalibun, Datin Dr Katharine Pearce, Dato Isaac Lugun, Datin Nikki Lugun, Datin Valerie Mashman and Datu Ose Murang for all their help in accommodating me, introducing me to many friends in Malaysia and helping me prepare for my research. Special thanks also go to Julia Raja and Lynette Smith for helping to teach me some useful Kelabit phrases before I arrived in Malaysia and for putting me in touch with family and friends.

I am also very grateful to those who helped me with institutional support whilst in Malaysia. Firstly, I thank the Sarawak State Planning Unit for approving my application to conduct research in Sarawak. Secondly, I would like to thank the Miri Resident for approving extended periods of research in the Kelabit Highlands. Thirdly, I am grateful to Ipoi Datan and the staff at the Sarawak Museum, who helped me to obtain a research visa and gave me access to their wonderful archival resources. Fourthly, I would like to thank Dr Norazuna Bt Norahim and colleagues at the UNIMAS Centre for Languages who offered help and advice whilst in Kuching. Particular thanks go to all the participants who recorded pear stories in Kuching Sarawak Malay and to Mohamad Fairuz Ali Ayen for kindly transcribing and translating the texts. Furthermore, I would like to thank Miss Bibi Aminah Abdul Ghani and colleagues at Curtin University, Sarawak Malaysia for their help and guidance whilst in Miri and all the participants who recorded pear stories in Miri Sarawak Malay. Many thanks to Prof. Dr Alvin Yeo, Dr Soubakeavathi Rethinasamy and colleagues at the Institute of Social Informatics and Technological Innovations, UNIMAS, Sarawak, for inviting me to present at the ISITI, and sharing their experiences of working with languages in Sarawak. Finally, thanks go to Dr Roger Harris for inviting me to join the eBario Knowledge Fair in November 2013.

Next, I would like to thank the many researchers who have previously worked in Sarawak for their helpful guidance and support throughout the project. Firstly, I would like to thank Monica Janowski, for introducing me to Bario, encouraging me to help document the Kelabit language and readily sharing her expertise along the way. Thanks also go to Borbala Nyiri, Lindsay-Lloyd Smith and Dr Jayl Langub for their willingness to answer practical questions and offer guidance. I would like to thank Matthew Amster for sending me a copy of the Kelabit Dictionary and Professor Robert

Blust, Dr Mary Tay and Professor Amsah Haji Omar for sharing their previous knowledge of working with the Kelabit language. I am grateful to Dr Wallace Chafe and Dr Jay Crain for sharing texts in Malay and Lun Bawang and to Colin Davis for making available his map of the Kelabit Highlands. Finally, I would like to give particular thanks to Beatrice Clayre for giving me feedback on my research and for our many fruitful discussions on Sa'ban, Kelabit and Lundayeh.

I owe many thanks to the Department of Linguistics at SOAS, University of London. In particular, I would like to acknowledge the support of my Supervisor, Peter Austin, who has given me guidance and feedback throughout my research. In addition, thanks go to my supervisory committee, Friederike Lüpke and Chris Lucas, for advice on preparation for fieldwork and academic writing during the upgrade process and throughout the PhD. Moreover, I would like to especially thank Candide Simard for invaluable advice on prosody and her contribution and guidance in relation to my work on clitic pronouns. Further thanks are due to Mandana Seyfeddinipur for encouraging me to apply to the Linguistics Society of America summer school where I received training in documentation, typology and theory, and to Irina Nikolaeva and Lutz Marten for insightful feedback on presentations at SOAS. I would also like to thank my fellow students at SOAS, particularly those involved in Language Landscape and the SOAS Grammar Group. It has been a great pleasure to know everyone who has made my experience at SOAS so special. Particular thanks go to Eleanor Ridge and Lauren Gawne for feedback on drafts of the thesis. Finally, I would like to thank Pak Sallehuddin bin Abdullah Sani for teaching me Indonesian and checking through the Indonesian examples in the thesis.

I am also very grateful to the wider academic community and all of the linguists who have taken time to comment on my research in progress. Firstly, I would

like to thank my two examiners, Mary Dalrymple and Elizabeth Zeitoun, who gave me many detailed comments and helped to improve the thesis no end. Secondly, I would like to thank the members of the South of England Lexical Functional Grammar (SE-LFG) group, including Mary Dalrymple and Louise Mycock, who shared their knowledge of syntactic theory and information structure. Thirdly, I would like to thank participants in the Linguistics Society of America Summer Institute 2015, particularly Edith Aldridge, Maria Polinsky, Amy Dahlstrom and Monica Macaulay, for their comments on Austronesian syntax, word order and Algonquian languages. Fourthly, I would like to thank I Wayan Arka, Paul Kroeger, René van den Berg and audiences at the 7th Austronesian and Papuan Languages and Linguistics (APLL7) Conference and the 12th International Borneo Research Council Conference for feedback on earlier versions of the work relating to voice in CHAPTER 3, as well as two anonymous reviewers from the Transactions of the Philological Society.

Finally, I wish to acknowledge the support of friends and family and loved ones who have supported me in everything I do. In particular, I am very grateful to Wendy Hemmings for proof reading the entire thesis and to Richard Hemmings, James Hemmings and Francis Newall for their support and patience over the years. It would not have been possible without you.

Table of Contents

Declaration	1
Abstract	2
Acknowledgements	4
Table of Contents	9
List of Tables	20
List of Figures	24
Abbreviations and Conventions	25
1. Introduction	29
1.1 Introduction.....	29
1.2 The Austronesian Language Family.....	33
1.3 Western Austronesian Voice.....	36
1.3.1 Philippine-type vs Indonesian-type.....	40
1.3.2 Asymmetrical Austronesian Voice Systems.....	46
1.3.3 Summary.....	48
1.4 Key Debates within Austronesian Syntax.....	49
1.4.1 The Subject Debate.....	49
1.4.1.1 Subject in Tagalog.....	50
1.4.1.2 Subject in Indonesian.....	52

1.4.1.3 Previous Accounts.....	54
1.4.2 The Alignment Debate.....	57
1.4.2.1 The Alignment Debate in Philippine-type Languages.....	59
1.4.2.1.1 The Accusative Hypothesis.....	59
1.4.2.1.2 The Ergative Hypothesis.....	60
1.4.2.1.3 The Philippine-type Alignment Hypothesis.....	65
1.4.2.2 The Alignment Debate in Indonesian-type Languages.....	68
1.4.2.2.1 The Accusative Hypothesis.....	68
1.4.2.2.2 The Ergative Hypothesis.....	70
1.4.2.2.3 The Symmetrical Hypothesis.....	71
1.4.2.3 Theory of Alignment Shift.....	74
1.4.3 Summary.....	75
1.5 Structure of the Thesis.....	76
1.6 Conclusion.....	77
2. The Kelabit Language.....	79
2.1 Introduction.....	79
2.2 The Kelabit Language.....	80
2.2.1 Classification.....	80
2.2.2 Dialect Geography.....	87
2.2.3 Ethnography.....	91
2.2.4 Language Vitality.....	92
2.3 Phonology and Orthography.....	94
2.3.1 Vowels.....	96
2.3.2 Consonants.....	98

2.3.3 Syllable Structure.....	101
2.3.4 Prosody and Stress.....	103
2.3.5 Phonological Processes.....	103
2.3.5.1 Nasal Assimilation and Nasal Substitution.....	103
2.3.5.2 Diphthongisation.....	105
2.3.5.3 Consonant Gemination.....	106
2.3.5.4 Vowel Reduction.....	107
2.3.5.5 Vowel Deletion.....	108
2.3.6 Summary.....	108
2.4 Morphology.....	109
2.4.1 Word Formation.....	110
2.4.1.1 Prefixation.....	110
2.4.1.1.1 <i>deN-</i>	111
2.4.1.1.2 <i>ke₁-</i>	112
2.4.1.1.3 <i>ke₂-</i>	113
2.4.1.1.4 <i>N-</i>	114
2.4.1.1.5 <i>ne₁</i>	117
2.4.1.1.6 <i>ne₂-</i>	118
2.4.1.1.7 <i>pe-</i>	119
2.4.1.1.8 <i>peN-</i>	122
2.4.1.1.9 <i>peneN-</i>	124
2.4.1.1.10 <i>pere-</i>	124
2.4.1.1.11 <i>se-</i>	125
2.4.1.1.12 <i>seN-</i>	126
2.4.1.1.13 <i>te₁-</i>	127

2.4.1.1.14 <i>te</i> ₂ -	129
2.4.1.2 Infixation.....	129
2.4.1.2.1 <i>-em-</i>	130
2.4.1.2.2 <i>-um-</i>	132
2.4.1.2.3 <i>-in-</i>	132
2.4.1.3 Suffixation.....	136
2.4.1.3.1 <i>-en</i>	136
2.4.1.3.2 <i>-um</i>	142
2.4.1.3.3 <i>-q, -m, -n</i>	143
2.4.1.3.4 <i>-an</i>	145
2.4.1.4 Reduplication.....	146
2.4.1.4.1 Reduplication of Nominal Roots.....	146
2.4.1.4.2 Reduplication of Adjectival Roots.....	147
2.4.1.4.3 Reduplication of Verbal Roots/Stems.....	149
2.4.2 Major Word Classes.....	151
2.4.2.1 Nouns.....	152
2.4.2.2 Verbs.....	155
2.4.2.2.1 Intransitive Verbs.....	156
2.4.2.2.2 Transitive Verbs.....	159
2.4.2.2.3 Ditransitive Verbs.....	160
2.4.2.2.4 Distributional Characteristics.....	161
2.4.2.3 Adjectives.....	162
2.4.2.4 Adverbs.....	165
2.4.2.5 Prepositions.....	166
2.4.2.6 Auxiliaries.....	168

2.4.2.7 Deictic Terms.....	174
2.4.2.8 Pronouns.....	177
2.4.2.8.1 Possessive Pronouns.....	178
2.4.2.8.2 Impersonal Pronouns.....	181
2.4.2.8.3 Emphatic Pronouns.....	183
2.4.2.8.4 Inclusive Pronouns.....	185
2.4.2.9 Interrogatives.....	186
2.4.2.10 Relativisers.....	188
2.4.2.11 Conjunctions.....	190
2.4.2.12 Numerals.....	190
2.4.2.13 Quantifiers and Negators.....	192
2.4.2.14 Particles.....	193
2.4.2.14.1 Sentence Particles.....	193
2.4.2.14.2 Clause-final Particles.....	198
2.4.3 Summary.....	199
2.5 Syntax.....	199
2.5.1 Grammatical Functions.....	200
2.5.1.1 Subject.....	201
2.5.1.2 Non-subject Core Arguments.....	207
2.5.1.3 Obliques and Adjuncts.....	210
2.5.2 Periphrastic Voices.....	212
2.5.2.1 <i>Inan</i> Clauses.....	213
2.5.2.2 <i>Tu'en</i> Clauses.....	215
2.5.3 Multi-clausal Constructions.....	217
2.5.3.1 Relative Clauses.....	217

2.5.3.2 Cleft Constructions.....	220
2.5.3.3 Complement Clauses.....	222
2.5.3.4 Adjunct Clauses.....	226
2.5.3.5 Co-ordination.....	227
2.5.4 Summary.....	230
2.6 Conclusion.....	231
3. Voice Alternations.....	232
3.1 Introduction.....	232
3.2 Voice.....	233
3.2.1 Alternations in Syntactic Functions of Arguments.....	234
3.2.1.1 Active/Passive.....	235
3.2.1.2 Ergative/Antipassive.....	237
3.2.1.3 Western Austronesian Symmetrical Voice.....	241
3.2.2 Alternations in Semantic Properties of Arguments.....	243
3.2.2.1 Active/Middle.....	243
3.2.3 Alternations in Pragmatic Salience of Arguments.....	245
3.2.3.1 Inverse Systems.....	246
3.2.3.2 Focus Systems.....	251
3.2.3.3 Subject-Object Reversal.....	256
3.2.4 Summary.....	261
3.3 Methodology for Studying Voice.....	262
3.3.1 Morphosyntax.....	263
3.3.2 Semantics.....	264
3.3.3 Discourse.....	266

3.3.4 Summary.....	267
3.4 Western Austronesian Voice.....	268
3.4.1 Philippine-type.....	269
3.4.1.1 Morphosyntax.....	271
3.4.1.2 Semantics.....	274
3.4.1.3 Discourse.....	277
3.4.2 Indonesian-type.....	279
3.4.2.1 Morphosyntax.....	280
3.4.2.2 Semantics.....	284
3.4.2.3 Discourse.....	287
3.4.3 Borneo and Sulawesi.....	290
3.4.3.1 Morphosyntax.....	291
3.4.3.2 Semantics.....	297
3.4.3.3 Discourse.....	299
3.4.4 Summary.....	301
3.5 Kelabit Voice.....	302
3.5.1 Morphosyntax.....	302
3.5.2 Semantics.....	305
3.5.3 Discourse.....	312
3.5.4 Summary.....	314
3.6 Conclusion.....	316
4. Pronominal Systems.....	320
4.1 Introduction.....	320
4.2 Kelabit Pronouns.....	322

4.2.1 Previous Accounts.....	324
4.2.1.1 Case-based Analyses	324
4.2.1.2 Voice-based Analyses	326
4.2.2 Extending the Analysis to Kelabit.....	328
4.2.3 A Prosodic Distinction.....	333
4.2.4 Summary.....	336
4.3 Clitics in the Literature.....	337
4.3.1 Clitic Subtypes.....	339
4.3.2 Austronesian Clitic Phenomena.....	343
4.3.3 Possible Kelabit Clitic Patterns.....	350
4.3.4 Summary.....	353
4.4 Methodology for Identifying Clitics.....	353
4.4.1 Prosody.....	354
4.4.2 Prosodic Boundary Marking.....	357
4.4.3 Formulating Predictions.....	359
4.4.4 Experimental Design.....	360
4.4.5 Procedure for Data Collection.....	365
4.4.6 Data Analysis.....	367
4.4.7 Summary.....	369
4.5 Results.....	370
4.5.1 Word Boundary Effects in Kelabit.....	370
4.5.2 Comparing FORM 1 and FORM 2	375
4.5.2.1 Transitive Predicates.....	375
4.5.2.1.1 Statistical Analysis.....	381
4.5.2.2 Intransitive Predicates.....	382

4.5.2.2.1 Statistical Analysis.....	384
4.5.3 Comparing FORM 1 in all Contexts.....	385
4.5.4 Summary.....	389
4.6 Discussion	390
4.6.1 Prosody-Syntax Mismatch.....	390
4.6.2 Austronesian Clitic Typology.....	395
4.6.3 Summary.....	397
4.7 Conclusion.....	398
5. Word Order.....	400
5.1 Introduction.....	400
5.2 Word Order.....	402
5.2.1 Greenberg and the Six-way Typology.....	402
5.2.2 Dryer and the VO/OV Typology.....	404
5.2.3 Typological Correlations with Word Order.....	407
5.2.3.1 Verb-initial Languages.....	407
5.2.3.2 SVO Languages.....	409
5.2.4 Basic and Alternative Word Orders.....	410
5.2.5 Summary.....	413
5.3 Methodology for Studying Word Order.....	414
5.3.1 Describing Possible Word Orders.....	414
5.3.2 Establishing Basic Word Order.....	415
5.3.3 Analysing the Contexts in which Different Word Orders Occur.....	417
5.3.4 Summary.....	425
5.4 Word Order in Western Austronesian.....	425

5.4.1 Philippine-type languages.....	426
5.4.2 Indonesian-type languages.....	434
5.4.3 Transitional languages.....	438
5.4.4 Historical Change in Word Order.....	442
5.4.5 Summary.....	443
5.5 Word Order in Kelabit.....	444
5.5.1 Possible Word Orders.....	444
5.5.1.1 Intransitive Clauses.....	445
5.5.1.2 Actor Voice Clauses.....	447
5.5.1.3 Undergoer Voice Clauses.....	450
5.5.1.4 Bare Verbs.....	453
5.5.2 Establishing Basic Order.....	454
5.5.2.1 Frequency in Narratives.....	456
5.5.2.2 Frequency in News Reports.....	458
5.5.3 Factors Affecting Word Order in Kelabit.....	460
5.5.3.1 Information Structure and Elicitation.....	460
5.5.3.2 Information Structure in UV.....	465
5.5.3.3 Information Structure in AV.....	469
5.5.4 Summary.....	475
5.6 Conclusion.....	475
6. Conclusion.....	479
6.1 Introduction.....	479
6.2 Western Austronesian Typology.....	481
6.2.1 Voice and Alignment.....	483

6.2.2 Pronouns and Clitic Systems.....	487
6.2.3 Word Order.....	489
6.2.4 Summary.....	491
6.3 Wider Theoretical and Historical Debates.....	493
6.3.1 The Subject Debate.....	493
6.3.2 The Alignment Debate.....	495
6.3.3 Summary.....	497
6.4 Future Research.....	498
6.5 Conclusion.....	501
Bibliography.....	502
Appendix 1: Documentation and Description of Kelabit.....	559
A1.1 Introduction.....	559
A1.2 Previous Documentation.....	559
A1.3 The Kelabit Corpus.....	561
A1.3.1 Elicitation.....	562
A1.3.2 Experiment.....	565
A1.3.3 Text.....	566
A1.3.4 Speakers.....	569
A1.4 Summary.....	570
Appendix 2: Prosody Experiment.....	571
Appendix 3: Example Texts.....	579
A3.1 Folk Narratives.....	579
A3.2 News Reports.....	592

List of Tables

1.1 Tagalog Voice Morphology (Himmelmann 2002).....	40
1.2 Defining Characteristics of Philippine-type and Indonesian-type (Arka 2002)	41
1.3 Tagalog Subject Tests (Schachter 1976).....	52
1.4 Indonesian Subject Tests (Riesberg 2014).....	54
1.5 Tagalog Verbal Marking in Aldridge (2004).....	61
1.6 Tagalog Nominal Marking in Aldridge (2004).....	61
1.7 Core Properties in Indonesian (Arka 2005).....	73
2.1 Vowel Inventory in Kelabit.....	95
2.2 Consonant Inventory in Kelabit.....	95
2.3 Kelabit Orthography.....	96
2.4 Prefixes in Kelabit.....	111
2.5 Ordinal Numerals in Kelabit.....	114
2.6 Distributive Numerals in Kelabit.....	129
2.7 Infixes in Kelabit.....	130
2.8 Suffixes in Kelabit.....	136
2.9 Bound Possessive Pronouns.....	143
2.10 Prepositions in Kelabit.....	167
2.11 Kelabit FORM 1 Pronouns.....	177

2.12 Kelabit FORM 2 Pronouns.....	177
2.13 Kelabit Possessive Pronouns.....	180
2.14 Kelabit Emphatic Pronouns.....	183
2.15 Cardinal Numbers.....	191
3.1 Middle Situations (Kemmer 1994).....	244
3.2 Transitivity Parameters following Hopper & Thompson (1980).....	264
3.3 Topicality of Arguments (Cooreman 1987).....	266
3.4 Methodology for Comparing Voices.....	268
3.5 Proto-Austronesian Voice (Adelaar 2005: 6, following Ross 2002).....	269
3.6 Tagalog Verbal Paradigm (Katagiri 2005: 159).....	272
3.7 Scaled Average Topicality (SAT) of Arguments in Cebuano (Walters 1994: 134).....	278
3.8 Philippine-type Voice Systems.....	278
3.9 Scaled Average Topicality (SAT) in Indonesian.....	279
3.10 Voice in Indonesian.....	290
3.11 Transitivity in Uma (van den Berg 1995).....	293
3.12 Referential Distance in Pendau (Quick 2005: 230-231).....	299
3.13 Transitional Voice Systems.....	301
3.14 Kelabit Voice Markers.....	302
3.15 Kelabit Semantic Transitivity.....	306
3.16 Kelabit Actor Voice Semantics.....	308
3.17 Frequency of Voices.....	312
3.18 Topicality of Arguments in Kelabit.....	313
3.19 Cross-linguistic Topicality.....	313

3.20 Kelabit Structural Properties.....	315
3.21 Kelabit Voice.....	316
4.1 Kelabit FORM 1 Pronouns.....	322
4.2 Kelabit FORM 2 Pronouns.....	323
4.3 Kelabit Oblique Pronouns.....	323
4.4 Proto-Southwest Sabah (Lobel 2013: 103).....	324
4.5 Lundayeh Pronoun Sets (Clayre 2005: 24).....	326
4.6 Summarising Lundayeh Pronouns.....	327
4.7 An Ergative Analysis of Lundayeh.....	328
4.8 Comparing Lundayeh and Kelabit.....	331
4.9 Clitics, Words and Affixes.....	338
4.10 Simple Clitics, Special Clitics and Bound Words.....	342
4.11 Proto-Austronesian and Proto-Malayo Polynesian Clitic Pronouns (Ross 2006)	344
4.12 The Re-analysis of Austronesian Clitics.....	350
4.13 Prosodic Features and Acoustic Correlates.....	354
4.14 The Strict Layer Hypothesis.....	355
4.15 Variables in Prosody Study.....	361
4.16 Test Sentences per Context.....	365
4.17 Word-final Lengthening in Kelabit.....	370
4.18 Syllable Duration.....	372
4.19 Word-final Pitch effects in Kelabit.....	374
4.20 Mean F ₀	374
4.21 Raw Duration (ms) of Syllables in Context 3 & 4.....	377

4.22 Mean Duration (ms) of Syllables in Context 3 & 4.....	377
4.23 Mean Duration of Syllables in Context 3 & 4 - possible errors removed.....	381
4.24 Mean Duration of Syllables in Context 1 & 2.....	383
4.25 Mean Duration of Pronouns by Context.....	387
4.26 Mean Duration according to Position.....	387
4.27 Mean Duration according to Function.....	388
4.28 Mean Duration according to Voice.....	388
5.1 Distribution of Word Orders.....	404
5.2 Word Order in Hanis Coos (Dryer 1983).....	405
5.3 Typological Properties of Verb-Initial Languages (Dryer 1997).....	408
5.4 Comparing Verb-final, SVO and Verb-initial Languages (Dryer 1991).....	409
5.5 Word Order in Sama-Bajau Languages (adapted from Miller 2014: 306).....	439
5.6 Word Order Elicitation Game.....	445
5.7 Word Order in Narrative Intransitive Clauses.....	457
5.8 Word Order in Narrative Transitive Clauses.....	457
5.9 Word Order in Narratives by Voice Construction.....	457
5.10 Word Order in News Report Intransitive Clauses.....	458
5.11 Word Order in News Report Transitive Clauses.....	458
5.12 Word Order in News Reports by Voice Construction.....	458
6.1 Default and Non-Default Clauses.....	496
A1.1 Text Recordings.....	567
A1.2 Speakers by Place of Birth.....	569

List of Figures

1.1 The Austronesian Language Family © Encyclopaedia Britannica.....	34
1.2 Philippine-type vs Indonesian-type.....	46
2.1 The Languages of Malaysia (Lewis et al. 2016).....	83
2.2 The Languages of Sarawak (Lewis et al. 2016).....	84
2.3 Apad Uat Family Tree (cf. Blust 1993).....	86
2.4 The Kelabit-Kerayan Highlands © Colin Davis.....	89
4.1 Boxplot of Duration by Syllable Position.....	371
4.2 Boxplot of Duration by Syllable Position, separating Single Syllable Words	373
4.3 Boxplot of Mean F_0 by Syllable Position.....	374
4.4 Boxplot of Mean F_0 by Syllable Position, separating Single Syllable Words	375
4.5 Mean Duration of Syllables in Context 3 & 4.....	378
4.6 Mean Duration of Syllables in Context 1 & 2.....	384

Abbreviations and Conventions

In this thesis, the following conventions are adopted. Example sentences are glossed using the Leipzig Glossing Rules. Where examples are drawn from the literature, the gloss is adopted from the source with the following exceptions. Firstly, in order to facilitate comparison between Austronesian languages, verbal morphology glosses are adapted to AV and UV etc., except in sections outlining alternative analyses. Similarly, nominal morphology glosses are adapted to NOM/GEN or SUBJ/CORE/NON.SUBJ. These can both be understood to reflect an analysis of Western Austronesian languages as morphosyntactically symmetrical, which is motivated in the thesis (see SUBSECTION 1.3).

There are also primary examples in English, Kelabit, Javanese and Indonesian. English examples are based on my native speaker judgements, unless otherwise specified. Indonesian examples are adapted from published sources, including Musgrave (2002) and Shiohara (2012) and were checked by native speakers in London. Javanese examples were elicited during MA research into Javanese morphosyntax in 2011-2012 and are courtesy of Nanang Endrayanto. Finally, the Kelabit examples are drawn from the documentary corpus collected during the PhD, which is described in APPENDIX 1. Audio and video-recorded examples are given a reference specifying the data source (i.e. text vs. elicitation), the filename (in the format PPPDDMMYYYYRR_00, where PPP is a code for the place of the recording,

DDMMYYYY is the date of recording, RR is a code for the researcher who collected the recording and 00 the recording number) and a timecode (in the format 00:00:00.000-00:00:00.000). Hence, the reference in (1) can be understood as follows:

(1) **text, BAR22102013CH_04 00:05:33.310-00:05:35.850**

An example from a naturalistic text, collected by Charlotte Hemmings in Bario on 22/10/2013. The example begins approximately 5 minutes 33 seconds into the recording and ends at approximately 5 minutes 35 seconds.

Other place codes include PDA for Pa' Dalih and PUM for Pa' Umur. This reflects the place of recording and not necessarily the dialect of the speaker. Elicited examples from written fieldnotes rather than recorded elicitation sessions are given the notation (elicitation, fieldnotes). Unless otherwise specified, single-word examples are all taken from the documentary corpus.

All examples are glossed consistently using the abbreviations listed in the table below. Where morpheme boundaries are not relevant to the analysis, they are not represented. For example, the auxiliary *mileh* 'be able' can be subdivided into the root *ileh* 'knowledge' and the intransitive verb forming infix *-em-*. However, it is typically glossed simply as 'able' rather than *m-ileh* 'INTR-knowledge'.

As discussed in CHAPTER 2, full reduplication is a common word-formation strategy in Kelabit. As the entire root is reduplicated, it is difficult to know whether the reduplicated element follows or precedes the stem. By convention, reduplicated forms in Kelabit are glossed REDUP~stem on the basis that partial reduplication is prefixed to the stem. However, it could equally be understood as stem~REDUP, as is typical in the study of Indonesian (see Dalrymple & Mofu 2012). Further research is needed to explore which analysis is preferable for full reduplication.

Finally, Kelabit has a set of variant pronouns that are referred to as FORM 1 and FORM 2 (see SUBSECTION 2.4.2.8). A first singular FORM 1 pronoun is indicated in the gloss using 1SG.1 and a first singular FORM 2 pronoun with 1SG.2. These have some similarities with nominative and genitive pronouns in Philippine-type languages, as discussed in CHAPTER 4. However, they also differ from other Western Austronesian languages and therefore the more neutral glosses of FORM 1 and FORM 2 are adopted, following Clayre (2005). The details of the analysis are explained in CHAPTER 4. All other abbreviations are explained within the main body of the text.

1	first person	DET	determiner	NEG.DESI D	negative desiderative
2	second person	DIR	direct	NOM	nominative case
3	third person	DISTR	distributive	NON.FUT	non-future
3'	third obviative	DU	dual	NON.PST	non-past
ABIL	abilitative	DV	dative voice	NON.SER	non-serious
ABL	ablative case	EMPH	emphatic	NON.SUBJ	non-subject
ABS	absolute case	EQUATIVE	equative	OBJ	object
ACC	accusative case	ERG	ergative case	OBL	oblique
ACCID	accidental	EXCL	exclusive	PART	partitive
ACT	active	EXIST	existential	PASS	passive
ADV	adverbial	F	feminine	PAU	paucal
AF	actor focus	FAM	familiar	PERS	person
AGR	agreement	FUT	future	PFV	perfective
ANIM	animate	GEN	genitive	PL	plural
ANTIF	antifocus	HON	honorific	POSS	possessive
ANTIP	antipassive	IF	instrument focus	PREP	preposition
AOR	aorist	IMP	imperative	PRO	pronoun
APPL	applicative	IMPERS	impersonal pronoun	PRS	present
ART	article	INCL	inclusive	PST	past
ASP	aspect	IND	indicative	PT	particle
ASSOC	associative	INDP	independent order	PTCP	participle
AUX	auxiliary	INF	infinitive	Q	question
AV	actor voice	INS	instrumental	REAL	realis
BV	benefactive voice	INTR	intransitive	RECP	reciprocal
CAUS	causative	INV	inverse	REC.PST	recent past
CLF	classifier	IPFV	imperfective	REDUP	reduplication
CNG	connegative	IRR	irrealis	REFL	reflexive
COM	comitative	IV	instrumental voice	REL	relative
COMT	comment	LNK	linker	SG	singular
CONJ	conjunct order	LOC	locative	STAT	stative
CONTR	contrastive	LV	locative voice	SUBJ	subject
CORE	core argument	M	masculine	SUFFIX	suffix
DAT	dative case	MED	medial	TOP	topic
DEF	definite	MIDDLE	middle voice	TR	transitive
DEM	demonstrative	N	neuter	UV	undergoer voice
DESID	desiderative	NEG	negative		

Chapter 1

Introduction

1.1 Introduction

This thesis presents a study of voice alternations and related syntactic phenomena in Kelabit, a Western Austronesian language spoken in Northern Sarawak. It provides a basic sketch of the Kelabit language (CHAPTER 2) followed by three detailed case studies of voice alternations (CHAPTER 3), pronominal systems (CHAPTER 4) and word order (CHAPTER 5). This serves as an empirical base from which to explore the position of Kelabit within the typology of Western Austronesian, and the contribution that an analysis of Kelabit can make to ongoing theoretical debates in the study of Austronesian voice systems (CHAPTER 6). The data is drawn from a documentary corpus collected over a period of six months of primary linguistic fieldwork and includes both elicited examples and naturalistic texts in a variety of genres (APPENDIX 1-3).

The study contributes to the growing literature on the typologically rare systems of verbal marking in Western Austronesian languages (Himmelmann 2005a).² These encode alternations in the mappings of semantic roles to grammatical functions

² Western Austronesian can be understood in a typological or geographical sense, as defined in SUBSECTION 1.2.

that have been the subject of considerable debate (Adelaar 2013). Today, the alternations are typically referred to as ‘voice’ (Arka & Ross 2005, Blust 2013), although they have previously been known as ‘focus’ (Clayre 1991, Boutin 1988, Healy 1958), ‘orientation’ (Svelmoe G. & T. Svelmoe 1974), ‘registration’ (Antworth 1979), ‘pivot’ (Foley & Van Valin 1984), ‘perceptual centre of the sentence’ (Starosta 1986) and ‘topic’ (McKaughan 1958).³ The many terms used to describe the alternations stem from the fact that they differ in a number of ways from the active/passive and ergative/antipassive voice alternations found in other language groups (cf. Keenan & Dryer 2006, Polinsky 2013). The main differences are:

- (1) a. The symmetrical nature of the alternations
- b. The number of alternations
- c. The relative prominence of the undergoer

Unlike active/passive and ergative/antipassive alternations, Western Austronesian voice systems do not involve either increased morphological marking or detransitivisation (Riesberg 2014, see SUBSECTION 1.3). Consequently, proto-typical Western Austronesian voice systems are often described as morphologically and syntactically ‘symmetrical’, in the sense that each voice is equally marked and has two or more core arguments (Himmelmann 2005a). Moreover, Western Austronesian voice systems often involve more than two voice alternations, and many languages have been described as ‘patient prominent’ in that definite undergoers are preferentially mapped to subject (Foley & Van Valin 1984).

³ There are also contemporary theoretical accounts that treat the alternations as case-agreement (Rackowski 2002, Rackowski & Richards 2005), transitivity marking (Starosta 2009abc, Aldridge 2011) and nominalisation (Starosta et al 1982, Kaufman 2009, Kaufman to appear). See Kroeger (2007) and Reid & Liao (2004) for critical discussion of the terminology.

The features in (1) have prompted two key debates about Western Austronesian. The first debate concerns whether the grammatical function ‘subject’ is a relevant category in Western Austronesian languages (SUBSECTION 1.4.1). The second debate centres on the behaviour of three core arguments: the actor of a transitive clause (A), the undergoer of a transitive clause (U) and the single argument of an intransitive clause (S).⁴ It concerns whether Western Austronesian languages can be said to have accusative alignment (A=S); ergative alignment (U=S) or an alternative form of alignment altogether in which both of the former alignment systems co-occur in different contexts (SUBSECTION 1.4.2). Both debates rest on the extent to which the alternations are seen as symmetrical. Thus, they have important typological and theoretical implications (see SUBSECTION 1.4).

Western Austronesian languages are typically subdivided into either Philippine-type or Indonesian-type languages on the basis of structural differences (cf. Himmelmann 2005a, Arka & Ross 2005, SUBSECTION 1.3.1). Philippine-type languages are more conservative, and are said to have preserved many of their structural properties from Proto-Austronesian (cf. Blust 2013). In contrast, the structural properties of Indonesian-type languages are generally agreed to represent historical innovation (Adelaar 2005). Both Philippine-type and Indonesian-type languages are subject to the key debates outlined above. However, they can be shown to vary in some important regards, as discussed in SUBSECTION 1.4. For this reason, it has been proposed that Western Austronesian languages differ in their degree of symmetry (see Riesberg 2014) and in their basic alignment (see Aldridge 2011). In particular, Aldridge (2011) proposes that synchronic structural differences may reflect

⁴ The symbols A, U and S are adapted from Comrie (1981) and Dixon (1994). U is sometimes written as P or O.

the fact that Western Austronesian languages have undergone a shift in alignment from ergative in the Philippine-type languages to accusative in at least some Indonesian-type languages (cf. Aldridge 2012, SUBSECTION 1.4).

The languages of Northern Sarawak fall, genetically and geographically, between the Philippine-type languages and the Indonesian-type languages (Hudson 1994, SUBSECTION 1.2).⁵ Indeed, Clayre (2005: 17) argues that the Apad Uat language subgroup, which includes Kelabit, can be divided into languages with Philippine-type characteristics, such as Lundayeh, and those that resemble Indonesian-type languages, such as Sa'ban (see CHAPTER 2). Kelabit is said to be more innovative than Lundayeh and more conservative than Sa'ban (Blust 1993, SUBSECTION 2.2.1). Hence, it would seem to be at a point of transition between the different systems. This raises two central questions. Firstly, can transitional languages like Kelabit be captured by the existing two-way typology of Philippine-type and Indonesian-type? Secondly, what can transitional languages tell us about the nature of subjecthood, alignment and theories of diachronic shift? If Western Austronesian languages have undergone largescale structural changes like those proposed by Aldridge (2012), then we might expect to find evidence of intermediate stages in the transition. If so, categorising languages as either Philippine-type or Indonesian-type may obscure further distinctions that are vital to understanding Western Austronesian languages as a whole.

This thesis addresses these questions by analysing three syntactic phenomena that are known to vary across Western Austronesian languages: verbal morphology, pronominal systems and word order. It establishes fine-grained parameters of variation and compares Kelabit with Philippine-type languages, Indonesian-type languages and

⁵ This is also true of the languages of Borneo and Sulawesi more broadly (Ross 2002).

other transitional languages in Borneo and Sulawesi. Ultimately, the thesis demonstrates that the two-way typology is not sufficient to capture the full range of possibilities within Austronesian syntax. Moreover, it supports a view of diachronic shift beginning with the reanalysis of the actor voice construction, as illustrated in CHAPTERS 3, 4 and 5.

This chapter defines key concepts and introduces the typological and theoretical accounts of Western Austronesian voice that are assessed in relation to Kelabit in this thesis. SUBSECTION 1.2 introduces the Austronesian language family and defines Western Austronesian as a typological subgroup. SUBSECTION 1.3 introduces the nature of Western Austronesian voice and the major distinction between Philippine-type and Indonesian-type. SUBSECTION 1.4 summarises the key debates within Austronesian syntax and SUBSECTION 1.5 sets out the structure for the rest of the thesis.

1.2 The Austronesian Language Family

The Austronesian language family is spread over a large geographical area from Taiwan to New Zealand and Madagascar to Easter Island (Adelaar 2005). With 1,200 languages, it is the second largest language family in the world in terms of the number of languages, though many are spoken by fewer than 1,000 speakers (Blust 2013). Though there is disagreement among Austronesianists as to origins of the Austronesian peoples, the most widely accepted theory is that they originated somewhere in Mainland China, reaching Taiwan by roughly 4,000 BC (King 1993, Bellwood 1985). From Taiwan they are thought to have moved into the Philippines,

before settling Borneo from about 2,500 BC and later moving into Indonesia, Malaysia and onwards (Bellwood 1985, King 1993: 77).⁶



Figure 1.1 The Austronesian Language Family © Encyclopædia Britannica⁷

The Austronesian family can be classified into ten primary subgroups that share the common ancestor Proto-Austronesian (Blust 2013: 30).⁸

- (2) **Primary Subgroups**
- a. *Atayalic* (Taiwan)
 - b. *East Formosan* (Taiwan)
 - c. *Puyuma* (Taiwan)
 - d. *Paiwan* (Taiwan)
 - e. *Rukai* (Taiwan)
 - f. *Tsouic* (Taiwan)
 - g. *Bunun* (Taiwan)
 - h. *Western Plains* (Taiwan)
 - i. *Northwest Formosan* (Taiwan)
 - j. *Malayo-Polynesian* (Extra-Formosan)

⁶ The theory is supported by archaeological, anthropological and linguistic evidence. Dyen (1965) and Kern (1889) present alternative, though less established, proposals such as coastal Vietnam and New Guinea (cf. Asmah 2004: 12).

⁷ Austronesian languages: major divisions of Austronesian languages [IMAGE]. Encyclopædia Britannica. Retrieved 3 March 2016, from <http://media1.library.eb.co.uk/eb-media/04/2004-004-7102F813.gif>

⁸ See Ross (2009) and Aldridge (2016) for alternative proposals.

The first 9 branches are found exclusively on Taiwan and are collectively referred to as the Formosan languages.⁹ All of the languages outside of Taiwan belong to the Malayo-Polynesian subgroup and share the common ancestor Proto Malayo-Polynesian. Malayo-Polynesian is typically further subdivided into two main branches: Western Malayo-Polynesian and Central-Eastern Malayo-Polynesian (Blust 2013: 31). Western Malayo-Polynesian includes roughly 500-600 languages spread from the Philippines across to Madagascar (see FIGURE 1.1).¹⁰

In this thesis, I refer to the Formosan and Western Malayo-Polynesian languages collectively as Western Austronesian. This is not a genetic subgroup, established by shared innovations from a proto-language. Rather it is a typological grouping that distinguishes the Austronesian languages with symmetrical voice systems from the Central-Eastern Malayo-Polynesian languages, particularly Oceanic languages, that do not tend to have this feature (SUBSECTION 1.3.2, Himmelmann 2005a).¹¹ The chapter will now discuss the nature of symmetrical voice systems, and introduce a key distinction between Philippine-type and Indonesian-type languages. More information on genetic classification within Borneo can be found in SUBSECTION 2.2.1.

⁹ There are 15 surviving languages in Taiwan and around 42 or 43 dialects (Elizabeth Zeitoun, p.c.).

¹⁰ Subgrouping within Western-Malayo-Polynesian has been more problematic. The following groups are among those more widely accepted: a Philippine group (which includes most of the languages of the Philippines, except the Sama-Bajau languages); a North Sarawak Group (which includes Kelabit and the languages of Northern Sarawak); a Barito Group (which includes the languages of Southeast Kalimantan and Malagasy of Madagascar); a Malayo-Chamic group (which includes the Malayic languages spoken in island South East Asia, as well as the Chamic languages of mainland SEA) and a Celebic Group (which includes a number of the languages of Sulawesi) (see Blust 2013, SUBSECTION 2.2.1).

¹¹ Note that symmetrical voice systems are not always assumed for all Western Austronesian languages, as discussed in more detail in SUBSECTION 1.4. Formosan languages, in particular, are traditionally analysed as asymmetrical (see Starosta 2009a), though symmetrical analyses have been proposed more recently in Chang (2006) and Kuo (2015). Moreover, Naess (2014) discusses a potentially symmetrical voice system in the Oceanic language Äiwoo. Nonetheless, broadly speaking, Western Austronesian languages have complex systems of verbal morphology, whereas Oceanic languages generally do not.

1.3 Western Austronesian Voice

Western Austronesian voice systems are described as ‘symmetrical’ because they seem to involve two or more voices that are morphologically and syntactically alike (Himmelman 2005a). In other words, each voice is equally morphologically marked and each voice is equally transitive. This can best be understood by comparing symmetrical voice alternations with asymmetrical alternations such as the active/passive alternation and the ergative/antipassive alternation. In these alternations, the active/ergative voice is typically analysed as basic whilst the passive and antipassive are viewed as derived. This analysis follows from the cross-linguistic tendency for passives and antipassives to be morphologically marked in contrast to active/ergative variants (Siewierska 1984: 30, Keenan 1985: 250-251, Keenan & Dryer 2006).¹² Moreover, passives and antipassives are typically marked in terms of their distribution, frequency and productivity (Comrie 1988) and both passivisation and antipassivisation can be seen as detransitivising processes.

To illustrate, let us consider the active/passive alternation in the Mon-Khmer language, Sre, shown in (3). It is morphologically ‘asymmetrical’ since the passive involves additional morphological marking compared with the active. Furthermore, it is syntactically ‘asymmetrical’ as the passive involves detransitivisation.

(3) *Sre* (Mon-Khmer)

a. **Active**

Cal pa? mpon.
wind open door
‘The wind opened the door.’

b. **Passive**

Mpon gə-pa? mə cal.
door PASS-open by wind
‘The door was opened by the wind.’ (Manley 1972)

¹² See Cobbinah & Lüpke (2012) for discussion of passives without morphology.

The active voice in (3a) is morphologically unmarked for voice, and syntactically transitive, with two core arguments expressed as nouns. In the passive voice in (3b), however, the predicate is marked with the prefix *gə-*. Moreover, the clause is intransitive and the agent-like argument expressed as an oblique through a prepositional *by*-phrase. Hence, the passive appears detransitivised.

A similar contrast is seen in ergative/antipassive alternations, such as that of West Greenlandic, shown in (4). Again, the alternation is morphologically ‘asymmetrical’ as the antipassive involves additional morphological marking. Similarly, the alternation is syntactically ‘asymmetrical’ as the antipassive is detransitivised. However, in contrast with the passive, it is not the agent-like argument that is demoted, but rather the absolutive patient-like argument, as shown in (4b) (cf. Polinsky 2013):

- (4) *West Greenlandic* (Eskimo-Aleut)
- a. **Ergative**
- | | | |
|---------------------------|----------|-----------------|
| Arna-p | niqi | niri-vaa. |
| woman-ERG | meat.ABS | eat-IND.3SG.3SG |
| ‘The woman ate the meat.’ | | |
- b. **Antipassive**
- | | | |
|-----------------------|----------|----------------------------|
| Arnaq | niqi-mik | niri-nnig-puq. |
| woman.ABS | meat-INS | eat-ANTIP-IND.3SG |
| ‘The woman ate meat.’ | | (Keenan & Dryer 2006: 359) |

The ergative verb form in (4a) is unmarked for voice. It is transitive and has two core arguments: an ergative and an absolutive. These both trigger pronominal marking or agreement on the verb. The antipassive in (4b) is signalled through the addition of the *-nnig* suffix. There is evidence of detransitivisation as the absolutive argument *niqi* is expressed in the oblique instrumental case. Moreover, the verb in the antipassive

construction only agrees with the absolutive argument. Hence, the ergative/antipassive alternation is also morphologically and syntactically asymmetrical.

However, in many Austronesian languages similar constructions appear to be ‘symmetrical’ (cf. Himmelmann 2005a). That is, neither construction is morphologically or syntactically more basic than the other.¹³ This can be seen in Indonesian in (5), which has two ‘voices’: one in which the agent-like argument (henceforth *actor*) is mapped to subject and one in which the patient-like argument (henceforth *undergoer*) is mapped to subject. These are referred to as actor voice (AV) and undergoer voice (UV) respectively:

- (5) *Indonesian*
- a. **Actor Voice (AV)**
 Hasan mem-beli ikan.
 Hasan AV-buy fish
 ‘Hasan bought fish.’
- b. **Undergoer Voice (UV)**
 Ikan di-beli Hasan.
 fish UV-buy Hasan
 ‘The fish was bought by Hasan.’¹⁴

(adapted from Musgrave 2002: 37)

In (5), both actor voice (AV) and undergoer voice (UV) are morphologically and syntactically equivalent. They are both overtly marked (with the *meN-* and *di-* prefixes respectively) and are both transitive, taking two nominal arguments, *ikan* ‘fish’ and *Hasan*. These are core in both voices and are expressed without oblique case-marking or prepositional phrases, unlike the passive and antipassive illustrated above. For this

¹³ This oversimplifies the situation somewhat in order to illustrate morphosyntactic differences between symmetrical and asymmetrical alternations. Further details on Austronesian voice systems, and a more precise definition of ‘basic’ status are given in CHAPTER 3.

¹⁴ It should be noted that there are a number of distinct constructions in Indonesian that map the undergoer to subject that differ in their syntactic properties. These are further discussed in SUBSECTION 1.3.1.

reason, many refer to the alternations illustrated in (5) as morphologically and syntactically symmetrical (Himmelman 2005a).

A largely similar situation can be seen in languages like Tagalog:

- (6) *Tagalog*
- a. **Actor Voice (AV)**
 B<um>ili ang lalake ng isda sa tindahan.
 <AV>buy SUBJ man CORE fish OBL store
 ‘The man bought fish at the store.’
- b. **Undergoer Voice (UV)**
 B<in>ili-Ø ng lalake ang isda sa tindahan.
 <PFV>buy-UV CORE man SUBJ fish OBL store
 ‘The man bought the fish at the store.’
- c. **Locative Voice (LV)**
 B<in>ilih-an ng lalake ng isda ang tindahan.
 <PFV>buy-LV CORE man CORE fish SUBJ store
 ‘The man bought fish at the store.’
- d. **Instrumental Voice (IV)**
 Ip<in>am-bili ng lalake ng isda ang pera.
 <PFV>IV-buy CORE man CORE fish SUBJ money
 ‘The man bought fish with the money.’
- e. **Benefactive Voice (BV)**
 I-b<in>ili ng lalake ng isda ang bata.
 BV<PFV>buy CORE man CORE fish SUBJ child
 ‘The man bought fish for the child.’ (Arka 2002)

The examples in (6) demonstrate an alternation in the mapping of semantic roles to grammatical functions similar to the Indonesian alternation in (5). The verb forms are all equally marked – as summarised in TABLE 1.1 below – and each construction seems to be transitive as they all involve a subject function with *ang* marking and other core nominals, marked with *ng*. Thus, much in the same way as alternations like (5) can be described as morphologically and syntactically symmetrical, so too can the Tagalog alternation, shown in (6).

Table 1.1 Tagalog Voice Morphology (Himmelman 2002)

	Realis	Irrealis
Actor Voice	<i>-um-/N-</i>	<i>-um-/M-</i>
Undergoer Voice	<i>-in-</i>	<i>-in</i>
Locative Voice	<i>-in- -an</i>	<i>-an</i>
Benefactive Voice	<i>i- -in-</i>	<i>i-</i>

Consequently, a wide range of Western Austronesian languages can be seen to share the property of having symmetrical voice alternations. Let us now explore the differences between the Indonesian alternations in (5) and the Tagalog alternations in (6) that have motivated a two-way typology of Western Austronesian into ‘Philippine-type’ languages and ‘Indonesian-type’ languages.

1.3.1 Philippine-type vs Indonesian-type

Thus far, I have focused on the properties that are shared by the voice systems of Indonesian and Tagalog. There are also a number of differences, which are discussed in more detail in CHAPTER 3. For now, the most notable difference is the number of alternations. In addition, Tagalog has the well-documented property of being ‘patient prominent’ (cf. Foley & Van Valin 1984). This means that there is a preference for UV wherever the undergoer is definite and a restriction against definite undergoers in any other voice (SUBSECTION 1.4.2.1.2). These differences have prompted many people to classify alternations such as (5) as ‘Indonesian-type’ and alternations such as (6) as ‘Philippine-type’. Yet, although the terms are prevalent in the literature (cf. Himmelman 2005a, Arka 2002), it is not always clear what the classifications would mean beyond a distinction between a multi-voice system on the one hand, and a two-voice system on the other. Neither is it clear how to establish whether a particular

voice system should be considered ‘Indonesian-type’ or ‘Philippine-type’ (cf. Brickell 2014).

Most attempts at making the typology more explicit draw upon a list of structural properties that seem to cluster around symmetrical voice languages in the Philippines and the symmetrical voice languages in Indonesia (Himmelmann 2005a, Arka 2002, Ross & Arka 2005). One such example is Arka (2002) who suggests the following defining characteristics:¹⁵

Table 1.2 Defining Characteristics of Philippine-type and Indonesian-type (Arka 2002)

	Indonesian Type	Philippine Type
Symmetrical alternations	Y	Y
True passive	Y	N
Applicative suffixes	Y	N
Micro roles with voices	N	Y
Mood marking morphology	N	Y
Case marking	N	Y

In Arka’s (2002) typology, both ‘Indonesian-type’ and ‘Philippine-type’ languages share the property of symmetrical alternations, as discussed in SUBSECTION 1.3. However, they differ in the five remaining properties in TABLE 1.2. Firstly, Indonesian-type languages, in addition to symmetrical alternations, also have a construction resembling the passive. This can be seen in (7):

- (7) *Indonesian*
 a. **Actor Voice (AV)**
 Hasan mem-beli ikan.
 Hasan AV-buy fish
 ‘Hasan bought fish.’

¹⁵ Arka (2002) uses the terms ‘Indonesian-type’ and ‘Tagalog-type’

- b. **Undergoer Voice (UV)**
 Ikan di-beli Hasan.
 fish UV-buy Hasan
 ‘The fish was bought by Hasan.’

- c. **Passive (PASS)**
 Ikan di-beli oleh Hasan.
 fish PASS-buy by Hasan
 ‘The fish was bought by Hasan.’

(adapted from Musgrave 2002: 37)

Although both UV and the passive use the same morphological marking, namely the *di-* prefix, (7c) differs from (7b) in that it is syntactically intransitive. The actor argument, *Hasan*, is optional and not expressed as a core argument but rather as an oblique in the prepositional phrase headed by the preposition *oleh*.¹⁶ As a result, (7c) resembles the passive construction illustrated in (3) for Sre. In other languages, such as Sasak and Balinese, UV and passives have different morphological marking (Austin, p.c.). Indeed, in Balinese the UV construction is morphologically unmarked, as illustrated in SUBSECTION 1.4.2.2.2.

In fact, Indonesian has four different constructions in which the undergoer is mapped to subject (see Riesberg 2014). In addition to (7b) and (7c), where the actor is a proper noun, distinct constructions are used when the actor is a third person pronoun or a first/second person pronoun:

- (8) *Indonesian*
 a. ***di-V-nya* UV**
 Ikan di-beli=nya.
 Fish UV-buy=3SG
 ‘The fish was bought by him.’

¹⁶ See Donohue (2007b) for discussion of syntactic differences between constructions like (7b) and (7c). For example, adverbials can intervene between the verb and the PP actor but not the verb and the NP actor.

- b. *pro*=V UV
 Ikan saya=beli.
 Fish 1SG=buy
 ‘The fish was bought by me.’

(adapted from Musgrave 2002: 38)

The four constructions are said to differ in their degree of transitivity, with the two constructions in (8) typically considered the most transitive (see SUBSECTION 3.3).¹⁷ Hence, some authors restrict the term UV to the *di*-V=*nya* and *pro*=V constructions in Indonesian, and refer to both constructions in (7b) and (7c) as passives (see Riesberg 2014, Arka & Manning 2008). Others also include the *di*-V NP construction in (7b) as undergoer voice, and restrict the term passive to cases in which the actor is also formally oblique, i.e. *di*-V PP constructions like (7c) (see Himmelmann 2005a, Donohue 2007b). Finally, Kroeger (2014) argues that only *pro*=V constructions are UV, and that all *di*-V constructions are passive (see SUBSECTION 3.4.2). Hence, the nature of morphological and syntactic symmetry is somewhat more complicated than described in SUBSECTION 1.3, as discussed in more detail in CHAPTER 3.

The second property that characterises Indonesian-type languages is the use of applicative constructions (cf. Himmelmann 2005a, Brickell 2014).¹⁸ Applicatives take oblique arguments and promote them to direct object status. The suffix *-kan* in Indonesian, for example, marks a construction in which a benefactive argument, like *Dogol*, is promoted to direct object:

- (9) *Indonesian*
 a. **Actor Voice (AV)**
 Hasan membeli ikan untuk Dogol.
 Hasan AV.buy fish for Dogol
 ‘Hasan bought fish for Dogol.’

¹⁷ Arka (2005:48) devises a quantitative method of defining core argument status, discussed in SUBSECTION 1.4.2.2.3. The actor argument is most core in *di*-V=*nya* constructions, followed by *di*-V NP and finally *di*-V PP.

¹⁸ Note that these applicative suffixes are typically multifunctional (see Hemmings 2013)

- b. **AV Applicative**
 Hasan membeli-kan Dogol ikan.
 Hasan AV.buy-APPL Dogol ikan
 ‘Hasan bought Dogol some fish.’

- c. **UV Applicative**
 Dogol di-beli-kan Hasan ikan.
 Dogol UV-buy-APPL Hasan fish
 ‘Dogol was bought some fish by Hasan.’

(adapted from Shiohara 2012)

The applicative *-kan* takes the peripheral benefactive, *Dogol*, and promotes it to a core argument. This can be seen by the fact that *Dogol* appears as part of the prepositional phrase *untuk Dogol* ‘for Dogol’ in the non-applicativised version in (9a), but is realised as an NP in the core-argument position directly following the verb in (9b). The applicative can apply both in AV, as in (9b), and UV, as in (9c), in which the benefactive argument is then mapped to subject. Indeed, the use of applicatives is the only way in Indonesian-type languages to map so-called micro roles, like the benefactive, to subject, since the voice system involves only two, more generalised voices for the actor and the undergoer (see Arka 2002).

In contrast, Philippine-type languages do not have applicatives but instead have voice marking for more specific semantic roles.¹⁹ This can be seen in (6c), (6d) and (6e), repeated as (10) below:

- (10) *Micro-role Voices in Tagalog*
 a. **Locative Voice (LV)**
 B<in>ilih-an ng lalake ng isda ang tindahan.
 <PFV>buy-LV CORE man CORE fish SUBJ store
 ‘The man bought fish at the store.’

¹⁹ Aldridge (2004) treats such voices as applicative constructions (see SUBSECTION 1.4.2.1.2). However, unlike applicatives in Indonesian, they cannot attach to AV stems.

b. **Instrumental Voice (IV)**

Ip<in>am-bili	ng lalake	ng isda	ang pera.
<PFV>IV-buy	CORE man	CORE fish	SUBJ money

‘The man bought fish with the money.’

c. **Benefactive Voice (BV)**

I-b<in>ili	ng lalake	ng isda	ang bata.
BV<PFV>buy	CORE man	CORE fish	SUBJ child

‘The man bought fish for the child.’ (Arka 2002)

In each construction, a micro-role is mapped to subject, as shown through the *ang*-marking. This identifies the locative as subject in LV in (10a), the instrument as subject in IV in (10b), and the benefactive as subject in BV in (10c). As a result, Philippine-type voice systems typically involve a higher number of voice alternations than Indonesian-type voice systems.²⁰

Moreover, Philippine-type languages also differ from Indonesian-type languages in that they have portmanteau voice and mood-marking morphology. In other words, the voice markers for Tagalog summarised in TABLE 1.1 not only express voice, but also realis and irrealis mood (Himmelman 2002). Finally, Philippine-type languages are said to have case-marking of dependent nominal arguments. The case-marking distinction is seen in (6) in that the argument mapped to subject function takes prenominal *ang*-marking. Core arguments that are not mapped to subject function are marked with *ng* and other semantic arguments with *sa*. Thus, Arka (2002) defines Philippine-type and Indonesian-type in terms of a set of shared syntactic properties that are central to the respective voice systems.

Further typological characteristics that are not discussed in Arka (2002), but are sometimes used to distinguish Philippine-type and Indonesian-type languages,

²⁰ For example, Prentice (1971: 32) posits five voice alternations for Timugon; Walton (1983: 9) posits five for Sama; Hurlbut (1988) posits seven in Eastern Kadazan and Kroeger (1988) posits six in Kimaragang. See Boutin (1988) for a discussion of the problems of identifying voice constructions.

include clitic systems and word order (cf. Himmelmann 2005a, Donohue 2007a, Billings & Kaufman 2004). Philippine-type languages are said to have second-position enclitics, whilst Indonesian-type languages tend to have verb-adjacent proclitics (see CHAPTER 4). Similarly, Philippine-type languages are typically verb-initial, whereas Indonesian-type languages are typically SVO, as can be seen in (5) and (6) (see CHAPTER 5). Consequently, one could compare Western Austronesian languages not just in terms of voice, and the properties in TABLE 1.2, but in the interacting categories of clitic pronouns and word order (FIGURE 1.2):

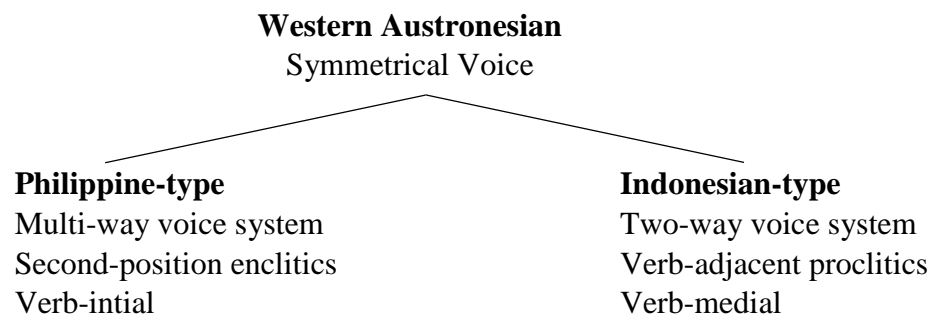


Figure 1.2 Philippine-type vs Indonesian-type

1.3.2 Asymmetrical Austronesian Voice Systems

In contrast to the ‘Philippine-type’ and ‘Indonesian-type’ languages described above, there are a number of Austronesian languages, particularly in the Central and East-Central Malayo-Polynesian branch, which have asymmetrical voice alternations, like the passive alternation discussed in (3). For example, in Bima, spoken in the Eastern part of Sumbawa, the construction that maps the undergoer to subject appears to be a proto-typical passive, and is indicated through the addition of the marker *di-* for irrealis mood and *ra-* for realis mood:²¹

²¹ Nb. As discussed in TABLE 1.2 and CHAPTER 4, languages in Indonesia are not typically analysed as having case-marking in pronouns. This also applies to Bima.

- (11) *Bima*
- a. **Active**
 Iwa nahu sepe-na buku ede.
 friend 1SG borrow-3SG book DEM
 ‘My friend borrowed that book.’
- b. **Passive**
 Buku ede ra-sepe ba iwa nahu.
 book DEM PASS.REAL-borrow by friend 1SG
 ‘That book was borrowed by my friend.’ (Arka 2009: 255)

Much like (3), in (11b) the passive morphology is accompanied by the demotion of the agent to the post-verbal position and oblique status, as can be seen by its realisation as a PP *ba iwa nahu* ‘by my friend’. Thus, (11b) is lower in transitivity than the UV constructions in (5) and (6), and (11) constitutes an asymmetrical alternation.

Further eastwards, there are languages with no morphological voice alternations at all. For example, consider the Eastern Flores language, Sikka:

- (12) *Sikka*
- a. **Actor Voice?**
 Petrus piru Siti.
 Petrus kiss Siti
 ‘Petrus kisses Siti.’
- b. **Undergoer Voice?**
 Petrus Siti piru.
 Petrus Siti kiss
 ‘Petrus kisses Siti.’ (Shibatani 2009)

Two different ways of expressing the proposition, ‘Petrus kisses Siti’, are shown in Sikka in (12). In both (12a) and (12b), the verb form *piru* ‘kiss’ is unchanged. However, in (12a) the undergoer follows the verb, whilst in (12b) it precedes the verb. It has been suggested that (12b) is a UV construction in which the undergoer is mapped to subject and the actor remains a core argument (see Arka & Wouk 2014, Sedeng

2000).²² Unlike the Indonesian and Tagalog constructions in (5b) and (6b), the contrast is expressed syntactically through a change in word order, rather than an alternation in the form of the verb. However, many of the debates surrounding Philippine-type and Indonesian-type languages discussed in SUBSECTION 1.4 also apply to languages without overt morphological distinctions between voice alternations (Nagaya 2009b). Hence, ‘voice’ could well be fundamental to the structure of Austronesian as a whole.

1.3.3 Summary

In this section, I have defined Western Austronesian voice alternations as ‘symmetrical’ in the sense that the voices are morphologically and syntactically equivalent. This contrasts with active/passive and ergative/antipassive alternations, in which derived voices involve additional morphological marking and detransitivisation or demotion of a core argument. I then introduced an important distinction between ‘Philippine-type’ and ‘Indonesian-type’ languages. Both share the property of symmetrical voice, but differ in the nature of their voice systems, clitic systems and word-order typology, as will be discussed in more detail in CHAPTERS 3, 4 and 5. Finally, I discussed a selection of Austronesian languages with asymmetrical and/or morphologically unmarked alternations and argued that many of the key debates within Western Austronesian or symmetrical voice languages apply equally to this group, though they are not further discussed in this thesis. I now turn to explore some of the key debates in Austronesian syntax, bearing in mind that any account would have to explain both the symmetrical nature of the voice alternations and the ‘patient prominence’ of Philippine-type systems.

²² Though see Nagaya (2009b) for discussion of the similarities between a similar construction in Lamaholot and a topicalisation construction.

1.4 Key Debates within Austronesian Syntax

In the previous section, I defined the concept of symmetrical voice alternations. This has led to two major debates within Western Austronesian syntax, namely the subject debate (SUBSECTION 1.4.1) and the alignment debate (SUBSECTION 1.4.2).

1.4.1 The Subject Debate

The first major debate in Western Austronesian is whether or not these languages can be said to have a grammatical subject. Subjects are typically defined as having a set of morphological and behavioural properties (cf. Keenan 1976).²³ However, typical subject properties are split between two arguments in Western Austronesian languages: the actor (i.e. the highest thematic role) and the argument selected as prominent by the verbal morphology (i.e. the actor in AV, the undergoer in UV and so on). This led Schachter (1976) to propose that ‘subject’ was not a relevant notion, and that the prominent argument was better described as ‘topic’. However, since the definition of ‘topic’ in the Austronesian literature is not equivalent to information-structure topics (cf. Lambrecht 1994), the matter has remained controversial (cf. Guilfoyle et al. 1992, Kroeger 1993, Schachter 1995, Liao 2004, Cole & Hermon 2005, Shibatani 2008, Nagaya 2009b, 2010, Pearson 2005). In the following sections, I illustrate the split subject properties of Western Austronesian through the examples of Tagalog and Indonesian, before discussing possible analyses.

²³ Though see Himmelmann (2005a) for alternative methods of identifying ‘subject’ and Weber (2011) on the cross-linguistic applicability of such tests.

1.4.1.1 Subject in Tagalog

Much has been written on the question of subjects in Tagalog and other Philippine-type languages (Schachter 1976, Kroeger 1993). This is because Tagalog appears to split Keenan's (1976) subject properties between the *ang*-marked NP (whose semantic role differs depending on the voice construction) and the actor (which remains constant). The split can be seen if we consider the patterns of relativisation and reflexivisation.

Keenan & Comrie (1979) propose an accessibility hierarchy, which states that if only one clausal argument can be relativised on, then this argument will be the grammatical subject. Hence, the ability to be relativised on can be considered a characteristic of subjects. In Tagalog, only the *ang*-marked argument can be relativised on, as illustrated in (13):²⁴

(13) *Tagalog Relative Clauses*

a. **Actor Voice**

Matalino ang lalaki[=ng bumasa ng diyaryo].
 Intelligent NOM man=LNK AV.read GEN newspaper
 'The man who read a newspaper is intelligent.'

b. *Interesante ng diyaryo[=ng bumasa ang lalaki].
 Interesting GEN newspaper=LNK AV.read NOM man
 For: 'The newspaper that the man read is interesting.'

c. **Undergoer Voice**

Interesante ang diyaryo[=ng binasa ng lalaki].
 Interesting NOM newspaper=LNK UV.read GEN man
 'The newspaper that the man read is interesting.'

²⁴ The same patterns apply to *wh*-questions, and what is sometimes called 'extraction' more generally, in that only the *ang*-marked argument can correspond to a *wh*-word in initial position:

(i) Sino ang b<in>igy-an ng lalaki ng bulaklak?
 who NOM <PFV>give-LV GEN man GEN man
 'Who did the man give flowers to?'

(ii) *Sino ang i-b<in>igay ng lalaki ang bulaklak?
 *Sino ang nagbigay ang lalaki ng bulaklak?

(Rackowski & Richards 2005: 566)

- d. *Matalino ng lalaki[=ng binasa ang diyaryo].
 Intelligent GEN man=LNK UV.read NOM newspaper
 For: ‘The man who read the newspaper is intelligent.’
 (Schachter 1976: 500)

In actor voice in (13a) and (13b), only the *ang*-marked actor can be relativised. In undergoer voice in (13c) and (13d), only the *ang*-marked undergoer can be relativised. This would suggest that the *ang*-marked argument is subject.

However, another common test of subjecthood is control of reflexive binding. As Schachter (1976) discusses, cross-linguistically subjects tend to control reflexive binding. In Tagalog, it is the actor that binds reflexives, regardless of whether it is the *ang*-marked element in the clause or not:

- (14) *Tagalog Reflexive Binding*
- a. **Actor Voice** (actor = *ang*-marked)
 Nag-aalala ang lolo sa kaniyang sarili.
 AV-worry NOM grandfather DAT his self
 ‘Grandfather worries about himself.’
- b. **Undergoer Voice** (actor ≠ *ang*-marked)
 Inaalala ng lolo ang kaniyang sarili.
 UV.worry GEN grandfather NOM his self
 ‘Grandfather worries about himself.’ (Manning 1996: 13)

In both AV in (14a) and UV in (14b), the actor controls reflexive binding. Hence, reflexivisation would seem to suggest that the actor is subject, not the *ang*-marked NP.

The tests which Schachter (1976) used to identify ‘subject’ in Tagalog – and the argument selected by these tests – can be summarised as follows:

Table 1.3 Tagalog Subject Tests (Schachter 1976)

<i>Ang</i> -marked NP	Agent-like argument
Obligatory argument	Antecedent for reflexives
Floating Quantifiers	Control
Relativisation	Imperative addressee
Agreement marking	

Given the split, Schachter (1976) concluded that ‘subject’ was not a category applicable to the languages of the Philippines and that ‘reference-related’ subject properties were associated with the *ang*-marked argument, which he analyses as ‘topic’, whilst ‘role-related’ properties were associated with the actor (Schachter 1976: 514).

1.4.1.2 Subject in Indonesian

A largely similar situation is found in Indonesian. Like Tagalog, subject properties are split between the actor and the argument selected as prominent by the verbal morphology. For Indonesian, since there is no case-marking, this is typically the pre-verbal argument. Once again, we can see the split if we compare the relativisation patterns in (15) and reflexivisation patterns in (16):²⁵

- (15) *Indonesian Relative Clauses*
- a. **Actor Voice**
 Hasan [yang mem-beli ikan].
 Hasan REL AV-buy fish
 ‘It was Hasan who bought fish.’
- b. *Ikan [yang mem-beli Hasan].
 fish REL AV-buy Hasan
 For: ‘It was fish that Hasan bought.’

²⁵ Similar patterns apply for other Indonesian-type languages, such as Balinese (see Arka 2003), though Balinese is not morphologically symmetrical (SUBSECTION 3.4.2.1).

c. **Undergoer Voice**

Ikan [yang di-beli Hasan].
 fish REL UV-buy Hasan
 ‘It was fish that Hasan bought.’

d. *Hasan [yang di-beli ikan].

Hasan REL UV-buy fish
 For: ‘it was Hasan who bought fish.’

(adapted from Musgrave 2002: 59)

(16) *Indonesian Reflexives*

a. **Actor Voice**

[saya] menyerah-kan [diri saya] ke polisi.
 1SG AV.surrender-APPL self 1SG to police
 ‘I surrendered myself to the police.’

b. **Undergoer Voice (pro=V)**

[diri saya] [saya] serah-kan ke polisi.
 self 1SG 1SG UV.surrender-APPL to police
 ‘I surrendered myself to the police.’

c. **Undergoer Voice (di-V-nya)**

[diri-nya] di-serah-kan=[nya] ke polisi.
 self-3SG UV-surrender-APPL=3SG to police
 ‘He/she surrendered himself to the police.’

(Arka & Manning 1998)

The relativisation data in (15) follow exactly the same patterns as Tagalog. Only the argument indicated in the verbal morphology can be relativised on. This would suggest that the actor is subject in AV and the undergoer is subject in UV. However, like Tagalog, the reflexivisation patterns in (16) suggest that the actor controls reflexive binding, regardless of whether it is selected by the voice-marking as in (16a), or not, as in (16b) and (16c).²⁶

²⁶ Interestingly, this is not possible for passive *di-* clauses, as in (i), or UV clauses where the agent is a full NP or proper noun, as in (ii):

(i) Passive *di-* clauses (*di-V* PP)
 ?*Diri-nya di-serah-kan ke Polisi oleh Amir.
 self-3SG PASS-surrender-APPL to police by Amir
 For: ‘Amir surrendered himself to the police.’

Following Riesberg (2014), different subject tests and their results for Indonesian can be summarised as follows:

Table 1.4 Indonesian Subject Tests (Riesberg 2014)

Pre-verbal argument	Actor argument
Relativisation	Reflexivisation
Control	
Raising	

Hence, split subject properties are a feature of both Philippine and Indonesian-type languages.

1.4.1.3 Previous Accounts

Whilst Schachter (1976, 1995) argued that split subject properties were sufficient reason to abandon the notion of ‘subject’, a number of accounts have since been proposed to maintain the idea of grammatical functions and still account for split subject properties in Western Austronesian. One such account is Manning’s (1996) inverse mapping theory.²⁷ He argues that only Schachter’s (1976) ‘reference-related’

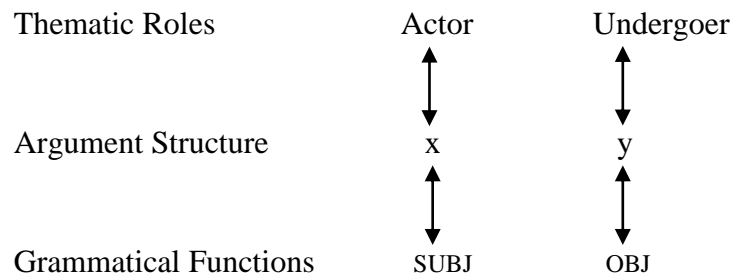
-
- (ii) UV *di*- clauses (*di*-V NP)
 ?*Diri-nya tidak di-per-hati-kan Amir.
 self-3SG NEG UV-TR-heart-APPL Amir
 For: ‘Amir did not take care of himself.’ (Arka & Manning 1998)

This is argued to support an analysis whereby passive *di*- and UV *di*- are not two separate forms, but rather Indonesian is in a state of transition from ergative to accusative (see SUBSECTION 1.4.2 and CHAPTER 3 for further discussion). This transition may have occurred at different rates for pronominal and nominal arguments. See also Kroeger (2014) who argues that the binding properties may relate to pragmatic/discourse rather than syntactic status. He argues that binding is possible when the actor is pronominal and inherently topical, as in (16b) and (16c), but not when the actor is a full NP or proper noun, as in (i) and (ii).

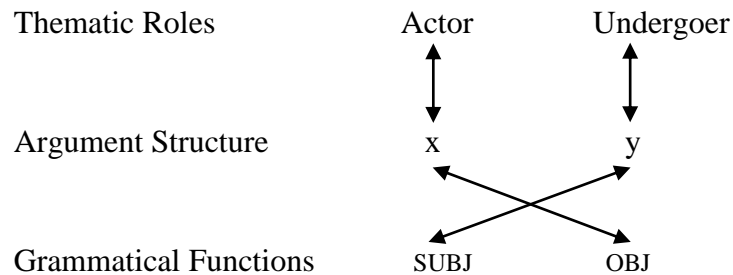
²⁷ See Falk (2006) for an account in Lexical Functional Grammar (LFG) that splits the notion of subject into two categories: i) the most prominent function of a predicate (GF) and ii) the ‘sentence-topic’ or ‘pivot’ that controls cross-clausal continuity (PIV). He argues that role-related properties are characteristic of the highest grammatical function, whilst reference-related properties are characteristic of the pivot. Furthermore, he argues that in Philippine-type languages, the voice-marking morphology specifies which grammatical function is associated with the pivot function, rather than (i) and (ii) being associated by default, as in syntactically accusative languages. Consequently, both Falk (2006) and Manning (1996) redefine ‘subject’.

properties (i.e. those that relate to the *ang*-marked NP) are important in the identification of subjects. The other properties can be handled at argument structure and relate to the highest thematic role. In syntactically accusative languages, the highest thematic role and the highest grammatical function will normally equate. However, in syntactically ergative and Philippine-type languages an inverse mapping is possible. This is illustrated in (17) (cf. Falk 2006):

(17) a. **Syntactically Accusative - Default Mapping**

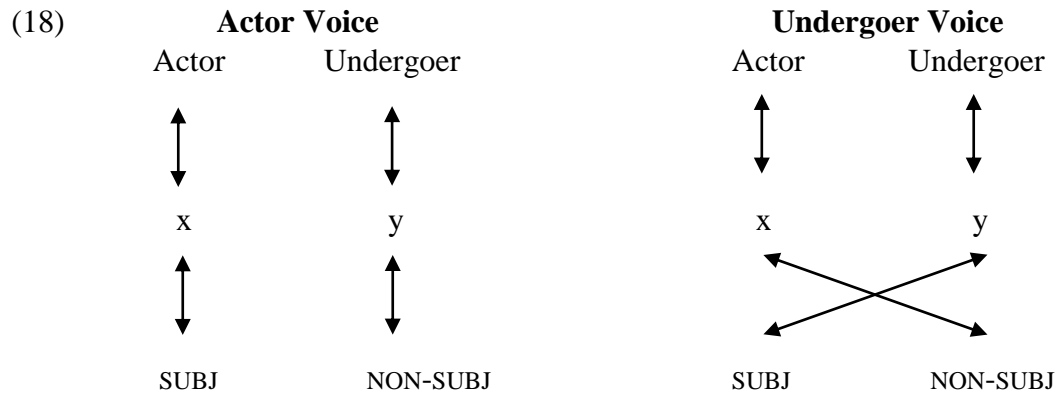


b. **Syntactically Ergative – Inverse Mapping**



Hence, the split subject properties follow from the fact that actor and subject do not always align.

The inverse-mapping approach is extended to Western Austronesian languages in different guises by Kroeger (1993), Arka & Manning (1998) and Riesberg (2014), among others. Essentially, the accounts allow both the mapping in (17a) and the mapping in (17b), depending on voice morphology. This is illustrated in (18):



Consequently, although split subject properties have been controversial in the Austronesian literature, it is possible to provide a theoretical account of the voice alternations, whilst maintaining a notion of grammatical subject. For this reason, I follow Riesberg (2014) and Kroeger (1993) in referring to the argument selected by the verbal morphology as subject in this thesis (see SUBSECTION 2.5.1 for arguments relating to Kelabit). In other words, the actor is treated as subject in AV and the undergoer as subject in UV. The status of the non-subject core argument is sometimes less clear-cut, particularly given the controversy of mapping an actor role to object function in UV (see Riesberg 2014).²⁸ Hence, I adopt the more neutral terminology of non-subject argument. This can be read as equivalent to the terms pivot and non-pivot used in Arka (2002 etc.).

In summary, typical subject properties identified in syntactically accusative languages appear to be split in both Philippine-type and Indonesian-type languages. Some properties are associated with the actor, regardless of the voice construction and other properties are associated with whichever argument is highlighted by the voice morphology.

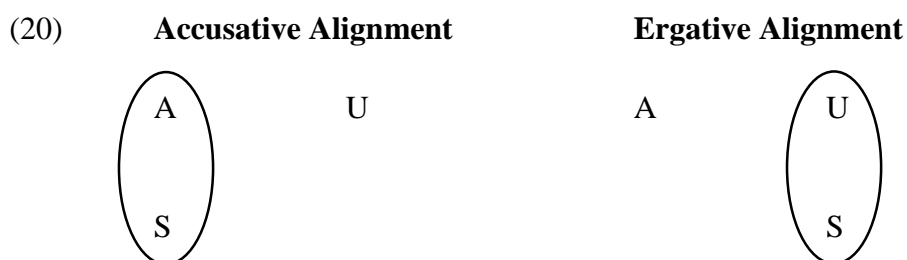
²⁸ Note that it is not unheard of for an actor to be treated as an object or internal argument. This is proposed for Norwegian existential clauses (see Lødrup 2000) and for inverse constructions in the Mapudungan language of Chile (Arnold 1997)

1.4.2. The Alignment Debate

The second key debate within Western Austronesian linguistics is the question of alignment. In particular, the debate concerns whether Western Austronesian languages can be considered to have accusative alignment, ergative alignment or an alignment system that is unique to Austronesian:²⁹

- (19) **The Western Austronesian Alignment Debate**
- a. The Accusative Hypothesis
 - b. The Ergative Hypothesis
 - c. The Philippine-type/Symmetrical Alignment Hypothesis

The two most prominent alignment systems found cross-linguistically are accusative alignment and ergative alignment. These can be schematised in (20):



In an accusative system, the actor (A) argument of a transitive clause is treated in the same way as the single (S) argument of an intransitive clause, and the undergoer (U) is treated differently, both in terms of morphological marking, and in syntactic behaviour. In an ergative system, U is treated the same as S, and A is treated differently. The difference can be illustrated using the example of Latin in (21), which has accusative alignment, and Dyirbal in (22), which has ergative alignment:

²⁹ Another alternative is to analyse Western Austronesian languages as ‘active’ (see Drossard 1984 on Tagalog). This possibility is not further explored in the thesis.

- (21) *Latin*
- a. **Intransitive**
 Domin-us veni-t.
 master-NOM (S) come-3SG.PRS
 ‘The master comes.’
- b. **Transitive**
 Domin-us serv-um audi-t.
 master-NOM (A) slave-ACC (U) hear-3SG.PRS
 ‘The master hears the slave.’ (Dixon 1994: 9)
- (22) *Dyirbal*
- a. **Intransitive**
 ŋuma-Ø banaga-nyu.
 father-ABS (S) return-NON.FUT
 ‘Father returned.’
- b. **Transitive**
 ŋuma-Ø yabu-ŋgu bura-n.
 father-ABS (U) mother-ERG (A) see-NON.FUT
 ‘Mother saw father.’ (Dixon 1994: 10)

In Latin, the A argument of a transitive clause and the S argument of an intransitive clause both receive nominative case, whilst the U argument of a transitive clause is treated differently and receives accusative case. In Dyirbal, however, it is the A argument that is treated differently, receiving ergative case, whilst both the S and the U argument receive absolutive case. These differences also extend beyond morphology to the level of syntax, where core arguments function together in constraints on clause-combining (Dixon 1994).

In order to establish which system of alignment obtains for a particular language, it is necessary to compare transitive and intransitive clauses. However, as illustrated in SUBSECTION 1.3, Western Austronesian languages seem to have two or more types of transitive clause. If AV is compared with intransitive clauses, the alignment appears to be accusative. If UV is compared with intransitive clause, the

alignment appears to be ergative. For this reason, alignment in Western Austronesian languages is sometimes referred to as ‘Philippine-type’ in contrast to the accusative and ergative alignment systems described above (cf. Tallerman 2005). Ultimately, the debate rests on the extent to which the alternations are considered symmetrical, and whether there is any evidence for considering either AV or UV the basic transitive clause-type. This section outlines different synchronic accounts of alignment in Philippine-type and Indonesian-type languages, before introducing a diachronic account, proposed by Aldridge (2011).

1.4.2.1 The Alignment Debate in Philippine-type Languages

1.4.2.1.1 The Accusative Hypothesis

Early analyses of Philippine-type languages, such as F. Blake (1925) and Bloomfield (1917), tended to assume that the languages were nominative/accusative.³⁰ Under an accusative analysis, AV is considered a transitive clause, and all other voices are considered intransitive variations of the passive. Nominal *ang*-marking is assumed to indicate nominative case. This could be represented schematically as follows:

- (23) *The Accusative Analysis in Tagalog*
- a. **Active/Transitive (AV):**
 B<um>ili ang lalake ng isda sa tindahan.
 <ACT>buy NOM man (A) ACC fish (U) in store
 ‘The man bought fish at the store.’
- b. **Intransitive:**
 D<um>ating ang babae.
 <INTR>arrive NOM woman (S)
 ‘The woman arrived.’

³⁰ See also Wolfenden (1961), Llamzon (1968), Johnson (1977) and Bell (1983) for more recent adaptations.

c. **Passive (UV):**

B<in>ili ng lalake ang isda sa tindahan.
<PASS>buy OBL man NOM fish (S) in store
'The fish was bought by the man at the store.'

(adapted from Aldridge 2004: 2)

If *ang*-marking is treated as nominative case, and the UV actor as an adjunct, then Tagalog can be considered to have canonical accusative alignment, where A and S are marked alike (with *ang*) and U is marked differently (with *ng*).

However, there are several reasons why such an analysis is problematic, as discussed extensively in Foley (2008) and Riesberg (2014). Firstly, the UV actor in Tagalog is not demoted and remains a core argument of the clause (see SUBSECTION 1.4.2.1.3). Secondly, the UV clause is not more morphologically marked than AV, as would be expected of a passive. Thirdly, the situation in Tagalog would be typologically unusual in having a single active voice and four passives. Finally, the analysis fails to account for the 'patient-prominence' described in Philippine-type languages. As a result, such analyses have lost favour in recent years.

1.4.2.1.2 The Ergative Hypothesis

A second approach is to treat Philippine-type languages as having ergative alignment (see Gerds 1988, Gault 1999, De Guzman 1988, T. Payne 1982, B. Blake 1988, Ceña 1977, Starosta 2009abc, Aldridge 2004, 2011, 2012).³¹ Under an ergative analysis, UV is analysed as the basic transitive clause and AV as an antipassive.³² All other voices are treated as applicatives and hence derivational rather than inflectional (cf. Reid & Liao 2004: 453, Mithun 1994). For example, Aldridge (2004: 2) interprets the verbal

³¹ Starosta et al (1982), Ross (2009) and Aldridge (2016) argue that ergativity arose through the reanalysis of an earlier system of nominalisations. This is supported by the fact that verbal affixes like *-in-* and *-an* only occur as nominalisations in Tsou, Rukai and Puyuma (Aldridge 2016).

³² Starosta (2009c) argues that AV is an extended intransitive clause, in the sense of Dixon (1994).

affixes in Tagalog – shown in TABLE 1.5 – as marking transitivity rather than symmetrical voice alternations. Similarly, she interprets the nominal markers – shown in TABLE 1.6 – as marking ergative and absolutive case:

Table 1.5 Tagalog Verbal Marking in Aldridge (2004)

Verbal affix	Traditional analysis	Aldridge's analysis
<i>-in (-in-)</i>	Undergoer Voice	Transitive marker
<i>-an</i>	Locative Voice	Applicative marker
<i>i-</i>	Benefactive Voice	Applicative marker
<i>-um-, mag-</i>	Actor Voice	Intransitive marker

Table 1.6 Tagalog Nominal Marking in Aldridge (2004)

Nominal marker	Aldridge's analysis
<i>ang</i>	absolutive case
<i>ng</i>	ergative/oblique case
<i>sa</i>	preposition

This allows her to posit a system of ergative alignment, illustrated in (24):

(24) *The Ergative Analysis in Tagalog*

a. **Transitive (UV):**

B<in>ili ng babae ang isda.
 <TR.PFV>buy ERG woman (A) ABS fish (U)
 'The woman bought the fish.'

b. **Intransitive:**

D<um>ating ang babae.
 <INTR.PFV>arrive ABS woman (S)
 'The woman arrived.'

c. **Antipassive (AV):**

K<um>ain ang babae ng isda.
 <ANTIP.PFV>eat ABS woman (S) OBL fish
 'The woman ate (a) fish.' (Aldridge 2004: 2)

In Aldridge's (2004) account, *ang*-marking is reanalysed as absolutive case, and *ng*-marking as both an ergative and an oblique marker, which is a syncretism commonly

found in ergative languages (Kaufman to appear).³³ The UV infix *-in-* is taken to indicate a transitive clause, whilst the AV infix *-um-* is analysed as marking an antipassive and syntactically intransitive construction (see also Starosta 2009abc). Assuming this, U and S are marked in the same way (with *ang*) and A is marked differently (with *ng*). Hence, there is proto-typical ergative alignment, following the diagram in (20).

The main argument in favour of the ergative hypothesis is that there are semantic similarities between AV constructions and antipassives in other languages (cf. Cooreman 1994, T. Payne 1982). It is well documented that the non-subject undergoer in an AV clause is typically interpreted as indefinite, nonspecific and non-presuppositional (see Bloomfield 1917, Kroeger 1993, Aldridge 2004 and Kaufman 2005 among others).³⁴ In UV clauses, conversely, the undergoer is typically definite, which has prompted many to describe the Philippine languages as ‘patient prominent’ (Foley & Van Valin 1984, see SUBSECTION 3.4.1 for further discussion). This results in a restriction against NPs with definite demonstratives expressing the undergoer of an AV clause, which is not found for undergoers in UV:

(25) *Tagalog Definite Undergoers*

a. **Antipassive (AV)**

*/?K<um>ain nito ang bata.
 <ANTIP>eat OBL.this ABS child

For: ‘The child ate this.’

Possible partitive interpretation: ‘The child ate from this.’

b. **Ergative (UV)**

K<in>ain ng bata ito.
 <TR.PFV>eat ERG child ABS.this

‘The child ate this.’

(Kaufman, to appear)

³³ Note that in other Philippine languages, such as Ivatan, there are separate markers for ergative and oblique case (cf. Reid 1966, Kaufman, to appear).

³⁴ There are situations in which definite undergoers do occur in AV. See Himmelmann (1991) for discussion.

In (25a), the definite demonstrative can only be used with a partitive reading. Otherwise, it is semantically infelicitous. There are no such semantic restrictions in (25b), however, where the use of the definite demonstrative is perfectly felicitous.

Exactly the same patterns are found in ergative languages, where the undergoer in an antipassive clause is typically indefinite, non-specific and non-presuppositional, whilst the undergoer in an ergative clause is typically definite, specific and presupposed. This can be illustrated from South Baffin Eskimo in (26):

- (26) *South Baffin Eskimo*
- a. **Antipassive**
- | | | |
|-------------------------|-----------------------|-------------|
| Joosi | quqiq-si-y-up | tutu-mik. |
| Joosi.ABS | shoot-ANTIP-PTCP-INTR | caribou-INS |
| ‘Joosi shot a caribou.’ | | |
- b. **Ergative**
- | | | |
|--------------------------------------|-------------------------|--------------------|
| Joosi-up | quqi-kkaniq-t-a-nga | tutu. |
| Joosi-ERG | shoot-again-PTCP-TR-3/3 | caribou.ABS |
| ‘Joosi shot the same caribou again.’ | | (Kalmar 1979: 124) |

In the antipassive in (26a), the undergoer ‘caribou’ is non-specific and indefinite. In the ergative clause in (26b), in contrast, the undergoer ‘caribou’ refers to a specific and given discourse referent. Hence, the Philippine-type restrictions are typical of ergative/antipassive alternations cross-linguistically.

However, there are also problems with this account. Firstly, unlike in canonical antipassive constructions and extended intransitives, the undergoer can be shown to be a core argument in AV (cf. Kroeger 1993, Riesberg 2014, SUBSECTION 1.4.2.1.3). Secondly, as discussed in Foley (2008), Himmelmann (2005a) and Kaufman (to appear), it is typologically unusual for antipassives to be expressed using the same morphology as basic intransitive predicates, though this is the analysis of Tagalog given in (24). Thirdly, if IV, LV and BV are treated as applicatives, we would need to

explain why peripheral arguments obligatorily have subject properties in these constructions and why the applicative markers do not co-occur with the irrealis UV suffix *-in*, if this is analysed as a marker of transitivity rather than voice (see Kaufman to appear).³⁵ Finally, AV constructions do not appear to be derived from UV constructions, although this is typical of antipassives (cf. Katagiri 2005).

Furthermore, as argued in Riesberg (2014), the main support for treating AV as an antipassive is the semantic restriction. However, the preference against definite undergoers illustrated in (25) is a tendency, rather than an outright constraint. There are cases, such as (27), where the undergoer must be interpreted as definite as the result of pragmatic inference:

- (27) *Tagalog*
 a. **Definite Undergoer in AV**
 Mag-bu~buslo ng bola si Gilbert.
 AV-REDUP~shoot CORE ball SUBJ Gilbert
 ‘Gilbert will shoot the ball.’ (Aldridge 2004: 3)

Thus, the definiteness restriction does not apply for all AV clauses, as might be expected from an antipassive.

Consequently, the ergative analysis also faces a number of problems. It could be considered preferable to the accusative hypothesis, as it provides a clear account of the ‘patient prominence’ effects and the core argument properties of the UV actor. However, it does not account for the fact that AV clauses appear to be transitive, rather than detransitivised versions of UV or extended intransitives, and leaves some typologically unusual patterns to be explained.

³⁵ Nb. An ergative analysis does not necessarily entail an applicative analysis (Elizabeth Zeitoun, p.c.).

1.4.2.1.3 The Philippine-type Alignment Hypothesis

On account of the problems associated with both the accusative and the ergative hypotheses, a third proposal has been made. This is that the languages of the Philippines have their own alignment systems – often referred to as Philippine-type alignment – in which both AV and UV are transitive clauses, and the alignment differs depending on whether AV or UV is compared with an intransitive. Under this analysis, the verbal morphology in Tagalog is treated as marking non-demoting and symmetrical voice alternations (see SUBSECTION 1.3). This has been the standard analysis in much of the typological literature and is adopted in Kroeger (1993), Himmelmann (2005a), Foley (2008) and Riesberg (2014), among others.

The Philippine-type alignment hypothesis can be supported by the fact that both the UV actor and the AV undergoer have core argument properties, unlike passive and antipassive constructions (see Kroeger 1993: 22). The actor in UV is clearly a core argument, as can be seen if we compare a passive construction with a UV construction in a language like Panguturan Sama:

- (28) *Panguturan Sama*
- a. **Undergoer Voice**
ø-bəlla dənda kiyakan kami.
UV-cook girl food 1PL.GEN
'The girl cooked our food.'
- b. **Passive**
B<i>lla uk dənda kiyakan kami.
<PASS>cook by girl food 1PL.GEN
'Our food was cooked by the girl.' (Kroeger 2004: 304)

Much like Indonesian in (7), the actor of the Sama passive clause in (28b) is marked with a preposition *uk*. In contrast, the actor in the UV clause in (28a) is a core NP, with no oblique marking. In addition, the agent of the passive clause can be omitted and

displaced to the end of a sentence, whilst the agent of the UV construction cannot (Kroeger 2004: 304). Hence, the actor in UV is a core argument, and UV is not a canonical passive.

Along similar lines, there are also morphosyntactic arguments for treating AV undergoers as core arguments. For example, Kroeger (1993) demonstrates that AV undergoers are treated differently from obliques in adjunct fronting. Unlike left dislocation or topicalisation constructions, in adjunct fronting there is no pause after the fronted element, and clitics, such as the pronoun *siya*, immediately follow the fronted constituent (Kroeger 1993: 43). Obliques can participate in adjunct-fronting, as shown in (29a), but AV undergoers cannot, as shown in (29b):

(29) *Tagalog Adjunct Fronting*

a. **Fronted AV Oblique**

[Sa pamamagitan ng sandok] siya kumuka ng sabaw.
 DAT use GEN ladle 3SG.NOM AV.PFV.take GEN soup
 ‘With the ladle, she took some soup.’

b. **Fronted AV Undergoer**

*[Ng balot] siya kumain.
 GEN balot 3SG.NOM AV.PFV.eat
 For: ‘The balot, he ate.’ (Kroeger 1993: 47)

Since the AV undergoer cannot participate in adjunct-fronting, Kroeger (1993) argues that *ng balot* ‘the balot’ does not function as an oblique, but rather a core argument of the verb.³⁶

A second argument for treating the AV undergoer as core is that it can control the reference of a gap in participial *nang* clauses, whereas obliques cannot (cf. Kroeger 1993):

³⁶ Aldridge (2004) argues that adjunct fronting is not sensitive to the core-oblique distinction but rather the definiteness of the argument. She argues that definite/specific arguments like ‘with the ladle’ can be fronted, whereas non-specific and indefinite arguments like *ng balot* ‘balot’ cannot be. Hence, she dismisses this as an argument for treating the AV undergoer as core.

(30) *Tagalog Participial nang clauses*

a. **AV undergoer as controller**

Nanghuli ng magnanakaw ang polis [nang
AV.PFV.catch GEN thief NOM police ADV

pumapasok sa bangko].

AV.IPFV.enter DAT bank

‘The police caught a/the thief when entering the bank.’

Interpretation 1: the police entered the bank

Interpretation 2: the thief entered the bank

b. **AV oblique as controller**

Bumista si Juan sa hari [nang nagiisa].

AV.PFV.visit NOM Juan DAT king ADV AV.IPFV.one

‘Juan visited the king alone.’

Only possible interpretation: Juan was alone

Ungrammatical: the King was alone

(Kroeger 1993: 47)

In (30a), the gap can be controlled by either the actor, *ang polis* ‘the police’, or the undergoer, *ng magnanakaw* ‘the thief’. Hence, it is ambiguous as to whether the policeman or the thief is entering the bank when the event takes place. In contrast, in (30b) the participial clause in brackets can only be controlled by the actor, *Juan*, and not the oblique, *sa hari* ‘the king’. In other words, the sentence cannot be understood as Juan visiting the king whilst the king is alone. Thus, it seems that the GEN marked NP ‘thief’ in (30a) is a core argument of the verb, whilst the DAT marked NP ‘king’ in (30b) is not.³⁷ Consequently, Kroeger (1993) argues that an account of AV as an

³⁷ Aldridge (2004) also argues that the evidence of *nang* clauses is not sufficient on the basis that obliques can control participial clauses in other contexts. For example, the participial clause is controlled by a *sa* PP in (i):

- (i) Nag-utos ang nanay sa anak=niya_i-ng [pro_i mag-bantay ng bahay].
AV.PFV-order NOM mother DAT child=3SG.GEN-LNK AV-watch GEN house
‘The mother ordered her child to watch the house.’ (the child watches the house)

In (i), the PP ‘to the child’ controls the gap in the participial clause despite not being a core argument of the verb. However, it is not clear to what extent these facts would revoke the patterns of *nang* clauses – where the AV undergoer can be shown to behave differently to obliques. Indeed, this could be a lexical fact about the verb *utos*. Hence, the data does invalidate Kroeger’s (1993) original argument.

antipassive construction cannot be upheld since AV constructions are no less transitive than UV ones.

Thus, the languages of the Philippines can be argued to have symmetrical voice alternations in which transitive clauses of both accusative (AV) and ergative (UV) type co-exist. This has the advantage of accounting for the core argument properties of both AV and UV, unlike the other alignment hypotheses. However, the question of the definiteness restriction and patient prominence remains unanswered, and treating alignment as symmetrical dissociates Western Austronesian languages from other voice alternations. These issues are returned to in CHAPTER 3.

1.4.2.2 The Alignment Debate in Indonesian-type Languages

Much like Tagalog, there has also been debate surrounding alignment in Indonesian-type languages. This section briefly sketches the main arguments for the accusative, ergative and symmetrical analyses and then introduces Aldridge's (2011) proposal for alignment shift.

1.4.2.2.1 The Accusative Hypothesis

Traditional analyses of Indonesian treat *meN-* verbs as active and both *di-* verbs and bare verbs with proclitic actors as variations on the passive (cf. Chung 1976, Sneddon 1996). This allows for an accusative analysis, as schematised in (31) using examples adapted from Musgrave (2002):

- (31) *Accusative Hypothesis for Indonesian*
- a. **Active (AV)**
- | | | |
|-----------|----------|-----------|
| Hasan (A) | mem-beli | ikan (U). |
| Hasan | ACT-buy | fish |
- ‘Hasan bought fish.’

b. **Intransitive**

Hasan (S) duduk.
Hasan sit
'Hasan sat.'

c. **Passive 1 (UV)**

Ikan (S) di-beli Hasan.
Fish PASS-buy Hasan
'The fish was bought by Hasan.'

d. **Passive 2 (UV)**

Ikan (S) saya=beli.
Fish 1SG=PASS.buy
'The fish was bought by me.' (adapted from Musgrave 2002)

Under an accusative hypothesis, both *di*-V NP and *pro*=V constructions are treated as passive clauses, in which the undergoer functions as the single (S) argument of an intransitive predicate. This is treated in the same way as the actor (A) of a transitive clause and the single (S) argument of basic intransitive clauses, since they all appear in pre-verbal position. In contrast, the undergoer (U) of a transitive clause comes post-verbally. Hence, Indonesian could be argued to have canonical accusative alignment in that S and A are treated alike, and U differently. This analysis is adopted in many contemporary accounts of Indonesian alignment, including Aldridge (2008, 2011), Cole et al (2008), Chung (2008) and Kaufman (to appear) on the basis that *di*-clauses share characteristics with passives (see SUBSECTION 3.4.2).

However, as discussed in Riesberg (2014), analysing UV as passive results in the typologically unusual situation whereby active is marked, and passive is unmarked in languages like Balinese (see SUBSECTION 1.4.2.2.2). Moreover, we would have to assume that there is a single active clause and multiple passive constructions, which is typologically rare (cf. Foley 2008). Hence, the analysis is somewhat problematic.

1.4.2.2.2 The Ergative Hypothesis

An alternative is to analyse Indonesian-type languages as having ergative alignment. This has been proposed for Indonesian on the basis that *di-* clauses represent foregrounded events in narrative (see Hopper 1979, 1983, and Verhaar 1988).³⁸ Similarly, Balinese has been described as ‘discourse ergative’ on the basis that UV is more frequent in discourse (Wechsler & Arka 1998, CHAPTER 3). The strongest argument for ergative alignment in languages like Balinese, however, is that UV is unmarked in contrast to AV. This could be taken to support an analysis of AV as a derived antipassive, which is schematised below using data from Arka (2000) and Artawa (2013):

- (32) *Balinese*
- a. **Antipassive (AV)**
Tiang (S) ng-lempag ipun.
1SG ANTIP-hit 3SG
‘I hit him.’
- b. **Ergative (UV)**
Ipun (U) lempag tiang (A).
3SG UV.hit 1SG
‘He was hit by me.’ (Arka 2000)
- c. **Intransitive**
Anak-e cenik ento (S) labuh.³⁹
child-DEF small that fall
‘The small child fell.’ (Artawa 2013: 7)

If AV is treated as a derived antipassive, then S and U are treated alike, occurring in pre-verbal position, whilst A is treated differently, occurring post-verbally. This resembles canonical ergative alignment.

³⁸ See also Ahmady (2009) on Pancor Ngenó-Ngené Sasak.

³⁹ Nb. Artawa (2013) argues that we should further distinguish between intransitive predicates in Balinese with an actor S, and those with an undergoer S.

However, such an account would be problematic as it would require analysing the AV as an antipassive. This does not seem motivated, as the undergoer can be shown to be a core argument in AV (SUBSECTION 1.4.2.2.3). Moreover, Balinese does not share the restriction against definite undergoers in AV, illustrated for Tagalog, since the undergoer in (32a) can be pronominal. This suggests that the undergoer is high in definiteness and/or referentiality, which would not be expected of antipassives (see SUBSECTION 3.2.1.2). Hence, there are only limited semantic arguments for such an analysis. For these reasons, Davies (1991) and Arka (1998) argue against the ergative hypothesis for Indonesian-type languages.

1.4.2.2.3 The Symmetrical Hypothesis

The final possibility is to assume that Indonesian-type languages are symmetrical, as proposed by Riesberg (2014) and Arka (2005). This is supported by the fact that there is syntactic evidence for treating the actor as a core argument in UV, and the undergoer as a core argument in AV (cf. Riesberg 2014). In other words, AV is not an antipassive and UV is not a passive construction.

One argument for treating AV undergoers and UV actors as core arguments comes from quantifier floating. In both Indonesian and Balinese, only core arguments can launch floating quantifiers (cf. Arka 2003). Hence, this is a test that can be used to distinguish core arguments from obliques. Consider the patterns from Indonesian in (33):

- (33) *Indonesian*
- a. **Quantifier Float launched by AV undergoer**
- | | | | | | |
|------|--------|-------------|-----|-----------|------------|
| Saya | mukul | anak~anak | itu | kemarin | semua-nya. |
| 1SG | AV.hit | child~REDUP | DEM | yesterday | all-3 |
- ‘I hit all the children yesterday.’

- b. **Quantifier Float launched by UV actor**
 Anak~anak kami pukul kemarin semua-nya.
 child~REDUP 1PL.EXCL UV.hit yesterday all-3
 ‘All the children were hit by us, yesterday.’
 Or ‘The children were hit by all of us, yesterday.’
- c. **Quantifier Float launched by intransitive oblique**
 *Orang~orang Sasak datang dengan anak-anak semua-nya.⁴⁰
 People~REDUP Sasak come with child-REDUP all-3
 For: ‘The Sasak people came with all their children.’
 (Musgrave 2002: 70)

In (33a), the quantifier *semuanya* ‘all’ is understood as modifying the undergoer argument of an AV clause, *anak-anak* ‘children’. In (33b), the quantifier *semuanya* ‘all’ can be understood to modify either the undergoer of a UV clause, *anak-anak* ‘children’, or the actor, *kami* ‘1SG.EXCL’. However, in (33c), the quantifier *semuanya* ‘all’ cannot be understood to modify *anak-anak* ‘children’ when they are expressed as an oblique PP, *dengan anak-anak* ‘with their children’. This suggests that the AV undergoer and the UV actor are core arguments, as they can both launch quantifier float, unlike obliques.

Arka (2005: 7) provides further support for the symmetrical analysis by calculating the core-index of arguments in AV and UV constructions in Indonesian and Balinese using both cross-linguistic and language specific tests, drawn from Arka (2003) and Musgrave (2002) among others. The core index is equivalent to the number of properties that an argument positively satisfies and ranges from 0-1. The tests are shown in TABLE 1.7:

⁴⁰ This could only mean ‘all the Sasak people came with their children’, where the floating quantifier is launched by the AV actor, rather than the oblique.

Table 1.7 Core Properties in Indonesian (Arka 2005)

Property	A _{AV}	U	A _{UV}	A _{PASS}
Quantifier Float with <i>semua</i>	√	√	*	*
Topicalisation of possessor phrase	√	√	n/a	*
Topicalisation with resumptive pronoun	√	√	√	*
Depictive predicate	√	√	*	*
Imperative actor (=zero)	√	n/a	√	*
Binding: binder of a core	√	*	√	*
Verbal marking: participates in voice alternation	√	√	√	*
Categorial marking	√	√	√	?
Obligatory	√	√	√	*
Proclitic on verb	√	√	√	*
Fixed structural position	√	√	√	*
Core Index	1	0.82	0.72	0.04

As shown in TABLE 1.7, Arka (2005) found that both the actor and the undergoer had a core index of over 0.80 in AV, whilst the actor has a core index of 0.72 in UV. This is much higher than the actor of a passive clause, which has a core index of 0.04. For this reason, Arka concludes that both arguments in AV and UV are core, and that UV is distinct from a passive construction.⁴¹

When Shibatani (2005) applied the core index tests to the Philippine-type language Cebuano, he showed that both the actor and the undergoer of UV clauses were highly core, whilst the undergoer of an AV clause had a low index of 0.09 – considerably less than the equivalent in Indonesian or Balinese (cf. Arka 2005: 14). Thus, it seems that AV and UV clauses in ‘Indonesian-type’ languages may be more ‘symmetrical’ than in the ‘Philippine-type’ languages (see SUBSECTION 3.4). Nonetheless, on the basis that Indonesian, like Tagalog, shows syntactic evidence for multiple transitive clause-types, it seems that the ergative and accusative hypotheses, at least in the canonical sense, cannot be upheld and that a symmetrical analysis of voice in Indonesian and Balinese provides a more adequate account of the data.

⁴¹ See Kroeger (2014) for critical discussion of other core-argument tests.

1.4.2.3 Theory of Alignment Shift

Before concluding this chapter, it should be noted that different accounts of synchronic alignment in Western Austronesian have led to the proposal that Western Austronesian languages have undergone alignment shift. Aldridge (2011) proposes that Western Austronesian languages have undergone a shift from ergative to accusative alignment on the basis that the more conservative Philippine-type languages, like Tagalog, appear to have ergative-like properties (SUBSECTION 1.4.2.1.2) and the more innovative Indonesian-type languages, like Indonesian, have developed accusative-like properties (SUBSECTION 1.4.2.2.1).⁴² She argues that the transition occurs through the reanalysis of antipassive AV as active/transitive, and the subsequent reanalysis of ergative UV as passive. Moreover, she argues that many of the typological differences between Philippine-type and Indonesian-type languages may follow from the various stages of reanalysis.

So far, such a proposal does not seem warranted as both Philippine-type languages and Indonesian-type languages can be argued to have symmetrical voice alternations in which AV and UV constructions are equally transitive. This distinguishes AV from a canonical antipassive and UV from a canonical passive in both typological groups, ruling out either of the canonical alignment systems at the level of morphosyntax. Nonetheless, semantic and discourse differences, particularly regarding ‘patient prominence’, are evident in the different voice systems and remain

⁴² The directionality of change in Aldridge (2011) can be assumed on the basis of widely accepted subgrouping within Austronesian (cf. Ross & Arka 2005, Blust 2013, SUBSECTION 1.2). The Philippine-type languages of Taiwan and the Philippines have an extensive system of verbal morphology and case-marking, as shown in SUBSECTION 1.3.1. An account of Austronesian that assumed a change from Indonesian-type to Philippine-type would have to account for how these systems developed out of the reduced voice-system and zero case-marking found in Western Indonesia. Occam’s razor suggests that a change by which morphosyntactic complexity is lost is simpler than one by which it is developed. What is perhaps more controversial about the Aldridge (2011) account is couching the change from Philippine-type to Indonesian-type in terms of alignment shift. This will be further discussed in CHAPTER 3.

to be further explored. Consequently, I return to the theory of alignment shift in CHAPTER 3.

1.4.3 Summary

In this section, I have highlighted two key debates in the study of Austronesian voice and presented a selection of arguments for the various hypotheses. The first debate concerns whether Western Austronesian languages have ‘subjects’ given the split subject properties in SUBSECTION 1.4.1. I argued, following Manning (1996), that subject could be identified as the argument selected by the voice morphology and that ‘role-related’ subject properties are better handled at argument structure. The second debate concerns the question of alignment and whether Western Austronesian languages have their own alignment systems or can be analysed as either ergative or accusative. I presented evidence from semantics for treating UV as ergative, and AV as antipassive in Philippine-type languages. Equally, I suggested that there were some similarities between UV and passive in Indonesian. However, I argued, following Kroeger (1993) and Riesberg (2014), that AV and UV are transitive in both language groups on the basis of morphosyntactic evidence, and that the alternations are therefore symmetrical, as outlined in SUBSECTION 1.3. This suggests that Western Austronesian languages are neither ergative nor accusative, at least not in the canonical sense. Finally, I introduced the hypothesis that Western Austronesian languages are undergoing a change from ergative to accusative alignment, which is reflected in typological differences between Philippine-type and Indonesian-type languages. This may well explain semantic and discourse differences, which are explained in more detail in CHAPTER 3. The rest of the chapter outlines the structure of the thesis and how

to explore what Kelabit can tell us about Austronesian syntax and syntactic typology more generally.

1.5 Structure of the Thesis

The rest of the thesis is structured as follows. CHAPTER 2 introduces further details about the Kelabit language and ethnographic context, based on linguistic fieldwork in the Kelabit Highlands. It provides a basic grammatical description, including phonology, morphology and syntax, in order to clarify data and analyses presented in the later chapters. Finally, it addresses the question of grammatical functions in Kelabit and the implications that this has for the subject debate.

CHAPTER 3 returns to the question of voice. It presents the Kelabit voice system and proposes a methodology for analysing alignment when voice alternations appear morphosyntactically symmetrical. This involves comparison of the different voices on morphological, syntactic, semantic and discourse levels with Philippine-type systems and Indonesian-type systems. It constitutes the first example of why a two-way typology struggles to capture synchronic variation in Western Austronesian and provides some support for the notion of alignment shift, at least on semantic and discourse levels.

CHAPTER 4 is concerned with pronominal systems. It presents a set of variant pronouns in Kelabit, which are used for actors in non-actor voices but do not appear to have the typical case patterns found in the Philippines. It establishes parameters of variation, both in terms of morphosyntax and prosody, and develops a methodology for analysing clitic phenomena in Kelabit. The results support the conclusion that Kelabit is intermediate between Philippine-type and Indonesian-type languages.

CHAPTER 5 examines word order and constitutes the final case study of variation in Western Austronesian. It demonstrates that there are different word-order patterns according to the voice construction in Kelabit, explores possible explanations for variation and compares word-order choices with Philippine-type and Indonesian-type languages. It reinforces the findings of the previous two chapters, namely that a two-way typology cannot capture the full extent of variation in Western Austronesian and that the Kelabit AV construction appears more innovative than UV.

Finally, CHAPTER 6 concludes, addressing the implications of the Kelabit voice system for the major debates within Western Austronesian syntax and typology. It comments on the extent to which a two-way classification of Western Austronesian languages as Philippine-type and Indonesian-type is adequate and proposes avenues for future research.

1.6 Conclusion

In this chapter, I introduced the typologically rare phenomenon of symmetrical voice alternations that seem to characterise Western Austronesian languages. I stated that Austronesian scholars typically subdivide Western Austronesian into Philippine-type and Indonesian-type on the basis of typological differences. I also demonstrated that both Philippine-type and Indonesian-type languages have been subject to two major debates regarding the nature of grammatical functions and the question of alignment. Finally, I introduced the hypothesis that Western Austronesian languages have undergone a shift in alignment from ergative to accusative. This sets the scene for the two central questions that are explored in this thesis through an analysis of the Kelabit voice system and related phenomena:

1. Is the two-way typology of ‘Philippine-type’ and ‘Indonesian-type’ sufficient to capture the variation within Austronesian languages?
2. What can Kelabit tell us about theoretical debates and theories of change between the more conservative Philippine-type languages and the more innovative Indonesian-type languages?

Kelabit provides an ideal opportunity to explore these questions as it is at a point of transition from Philippine-type to Indonesian-type (see SUBSECTION 2.2.1). If these two categories represent different points in an alignment shift and/or other historical changes, then Kelabit could not only provide evidence for intermediate stages in the transition but also tell us something very interesting about how these sorts of large-scale structural changes take place. Consequently, the rest of the thesis compares Kelabit with other Western Austronesian languages, beginning in the next chapter with an outline of the Kelabit language and its ethnographic context.

Chapter 2

The Kelabit Language

2.1 Introduction

In this chapter, I introduce the Kelabit language of Northern Sarawak, spoken on the island of Borneo in Sarawak, East Malaysia.⁴³ Borneo is characterised by huge linguistic diversity, yet there has been relatively little research done regarding the linguistic situation (see Ray 1913, Cense & Uhlenbeck 1958, Asmah 2004). Nonetheless, the indigenous languages of Borneo - and Northern Sarawak in particular - are worthy of more attention as they lie genetically and geographically between ‘Philippine-type’ languages and ‘Indonesian-type’ languages (Hudson 1978, see FIGURE 2.1). Consequently, they may reveal important information about developments within Austronesian and the ‘symmetrical voice’ languages as a whole (see CHAPTER 1).

This chapter presents a sketch grammar of Kelabit in order to contextualise the more detailed case studies in CHAPTERS 3, 4 and 5. There are very few existing resources on Kelabit, as discussed in APPENDIX 1. Hence, the description is based on

⁴³ The origin of the term ‘Kelabit’ is not known (though see Schneeberger 1979: 29 and Harrisson 1959b for potential etymologies). It was not originally used by speakers, who instead referred to *karuh tauh* ‘our language’ but has since come into use as a form of self-reference in the community within the last two generations (Saging 1976/77: 4-12).

primary linguistic fieldwork and documentation over a period of six and a half months between October-December 2013 and June-September 2014. The data were collected following the methodology of language documentation and description, as outlined in Himmelmann (1998, 2006a) and Woodbury (2003, 2011). This involves ‘the creation, annotation, preservation and dissemination of transparent records of a language’ (Woodbury 2011: 159). Consequently, a corpus of audio, video and written materials was collected, including elicitation sessions and texts in a variety of genres. More information on the methods used in compiling the corpus and the nature of the recordings can be found in APPENDIX 1.

This chapter is structured as follows. SUBSECTION 2.2 provides information on the classification of Kelabit and the sociolinguistic and ethnographic context in which it is spoken. SUBSECTION 2.3 gives a basic sketch of the phonology, discussing the phoneme inventory, syllable structure, stress and phonological alternations. SUBSECTION 2.4 gives a basic sketch of the morphology, including word formation processes and word classes, and SUBSECTION 2.5 discusses Kelabit syntax, including grammatical functions, periphrastic voices, multi-clausal constructions and the implications that these have for the subject debate (SUBSECTION 1.4.1).

2.2 The Kelabit Language

2.2.1 Classification

Like all of the languages of Borneo, Kelabit belongs to the Western Malayo-Polynesian branch of Austronesian (SUBSECTION 1.2). However, further subdivision has been problematic on account of the complex linguistic situation and

relative lack of systematic comparative work (cf. Kroeger 1998a, Asmah 2004).⁴⁴

Hudson (1978) divides the languages of Borneo into ten groups on the basis of shared innovations and lexicostatistical similarities:

(1) **Subgroups in Borneo**

- a. *Land Dayak*
- b. *Rejang-Baram*
- c. *Kenyah-Kayan*
- d. *Apo Duat*
- e. *West Barito*
- f. *Barito-Mahakam*
- g. *East Barito*
- h. *Malayic*
- i. *Tamanic*
- j. *Sabahan*

The first seven groups are indigenous to Borneo, though the Barito languages are thought to be related to Malagasy of Madagascar (see Adelaar 1995). Malayic, Tamanic and Sabahan are known as ‘exo-Bornean’ since they are closely related to other Western Austronesian languages spoken outside of Borneo. The Malayic languages in Southern Borneo are related to the languages of Western Indonesia.⁴⁵ The Tamanic languages in Central and Eastern Borneo are related to the languages of South Sulawesi and the Sabahan languages in Northern Borneo are related to the languages of the Philippines (Hudson 1978, Adelaar 1995).

Blust (1974a and elsewhere) argues that the Rejang-Baram, Kenyah-Kayan and Apo-Duat languages form a single subgroup, which he calls North Sarawak. Furthermore, he argues that the North Sarawak languages share a common ancestor with the more conservative Sabahan languages: North Borneo. He suggests that

⁴⁴ Though see Greenhill, Blust & Gray (2008) for a more recent approach using Phylogenetic methods and increased comparative data through the Austronesian Basic Vocabulary Database.

⁴⁵ See Adelaar (1992) for discussion of the proposal that Malayic languages are indigenous to Borneo.

Proto-North Borneo was spoken in coastal parts of Western Sabah around 2,000 BC before later splitting into the two groups.⁴⁶ If the North Borneo subgroup is assumed, then languages of Sarawak are mainly Malayic and Land Dayak in the south and North Borneo in the north, as shown in FIGURES 2.1 and 2.2.

⁴⁶ Based on the ‘Vowel Deletion Hypothesis’, which is a posited innovation to explain double reflexes of PAn voiced obstruents (see Blust 1974a). However, since no other phonological or morphological shared innovations have been discussed, this has not always been adopted (cf. Charles 1974, Kroeger 1998a:145, Hudson 1978).

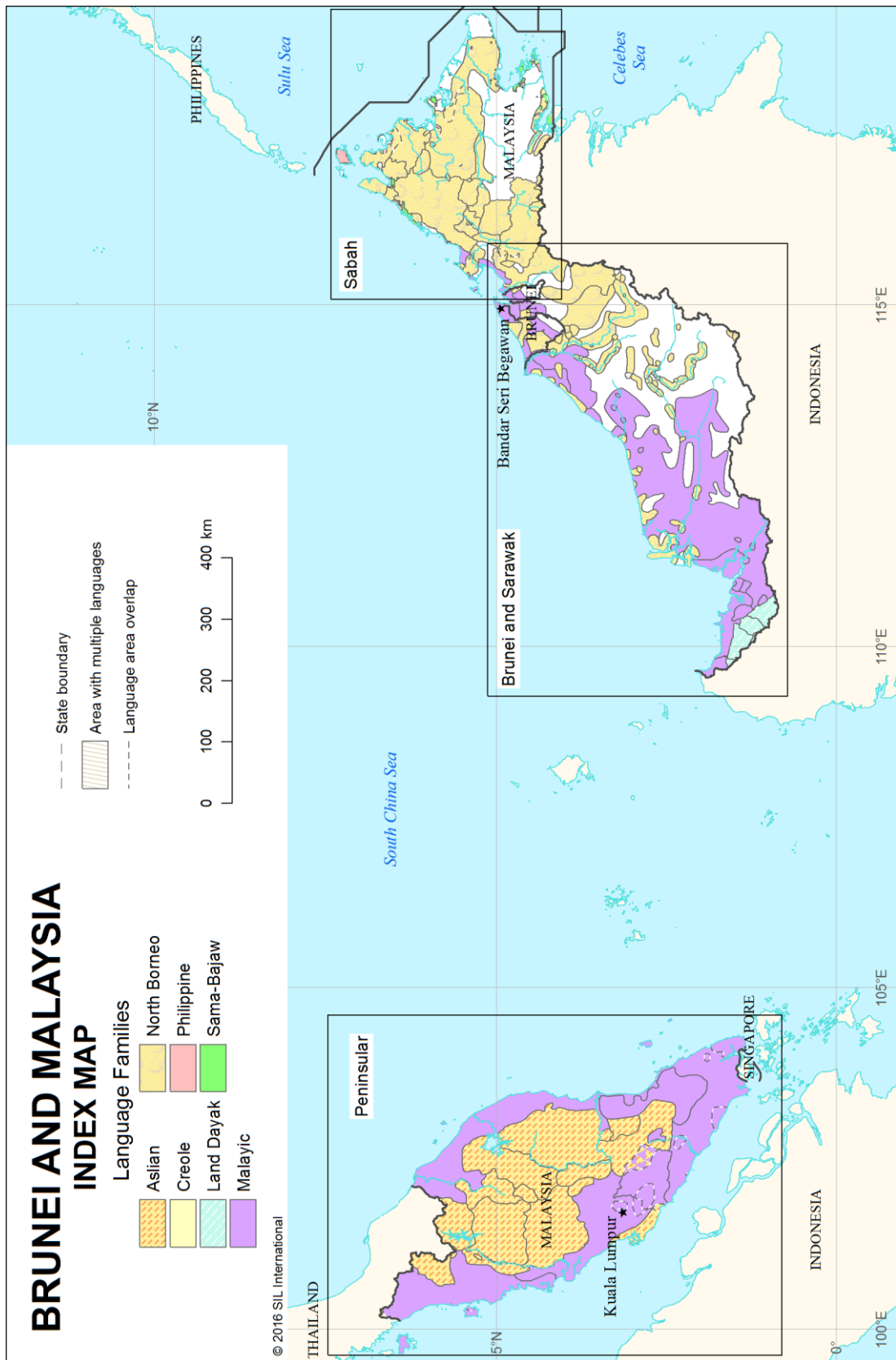


Figure 2.1 The Languages of Malaysia (Used by Permission Lewis et al 2016)

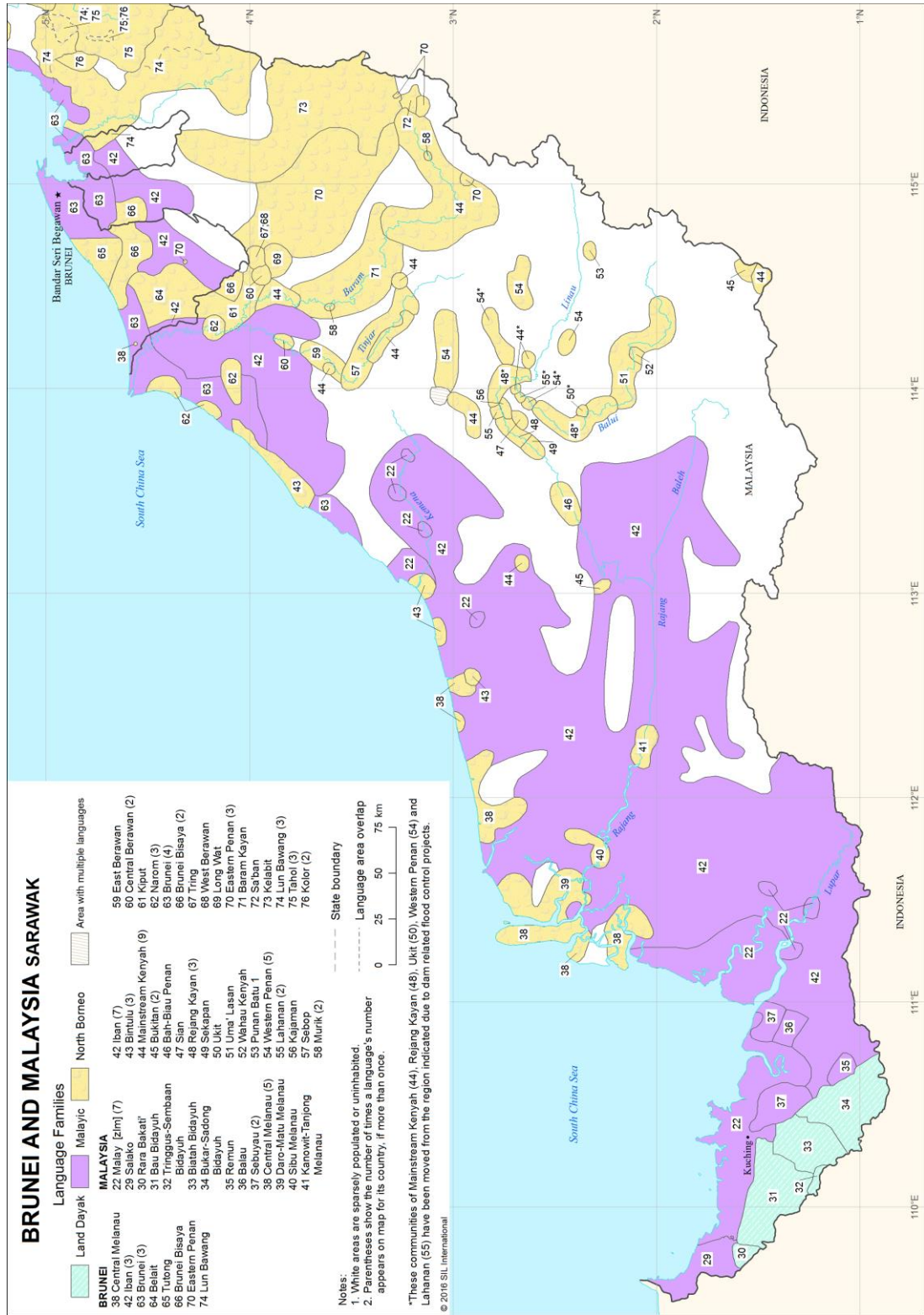


Figure 2.2 The Languages of Sarawak (Used by Permission Lewis et al 2016)

Kelabit is a North Borneo language and a member of Hudson's (1978) Apo Duat subgroup, which also includes Lun Bawang/Lundayeh, Tring and Sa'ban (Martin 1996, FIGURE 2.3). It is spoken mainly in the Fourth and Fifth divisions of Sarawak, Malaysia, though related languages are spoken across the border in parts of Kalimantan, Sabah and Brunei (Martin 1996, see FIGURE 2.2).⁴⁷ As is generally true of the languages of Borneo, Kelabit has historically been classified using a range of problematic and confusing labels, including 'Dayic', 'Orang Ulu', 'Kelabitic' and 'Murut'. Traditionally, all interior peoples, including Kelabit, were grouped with the indiscriminate term 'Dayak' creating great confusion (Roth 1896, Schneeberger 1979, King 1993). Today, the state government of Sarawak uses the term 'Orang Ulu' (meaning upriver people) to refer to several groups including the Kelabit, Kenyah, Kayan and Penan (cf. Kroeger 1998a). However, the term has no cultural or linguistic meaning but is merely a residual category referring to any non-Muslim group that is not part of the dominant Iban, Malay, Bidayuh or Melanau ethnicities (cf. Asmah 2004). Earlier works used the term 'Murut' to refer to Kelabit and related languages (Appell 1969, Pollard 1933, LeBar 1972). However, this created confusion with unrelated ethnic and linguistic groups in Sabah (cf. Prentice 1970: 370, Langub 1987, Bolang & Harrisson 1949). Blust (1974a) proposed 'Dayic' but this was disfavoured on account of confusion with the term 'Dayak'. Dyen (1965) and Kroeger (1998a) use 'Kelabitic' but this has been objected to as it favours one group over the others (cf. Hudson 1994). Finally, Hudson (1978) uses Apo Duat, a neutral term from the mountain range on the border with Indonesia. However, this is said to be a mishearing of *Apad Uat* (Eghenter & Langub 2008). Hence, I use the term Apad Uat in this thesis.

⁴⁷ It is thought that groups migrated to Western and Northern Borneo in the early nineteenth century, following the Trusan, Limbang and Padas rivers (Edwards & Stephens 1971). The Kelabit are also thought to be related to the Kerayan and Berian peoples of Brunei and Indonesia (cf. Bala 2002: 19).

The Apad Uat family tree is schematised in FIGURE 2.3:⁴⁸

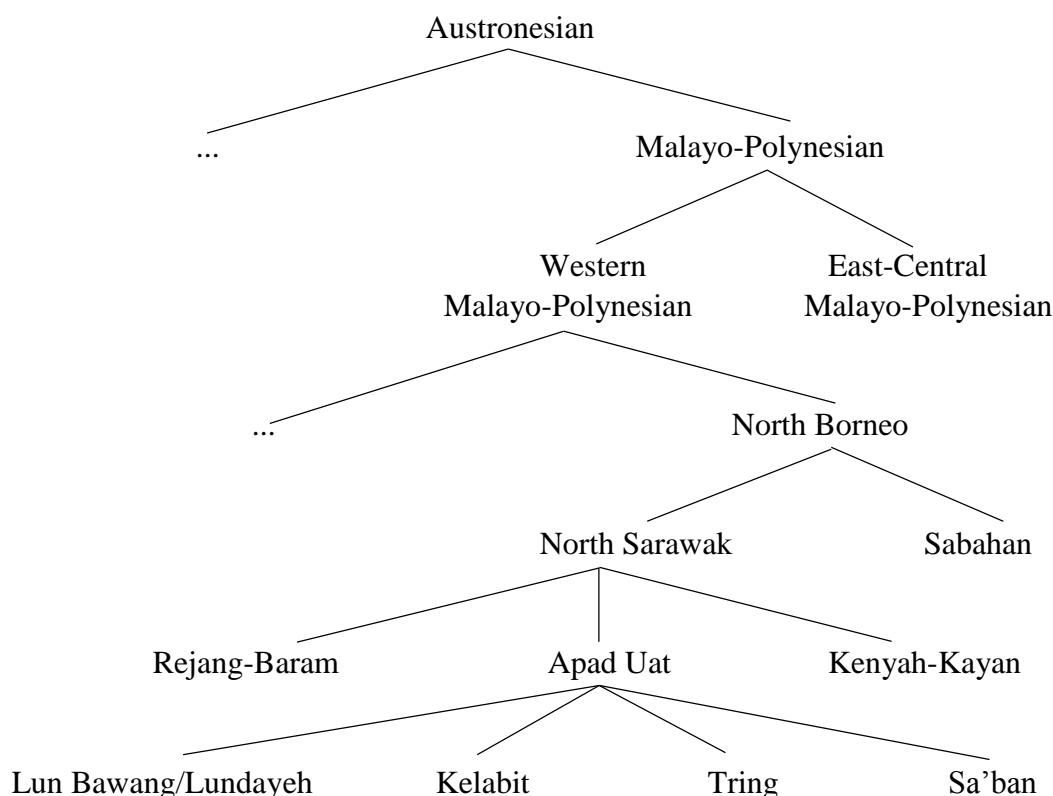


Figure 2.3 Apad Uat Family Tree (cf. Blust 1993)

The group is lexicostatistically cognate at 70% and shares phonological innovations, such as the merger of PAN phonemes *j, *D *d *Z and *z with /d/ (Hudson 1994: 22), and lexical innovations, such as *rudap* ‘sleep’ and *birar* ‘yellow’ (Hudson 1994: 21). Lun Bawang/Lundayeh is the most conservative of the Apad Uat languages, and Sa’ban the most innovative, whilst Kelabit and Tring fall somewhere between the two extremes (Blust 1993).⁴⁹ Among the languages of Sarawak, Lun Bawang is said to be

⁴⁸ Nb. The symbol ... reflects any number of additional subgroups that are not represented on the tree. This includes the Formosan languages, which have been subject to various subgrouping hypotheses, such as Blust (2013), Ross (2009) and Zeitoun & Teng (2014). It also includes other Western Malayo-Polynesian subgroups, such as the Philippine group, Barito group, Malayo-Chamic group and Celebic group. See SUBSECTION 1.2 for further discussion of subgrouping higher in the tree.

⁴⁹ Hudson (1978) originally distinguished Kelabit from all other Apad Uat languages. In contrast, Blust (1974a) initially singled out Lun Bawang/Lundayeh as distinct from Kelabit, Sa’ban and Tring but later added Sa’ban as a third branch based on considerable phonological innovations. Distinguishing Lun Bawang/Lundayeh from Kelabit and Sa’ban makes more sense than Hudson’s (1978) classification,

unique in displaying typical Philippine-type characteristics (Clayre 2005). In contrast, Clayre (2014) analyses Sa'ban as having a 'reduced voice system' similar to those found in Indonesian-type languages (see SUBSECTION 3.4.2). In CHAPTER 3, I discuss the Kelabit voice system and argue that it has both Philippine-type and Indonesian-type characteristics. Hence, the Apad Uat family appears at a point of transition, as alluded to in CHAPTER 1.

2.2.2 Dialect Geography

Kelabit is traditionally spoken in the Kelabit Highlands, a plateau in central Borneo which lies at the headwaters of the Baram River and is surrounded by the Apad Uat mountain range to the east and the Tamabu range to the west (Schneeberger 1945, Bala 2002: 13). Mount Murud, the highest peak in Sarawak, lies to the north of the Highlands at approximately 2,500 metres above sea level (Amster 2003: 253, Asmah 1983: 542). The villages in the Highlands are approximately 1,000 metres above sea level.

The Kelabit Highlands became an important military base during the Second World War and the Confrontation between Malaysia and Indonesia in the 1960s due to its proximity to the Indonesian border (see FIGURE 2.4). During this time, many of the villages close to the border were resettled around the longhouse of Bario Asal (Bala 2002). This area became known as Bario and has since emerged as the administrative centre of the Kelabit Highlands (Saging & Bulan 1989: 91).

Today there are 18 longhouse settlements in and around the Kelabit Highlands where dialects of Kelabit are spoken. The exact number of dialects remains to be

given that Lun Bawang verbal morphology, such as the stative prefix *ma-*, does not occur in either Kelabit or Sa'ban (cf. Clayre 1994). However, many people feel intuitively that Kelabit is quite different from Lundayeh, Kerayan and Sa'ban (Jayl Langub, p.c.). Relationships within the Apad Uat family remain to be further explored.

studied in further detail (cf. Blust 1993). However, Gerawat Nulun (p.c.) suggests that there may be four major dialects distinguished by the pronunciation of the word ‘day’: *edto*, *echo*, *eso* and *so*. The *edto* pronunciation is common in northern villages, like Bario, and the *so* pronunciation is typical of southern villages, like Pa’ Dalih. The rough location of several Kelabit villages is indicated in FIGURE 2.4.⁵⁰

⁵⁰ Nb. Spellings of village names differ from those used in the thesis and follow Colin Davis. Some villages are not shown, including Pa’ Ukat which is close to Pa’ Umur, and the villages outside of the Kelabit Highlands.

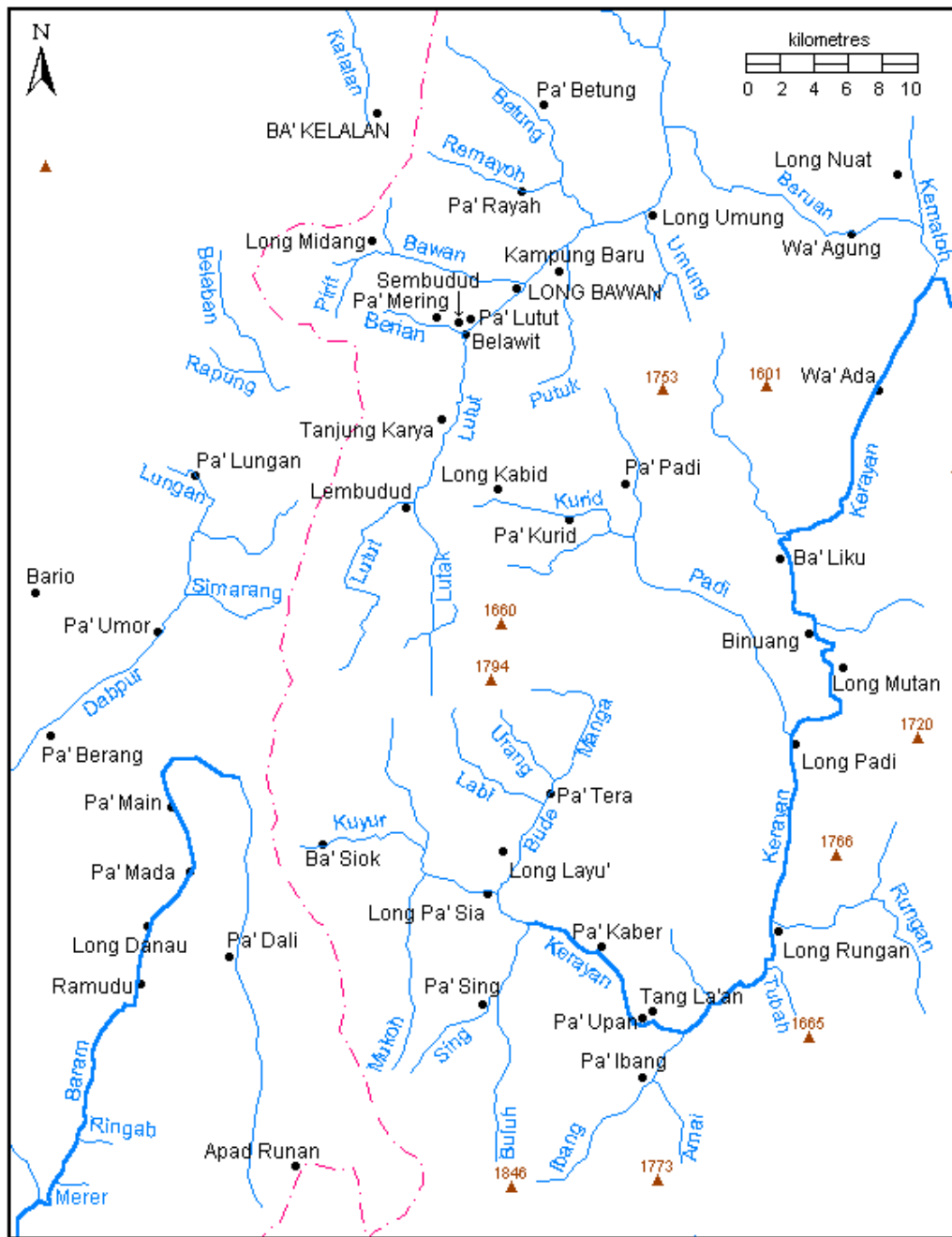


Figure 2.4 The Kelabit-Kerayan Highlands © Colin Davis

Bario is made up of a number of longhouse settlements in close proximity. These include Bario Asal, Ulung Palang⁵¹, Arur Dalan, Arur Layun, Pa' Ramapoh Atas and

⁵¹ Ulung Palang is the name given to the resettled community from Pa' Main. This was roughly in the middle of the Kelabit Highlands, between the Northern and Southern villages, as shown in FIGURE 2.4. It was the site of a salt spring (*main* 'salty') and one of the first schools in the Highlands. Today, there are primary schools in Bario and Pa' Dalih and a secondary school up to 14 in Bario.

Bawah, Pa' Derung, Padang Pasir and Kampung Baru.⁵² To the north of Bario, there are three villages located along the Debpur river, namely Pa' Ukat, Pa' Umur and Pa' Lungan. The first two are approximately half an hour's walk from Bario and accessible by car. Pa' Lungan is the northernmost Kelabit village and is four hours walk through the jungle.

Further downriver, towards the southern end of the Kelabit Highlands are three villages that lie along the Kelapang river: Pa' Mada, Pa' Dalih and Remudu. Although there are minor dialectal differences between the Kelabit spoken in the Northern villages, there is a very salient dialect boundary between Kelapang Kelabit – the dialect continuum spoken in Pa' Mada, Pa' Dalih and Remudu – and Bario Kelabit (Blust 1993). The most obvious differences are phonological. For example, Bario schwa sometimes corresponds to Kelapang /i/, Bario /u/ corresponds to Kelapang /o/, Bario /dt/ corresponds to Kelapang /s/ and Bario /d/ sometimes corresponds to Kelapang /r/ (see SUBSECTION 2.3.1 for discussion of allophonic variation in vowels):

(2) *Dialect Differences*

Bario Kelabit	Kelapang Kelabit	
[ŋadəl]	[ŋadil]	<i>ngadel</i> 'sharp'
[manək]	[manək]	<i>manuk</i> 'bird'
[əd ^h ɔ:]	[sɔ:]	<i>edto</i> 'day'
[dadan]	[radan]	<i>dadan</i> 'long time'

There are also lexical differences. For example, the adverb meaning 'later' is *na'an* in Bario Kelabit and *ano* in Kelapang Kelabit. However, the exact dialect differences remain to be further explored.

⁵² Place names in the Highlands are typically named for the rivers (*pa'*), streams (*arur*), and confluences (*long*) where the first longhouse settlements were built. Villagers would resettle every so often, according to traditional practice, and hence the present day sites of the villages may no longer be adjacent to the natural features for which they are named. Spellings of settlements follow the Bario Clinic rather than the orthography in Labang (2012) used in this thesis (SUBSECTION 2.3).

Finally, there are Kelabit settlements beyond the Kelabit Highlands, which are closer to coastal towns like Miri. The main Kelabit villages outside of the Highlands are Long Peluan, further south of Pa' Dalih, and Long Lellang, Long Seridan and Long Napir. In particular, Long Lellang Kelabit is known for its distinctive intonational patterns (Beatrice Clayre, p.c.). The analysis in this thesis is based on the dialect of Kelabit spoken in Bario, unless otherwise indicated.⁵³

2.2.3 Ethnography

The Kelabit are traditionally rice farmers and are known for their distinctive system of wet rice cultivation (Saging 1976/1977). This and other traditional practices, including salt-making, traditional arts and crafts, a megalithic tradition and secondary burial practices, distinguish the Kelabit from other groups in Northern Sarawak, such as the Kayan and Kenyah (Saging & Bulan 1989). For further discussion of traditional Kelabit practices and customs, the reader is referred to the detailed ethnographic accounts of Talla (1979), Saging (1976/77), Bala (2002) and Saging & Bulan (1989), written from the perspective of the Kelabit community, as well as several works written by outside academics, including Harrisson (1954, 1959a, 1960), Janowski (1988, 1991, 2003, 2012), Amster (1998, 1999, 2003, 2006), Mashman (2014) and Schneeberger (1979) etc.

In recent years, the community has undergone a number of largescale changes, including conversion to Christianity, outward migration, and urbanisation (Bala 2002, Amster 2006: 208, Lee & Bahrain 1993). From the 1960s, many Kelabit people migrated to urban centres to pursue educational and economic opportunities (Amster

⁵³ Speakers consulted during the fieldwork came from many of the different settlements around Bario, as discussed in APPENDIX 1. It is not clear whether there is variation between the different long houses around Bario or whether dialect levelling has taken place subsequent to resettlement.

2006). Consequently, many Kelabit now live in Miri, Kuching and other cities in Sarawak, Malaysia and beyond. This has had several implications for the Kelabit language, which are discussed in SUBSECTION 2.2.4.

2.2.4 Language Vitality

Lewis et al (2016) classify Kelabit, much like many of the languages of Sarawak, as endangered at the EGIDS level 6b (Threatened).⁵⁴ This entails that Kelabit is still used among those of child-bearing age but that the language is not always transmitted to the next generations. This classification is supported by Rethinasamy et al (2013a) and Martin & Yen (1994), who highlight important differences between Kelabit in the Highlands and Kelabit in town.

It is difficult to calculate the exact number of ethnic Kelabit, and even more difficult to assess the number of speakers within the total population. The total Kelabit population is listed as 5,900 in the Sarawak 2010 census and Mashman (2014) estimates a figure of 6,500 to allow for population growth outside of Sarawak. However, the population living within the Highlands is much smaller. The Bario Clinic listed a population of 1,089 in 2012, including the inhabitants of Bario, Pa' Ukat, Pa' Umur, Pa' Lungan, several Penan settlements and administrative offices. The Orang Ulu National Association estimates that the total Highlands population may be in the region of 1,200 (Rethinasamy 2014).

The Kelabit living in the Highlands tend to use Kelabit as a language of daily communication in various domains, including the home, the village centre, the airport and the church. Intergenerational transmission is high and the majority of children

⁵⁴ See Martin (1995), Clynes (2012) and McLellan (2014) for general overviews of language vitality in Borneo and Sarawak, and David and Dealwis (2006) on Teluga, Dealwis (2008) and Ting & Campbell (2007) on Bidayuh, Cullip (2000) and Coluzzi (2010) on Iban and Lun Bawang and Bibi (2006) on Bintulu.

acquire Kelabit in the home and use Kelabit as the medium of communication within the family (cf. Rethinasamy et al 2013a). Kelabit also serves formal functions at official meetings and traditional ceremonies such as the Naming Ceremony or *Irau Mekaa' Ngadan* (cf. Saging & Bulan 1989). The only domain in which Kelabit is not widely used is in school, where the medium of education is Malay, and in the clinic where Malay is used to communicate with medical staff.⁵⁵ Speakers generally report positive attitudes towards Kelabit as a language of local identity and solidarity (cf. Rethinasamy et al 2013a). Hence, Kelabit is reasonably vital within the Highlands, despite the relatively low number of speakers, according to scalar and multidimensional models of language vitality (cf. Austin & Sallabank 2011).

On the other hand, Martin & Yen (1994) paint a different picture for Kelabit in town. They describe a process of language shift, particularly among younger generations. Though there are many fluent speakers in urban centres like Miri, Martin & Yen (1994) report that intergenerational transmission of Kelabit is declining in town. They suggest that this is influenced by patterns of intermarriage. In families where both parents are Kelabit, they found that Kelabit is used to communicate with children 70% of the time. However, they found that Kelabit is only used 33% of the time by other respondents, who also use Malay, English and other local languages at home. Hence, intergenerational transmission is generally lower in urban centres.

Similarly, the domains of use are also more restricted in town. In particular, the use of Kelabit in religious settings is much less common. Moreover, younger speakers, particularly those who have multi-ethnic friendship groups, will often use English or Malay or code-switch in group settings (Martin & Yen 1994). Even at

⁵⁵ English is also used as a language of wider communication in the many homestays that cater for the growing tourist economy.

home, use of Kelabit differs depending on whether both parents are Kelabit or only one. For 85% of families where both parents are Kelabit, Kelabit is the main means of communication between spouses. In mixed marriages, on the other hand, English is used 65% of the time, and Malay 14% (Martin & Yen 1994: 155).

Nonetheless, there are ongoing community attempts to revitalise Kelabit, including the establishment of Kelabit camps in Miri and a Kelabit play school in Bario (Bulan & Labang 2008). Moreover, Kelabit is increasingly used in so-called new domains, such as on Facebook and Whatsapp, and increased air travel and better roads mean that the villages are now more accessible and more people return to the Highlands in the holidays and for special occasions to visit family and friends. Finally, a Kelabit-medium community radio station has been established in Bario that broadcasts for a few hours in the morning and evening (see Harris & Harris 2011).⁵⁶ Hence, the Kelabit language is something that the community are striving to preserve.

In summary, the vitality of the Kelabit language seems to differ in the Highlands and the town. In the villages, most people speak Kelabit on a regular basis for a wide range of functions. However, given the largescale patterns of migration away from the villages, and the processes of urban language shift, Kelabit can be classed as endangered. It is therefore important to preserve a record of the language and it is hoped that the documentary corpus and preliminary description provided in this chapter will help in this project, as well as in ongoing revitalisation efforts.

2.3 Phonology and Orthography

In this section, I outline basic aspects of Kelabit phonology. The analysis draws on primary linguistic fieldwork, as well as the insights in Asmah (1983), Blust (1974,

⁵⁶ <http://www.ebario.org/radio-bario.html>

1993, 2006) and Labang (2012). In order to interpret the data in the following chapters, it is important to understand the orthographic conventions used and how these relate to the phoneme inventory (see Blust 1993). The phoneme inventory in Kelabit is shown in TABLES 2.1 and 2.2. It is based on an analysis of minimal pairs:

Table 2.1 Vowel Inventory in Kelabit

	Front		Back
High	i		u
Low-Mid	e	ə	ɔ
Low	a		

Table 2.2 Consonant Inventory in Kelabit

	Bilabial	Alveolar	Palatal	Dorso-velar	glottal
Voiceless Plosives	p	t		k	ʔ
Voice Plosives	b	d		g	
Voiced Aspirates	b ^h	d ^h		g ^h	
Africates			ɟʒ		
Nasals	m	n		ŋ	
fricatives		s			h
liquids		l			
trills		r			
glides			j	w	

In general, sounds are represented using the IPA symbols in TABLE 2.1 and 2.2.

The exceptions are listed in TABLE 2.3 and follow Labang (2012):⁵⁷

⁵⁷ Kelabit orthography has yet to be standardised and a number of variant spelling systems exist. This dissertation employs the spelling system in Labang (2012) as it is the most formalised system currently available and is based on a phonological analysis. However, many speakers are strongly opposed to the use of ⟨q⟩ to reflect the glottal stop. The older spelling system, originally used in the Lun Bawang Bible translation and extended to Kelabit, uses an apostrophe for the glottal stop. This is used in many languages of Borneo and is felt to be a distinctive feature of Bornean languages (Poline Bala, p.c.). Hence, it has an identity function. The other major issue is how to reflect the vowel sounds [ə] and [e]. Labang (2012) and this thesis use the symbols ⟨e⟩ and ⟨ey⟩ respectively. However, the use of ⟨ey⟩ can prove difficult to interpret, particularly when followed by a glottal stop, as in the particle *tebeyq*. The old spelling system reserved ⟨e⟩ for [e] and used ⟨a⟩ for schwa, but this creates some confusion with the

Table 2.3 Kelabit Orthography

Phoneme	Orthography
/e/	ey
/ə/	e
/ɔ/	o
/b ^h /	bp
/d ^h /	dt
/g ^h /	gk
/dʒ/	j
/ʔ/	q (word-finally) and ‘ (elsewhere)
/ŋ/	ng
/j/	y

2.3.1 Vowels

Kelabit has six monophthong vowels (see Asmah 1983). Each of these is subject to phonologically-conditioned allophonic behaviour. Firstly, non-schwa vowels are realised as tense or long in open syllables, and lax or short in closed syllables:

(3)	Open-syllable	Closed-syllable
a.	<i>abi</i> [a-bi:] ‘all’	<i>ngabit</i> [ŋa-bit] ‘borrow’
b.	<i>emey</i> [ə-me:] ‘go’	<i>emeyq</i> [ə-mɛʔ] ‘goat’
c.	<i>tudo</i> [tudɔ:] ‘sit’	<i>betɔq</i> [bətɔʔ] ‘yet’/sentence particle
d.	<i>ayu</i> [aju:] ‘likely’	<i>ayuq</i> [ayʊʔ] ‘nature’/emphatic particle

Secondly, vowels are nasalised in the context of nasals and laryngealised in the context of glottal stops (cf. Blust 1974a).⁵⁸

In addition to monophthong vowels, there are several diphthongs. These occur less frequently than monophthongs and could be considered clusters of vowels:

[a] phoneme. Other suggestions include ⟨ae⟩ or ⟨ay⟩ for [e]. The spellings reflect pronunciation in Bario and may not apply for other dialects of Kelabit (see SUBSECTION 2.2.2).

⁵⁸ See Blust (1993, 2006) for discussion of other phonological alternations.

(4) **Diphthongs**

- a. /aɪ/ *main* [mam] ‘tasty, salty’
- b. /aʊ/ *raut* [raʊt] ‘play, joke’
- c. /ɔɪ/ *perekitoi* [pərəkitɔɪ] ‘hang oneself onto something’
- d. /ui/ *selangui* [səlaŋui] ‘snake’
- e. /iu/ *masiu* [masiu] ‘sell’

Finally, there may be a phonemic contrast in vowel length. A few minimal pairs can be found for /a/, /i/ and /u/, in environments where a tense vowel is not expected. Length is contrastive in Sa’ban (Beatrice Clayre p.c.):⁵⁹

(5) **Vowel Length**

- a. *lun* [lɔn] ‘people’
luun [lu:n] ‘on’
- b. *iih* [i:h] ‘who’
ih [ɪh] ‘definite particle’
- c. *maan* [ma:n] ‘collect water’
man [man]⁶⁰ ‘also’

Where vowel length creates a minimal pair, this is distinguished in the orthography with double letters, e.g. *luun* ‘on’ vs *lun* ‘people’.

The vowel phonemes vary in their phonotactic restrictions. The phonemes /i/, /a/ and /u/ are relatively unrestricted and occur word-initially, word-medially and word-finally.⁶¹ In contrast, /e/ and /ɔ/ are more restricted and occur mainly – or perhaps exclusively – in word-final position. They typically occur in open syllables or in closed syllables where the coda is a glottal stop. Moreover, they undergo vowel alternations with /ay/ and /aw/ in the context of suffixation, as discussed in SUBSECTION 2.4.1.3.1. Schwa is also subject to restrictions. It is found word-initially, word-medially and

⁵⁹ There is also the word [dɔ:ʔ] ‘good’ but I have not found a minimal pair. Blust (2006:315) suggests that vowel length may be allophonic.

⁶⁰ Sometimes also pronounced [mən].

⁶¹ Blust (1993) argues that /a/ is not found before /h/. However, the words [laʔah] ‘exceed’ and [naʔah] ‘before’ do exist, though it is also possible that /a/ is reduced to schwa in some dialects.

word-finally in closed syllables (typically with /h/ or /n/ in the coda) but does not occur in final open syllables, initially in pre-penultimate syllables or before a glottal stop or glide (cf. Blust 1993: 146). There is a tendency for all pre-penultimate vowels to be realised as schwa (see SUBSECTION 2.3.5.4).

2.3.2 Consonants

In contrast to the vowel inventory, the consonant inventory in Kelabit has proven relatively controversial, particularly in regards to the so-called voiced aspirates (see Asmah 1983, Blust 1974, 1993, 2006). Blust (2006: 316) suggests that the ‘voiced aspirates’ probably developed from geminate consonants. Nonetheless, he treats the consonants as single phonemes synchronically on the basis that minimal pairs can be found (cf. Blust 1993, 2006):

- (6) **Voiced Aspirates**
- a. **b^h**
tebpuh [təb^ho^h] ‘sugar cane’
tetepuh [tətəp^ho^h] ‘grandfather’
- b. **d^h**
tudtuq [tud^hoʔ] ‘salt’
tutuq [tutoʔ] ‘fall’
tuduq [tudoʔ] ‘seven’
- edten* [əd^hən] ‘UV.IRR.work’
eden [ədən] ‘only’
eten [ətən] ‘instruction’
- c. **g^h**
legkuq [ləg^hoʔ] ‘thunder’
lekuq [lekoʔ] ‘bracelet’
- migkuq* [mig^hoʔ] ‘hit a bruise’
miguq [migoʔ] ‘be shy’

Asmah (1983) suggests that voiced aspirates may be allophones of the voiced phonemes on the basis of phonological alternations (see SUBSECTION 2.3.5). Alternatively, Ladefoged & Maddieson (1996) suggest that they may be clusters of voiced and voiceless phonemes. Blevins (2006) proposes a similar analysis for Ida'an Begak of Eastern Sabah. However, this does not mean that a cluster analysis should necessarily apply to Kelabit.⁶²

Indeed, Blust (2006) argues convincingly that voiced aspirates are not consonant clusters for the following reasons. Firstly, consonant clusters do not occur elsewhere within a single syllable and are quite rare, even across morpheme and syllable boundaries. Secondly, the voiced aspirates alternate with voiced plosives and vice versa, which is difficult to explain if the sound is considered a cluster rather than a single phoneme (SUBSECTION 2.3.5.3). Thirdly, voiceless plosives are unaspirated, whilst the 'voiced aspirates' are aspirated, which would be difficult to explain on a cluster account. Fourthly, /t/ appears to be a dental phoneme, whilst /d/ and the voiced aspirate /d^h/ are alveolar. Finally, the /d^h/ voiced aspirate is realised as the single phoneme [s] in Kelapang Kelabit. Hence, behavioural properties favour a single consonant analysis. Whatever the status of the voiced aspirates, they are represented orthographically in this thesis using the symbols 'bp', 'dt' and 'gk', as is standard practice among the Kelabit community (TABLE 2.3).

As for the other consonants, voiceless plosives are unaspirated in Kelabit, and realised word-finally, in contrast to other Western Austronesian languages, such as Javanese (Hemmings 2012). The phoneme /t/ is dental [t̪] (cf. Blust 2006) and there is a phonemic distinction between /k/ and glottal stop, which is supported by numerous

⁶² In fact, there are some key differences between Ida'an Begak and Kelabit phonology (see Blust 2006). Firstly, Ida'an Begak allows other instances of consonant clusters and, secondly, there is no aspiration on the second consonant in Ida'an Begak, unlike Kelabit (cf. Goudswaard 2005).

minimal pairs, such as *reraq* [rəraʔ] ‘ant’ vs. *rerak* [rərak] ‘torn’. Finally, Blust (2006: 315) suggests that all consonants, including plosives, have lengthened allophones following schwa. This may be because schwa does not have a lengthened allophone in open syllables, and is probably linked to stress (SUBSECTION 2.3.4).

Voiceless plosives are relatively unrestricted in where they occur. However, they generally do not occur in the coda of non-final syllables (see SUBSECTION 2.3.3).⁶³ The voiced aspirates typically occur as the onset to a final syllable, mainly following schwa, whilst the voiced consonants typically do not occur in this context. This complementary distribution has been taken to support an allophonic analysis of the voiced aspirates, though they are not restricted to environments following schwa, as shown in (6). Finally, the glottal stop is phonemic word-medially and word-finally, but does not occur word-initially, except perhaps allophonically before vowel-initial roots (see Blust 2006). There is only one affricate, the voiced palatal [dʒ], which occurs infrequently, and typically as the onset to a final syllable. However, it does occur in kinship terms such as *ejaq* [ədʒaʔ] ‘friend/partner of your partner’s sibling’, which suggests that it is not a borrowed phoneme.

Like in many Western Austronesian languages, nasals play an important morphological role and are subject to homorganic nasal substitution (SUBSECTION 2.3.5.1). Kelabit has three nasal phonemes: bilabial, alveolar and dorso-velar. These are the least restricted of all consonant phonemes and can occur word-initially, word-medially and word-finally but also in the coda of non-final syllables under infixation:

- (7) **Nasal Consonants Word-Medially**
 a. *meta’ut* ‘AV.scare’ + *-in-* → *pinta’ut* [pin.taʔut] ‘UV.PFV.scare’

⁶³ This may be linked to the preference against consonant clusters in Apad Uat languages (cf. Blust 2006). However, consonant clusters do occur in Sa’ban along with other unusual consonants, such as voiceless nasals and geminate consonants word-initially (Clayre 1994).

Kelabit has two fricative phonemes, /s/ and /h/. The alveolar fricative /s/ is relatively unrestricted and occurs word-initially, word-medially and word-finally. However, /h/ only occurs word-finally and is deleted under suffixation (SUBSECTION 2.4.1.3.1). Finally, Kelabit has liquid, trill and glide phonemes. The alveolar liquid /l/ and trill /r/ are unrestricted in distribution. The two glide consonants, /w/ and /j/, tend to occur intervocalically.⁶⁴ They are sometimes added epenthetically between two vowels:

- (8) **Epenthetic Glides**
 a. *uih* [uwih] ‘1SG.1’
 b. *ieh* [ijəh] ‘3SG.1’

It is also possible to think of /i/ and /u/ as semi-vowels that are realised as glides when combined with other vowel phonemes. In this thesis, epenthetic glides are not represented orthographically, as they are predictable. However, some people do use the spellings *uwih* and *iyeh*.

2.3.3 Syllable Structure

Kelabit has a preference for CV and CVC syllable structure, as is common in many Austronesian languages (Blust 2013).⁶⁵ Tense vowels are only found in CVC syllables, as discussed in SUBSECTION 2.3.1.

The following syllable structures are found in root words:

- (9) a. **One-syllable root**
 CV *mey* [me:] ‘go’
 CVC *laq* [laʔ] ‘want’⁶⁶

⁶⁴ Though also occur initially, e.g. in *waluh* ‘eight’.

⁶⁵ CVCVC is reconstructed for the large majority of bases in PAn (Blust 2013: 595).

⁶⁶ VC is also found in functional items, such as the particle *ih*. However, I am not aware of any lexical item consisting only of a vowel + consonant in the coda.

- b. **Two-syllable root**
- | | |
|--------|--------------------------------|
| V.CV | <i>aba</i> [a.ba:] ‘log’ |
| V.CVC | <i>ideh</i> [i.dəh] ‘they’ |
| CV.CV | <i>laba</i> [la.ba:] ‘pass by’ |
| CV.CVC | <i>manuk</i> [ma.nək] ‘bird’ |

It appears that most lexical roots are bisyllabic. Monosyllabic words are typically grammaticalised particles and functional items, such as *laq* and *mey* in (9), and sometimes have bisyllabic variant pronunciations, such as *emey* and *elaq* (see SUBSECTION 2.4.2.6).⁶⁷ There are also some nominal items that are monosyllabic such as *war* ‘root’ and *wey* ‘rattan’.⁶⁸

There are a few roots that consist of more than two syllables:

- (10) **Three-syllable roots**
- | | | |
|----|----------------|--|
| a. | borrowed terms | <i>sekolah</i> [sə.kə.lah] ‘school’ (from Malay) |
| | | CV.CV.CVC |
| b. | body parts | <i>segerang</i> [sə.gə.raŋ] ‘rib’ |
| | | CV.CV.CVC |
| | | <i>demawid</i> [də.ma.wid] ‘pancreas’ |
| | | CV.CV.CVC |

However, such roots are rare. Most multisyllabic words are formed via morphological processes of prefixation, infixation and suffixation, which are discussed in SUBSECTION 2.4.1.

There is a strong preference for final syllables to be heavy or bimoraic. This is achieved either by a syllable coda or a lengthened vowel, as in *mey* [me:] and *wey* [we:]. There do not seem to be any one-syllable words formed exclusively from a vowel, lengthened or otherwise. In words of two or more syllables, the final syllable

⁶⁷ The elision of the initial schwa is emblematic of a process of grammaticalisation (see Heine & Kuteva 2007). The grammatical function of the forms *mey* and *laq* is discussed in SUBSECTION 2.4.2.6.

⁶⁸ These could be considered vowel-initial words where the first vowel /u/ is realised as a glide [w].

preferentially begins with a consonantal onset. If the bisyllabic word is formed from a vowel-initial root, an epenthetic glottal stop or glide is added. In all non-final syllables, there is a preference for open syllables.

2.3.4 Prosody and Stress

Both Asmah (1983: 551) and Blust (2006: 315) suggest that stress falls on the final syllable of the word in citation form, though this may vary in speech. Blust (1974a) initially noted a preference for stress on the penultimate syllable and suggests that stress patterns may have changed since his earlier work.⁶⁹ He suggests that a distinction between lexical stress and phrasal stress may constitute an areal feature in North Sarawak (Blust 2006: 315). Stress-shift appears to occur under suffixation (SUBSECTION 2.4.1.3). Both syllable structure and stress play a role in phonological processes, which is discussed in SUBSECTION 2.3.5.

2.3.5 Phonological processes

Several phonological processes apply in Kelabit and are further discussed in regards to word formation in SUBSECTION 2.4.1. In the following sections, I illustrate nasal assimilation and substitution, diphthongisation, consonant gemination and vowel reduction and deletion.

2.3.5.1 Nasal Assimilation and Substitution

Nasal assimilation and substitution are found in many Western Austronesian languages (cf. Davies 2010, Blust 2004, 2013). Nasal substitution is a process whereby

⁶⁹ It is possible that roots differ in their stress assignment – some with stress on the final syllable and some with stress on the penultimate syllable. This is reconstructed for Proto-Austronesian (see Ross 2002).

a nasal fully assimilates to the place of articulation of the initial consonant of the root to which it attaches and then replaces that consonant. In Kelabit, this occurs in the context of the actor voice nasal prefix (SUBSECTION 2.4.1.1.4).

- (11) **Nasal Substitution**
- | | | |
|----|-------------------------------|--|
| a. | N → [m] / _ [bilabial] | N- + <i>puwer</i> → <i>muwer</i> ‘AV.butcher’
N- + <i>bilaq</i> → <i>milaq</i> ‘AV.break’ |
| b. | N → [n] / _ [dental/alveolar] | N- + <i>terad</i> → <i>nerad</i> ‘AV.cut’
N- + <i>dinger</i> → <i>ninger</i> ‘AV.hear’
N- + <i>si’er</i> → <i>ni’er</i> ‘AV.see’ |
| c. | N → [ŋ] / _ [velar] | N- + <i>kiding</i> → <i>ngiding</i> ‘AV.lift’
N- + <i>gegkang</i> → <i>ngegkang</i>
‘AV.lift.up/fire’ |

The initial consonant is substituted regardless of whether that consonant is plosive or fricative, voiced or voiceless. This is similar to Madurese (Davies 2010). In other Western Austronesian languages, including Indonesian, Javanese, Balinese, Sundanese and Batak, only voiceless consonants are replaced. For roots beginning with voiced consonants, the nasal assimilates to the place of articulation, but does not replace the root consonant, resulting in prenasalised consonants or nasal clusters (cf. Davies 2010: 47).

Nasal assimilation, in contrast to nasal substitution, involves the assimilation of the nasal to the place of articulation of the following consonant without replacing the consonant. This applies mainly in the context of *-in-* infixation in complex stems formed with *pe-* and *te₁-* (SUBSECTION 2.4.1.1.7 and 2.4.1.1.13):

- (12) **Nasal Assimilation**
- | | | | |
|----|--------------------|--|--|
| a. | <i>-in-</i> → [ɪm] | AV form
[nəbukʊh]
N- + <i>te-</i> + <i>bukuh</i>
‘AV.knot’ | UV form
[sɪmbukʊh]
<i>te-</i> + <i>-in-</i> + <i>bukuh</i>
‘UV.knot’ |
|----|--------------------|--|--|

- | | | | |
|----|--------------------|--|--|
| b. | <i>-in-</i> → [ɪn] | [mətəlaq]
N- + <i>pe-</i> + <i>telaq</i>
'AV.throw.away' | [pɪntəlaq]
<i>pe-</i> + <i>-in-</i> + <i>telaq</i>
'UV.throw.away' |
| c. | <i>-in-</i> → [ɪŋ] | [nəgaʊ]
N- + <i>te-</i> + <i>gao</i>
'AV.unsettle' | [sɪŋgəʊ]
<i>te-</i> + <i>-in-</i> + <i>gao</i>
'UV.unsettle' |

Hence, nasal assimilation and substitution occur in the context of prefixation and infixation.

2.3.5.2 Diphthongisation

Another form of assimilation occurs in the context of the specificity particle *ih* (SUBSECTION 2.4.2.1). When the particle follows words ending in vowels – or V[h] as [h] is elided intervocally – the two vowels diphthongise under co-articulation:

- (13) **Diphthongisation**
- | | | |
|----|-------------------------|--|
| a. | [a] → [aɪ]/ _ <i>ih</i> | <i>suk na'ah ih</i> [sʊk naʔaɪ] 'the aforementioned' |
| b. | [ə] → [eɪ]/ _ <i>ih</i> | <i>teh midteh ih</i> [təmid ^h eɪ] 'sometimes' |

This is a regular process and examples (13a) and (13b) are interpreted as single words, rather than phrases. Nonetheless, they are written separately in this thesis, as in (13), in order to preserve the morphological connection with *na'ah* 'earlier' and *midteh* 'once'. The relativiser *suk* is discussed in SUBSECTION 2.4.2.10. *Teh* is sometimes analysed as a prefix *te-* and sometimes as a particle. See SUBSECTIONS 2.4.1.1.13, 2.4.1.3.3 and 2.4.2.14 for discussion.

The particle *ih* may derive from the medial demonstrative *dih* (SUBSECTION 2.4.2.7). This can be used wherever *ih* is used, with similar lexical meaning. However, no diphthongisation occurs when *dih* is used in place of *ih*.

- (14) **No Diphthongisation**
- a. *suk na'ah dih* [sʊk naʔah dɪh] ‘the aforementioned’
 - b. *teh midteh dih* [təmid^heh dɪh] ‘sometimes’

It may be that *ih* is a clitic and that diphthongisation occurs within the phonological word, rather than across word boundaries (see SUBSECTION 4.3). Alternatively, the intervening consonant blocks diphthongisation of the vowels.

2.3.5.3 Consonant Gemination

Consonant gemination or lengthening applies to consonants in the onset of the final syllable following schwa. It mainly occurs under *-en* suffixation (SUBSECTION 2.4.1.3.1). When the final consonant is voiced, the geminate allomorph is a voiced aspirate:

- (15) **Gemination of Voiced Stops after schwa**
- a. /b/ → [b^h] *rereb* → *rerebpen* [rərəb:^hən] ‘UV.IRR.baptise’
eseb → *sebpen* [səb:^hən] ‘UV.IRR.burn’
kekeb → *kekebpen* [kəkəb:^hən] ‘UV.IRR.cover’
 - b. /d/ → [d^h] *tuked* → *tekedten* [təkəd:^hən] ‘UV.IRR.put.at.angle’
lened → *lenedten* [lənəd:^hən] ‘UV.IRR.cook
(vegetables)’
 - c. /g/ → [g^h] *eleg* → *legken* [ləg:^hən] ‘UV.IRR.divorce’

If the final consonant is a voiceless stop, the consonant is lengthened following schwa, and fills both the coda of the penultimate syllable and the onset of the final syllable:

- (16) **Gemination of Voiceless Stops following schwa**
- a. /p/ → [p:] *rerep* → *rerepen* [rə.rəp:ən] ‘UV.IRR.lower’
 - b. /t/ → [t:] *bebpet* → *bebeten* [bə.bət:ən] ‘UV.IRR.hit’
 - c. /k/ → [k:] *tetek* → *teteken* [tə.tək:ən] ‘UV.IRR.chop’

This process appears to occur in order that penultimate syllables are bi-moraic, and final syllables have a filled onset. This may suggest that both the penultimate and final syllables bear stress under suffixation, whilst other syllables are unstressed. Indeed, Blust (2006) argues that stress shifts under suffixation.

Consonants are not geminated following non-schwa vowels. Instead, the vowel is lengthened and the consonant articulated in the onset of the following syllable:

- (17) **No Gemination following non-schwa vowels**
- a. *ukab* [u.kab] → *kaben* [ka:.bən] ‘UV.IRR.open’⁷⁰
 - b. *terad* [tər:ad] → *teraden* [tə.ra:.dən] ‘UV.IRR.cut’
 - c. *palug* [pa.lɔg] → *pelugan* [pə.lu:.gan] ‘UV.IRR.trick’

Hence, consonant gemination and vowel lengthening may both be triggered by stress.

2.3.5.4 Vowel Reduction

A related process is the reduction of all pre-penultimate vowels to schwa when they occur in unstressed syllables. Reduction typically occurs under suffixation and infixation:

- (18) **Vowel Reduction**
- a. Infixation *si'er* + *-in-* → *seni'er* [sə.ni.ʔər] ‘UV.PFV.see’
tatek + *-in-* → *senatek* [sə.na.tək] ‘UV.PFV.close’
 - b. Suffixation *badaq* + *-en* → *beda'an* [bə.da.ʔan] ‘UV.IRR.show’
pudut + *-en* → *peduten* [pə.du.tən] ‘UV.IRR.build’

⁷⁰ The form sounds correct to my primary consultant but it is not in regular use.

Vowel reduction sometimes also occurs in words like *ngimalem* ‘yesterday’, which is formed by the combination of the preposition *ngi* ‘at’ and *malem* ‘past’. The /i/ is variously pronounced [i] and [ə].⁷¹ In the orthographic representation in the thesis, ‘e’ is used in pre-penultimate syllables of UV perfective forms, whilst ‘i’ is used in words like *ngimalem*.

2.3.5.5 Vowel Deletion

Finally, vowels that comprise the initial syllable of a bi-syllabic root are deleted under suffixation:

- (19) **Vowel Deletion**
- a. *irup* ‘drink’ → *rup-en* ‘UV.IRR.drink’
 - b. *itun* ‘question’ → *tun-en* ‘UV.IRR.question’
 - c. *eseb* ‘burn’ → *sebp-en* ‘UV.IRR.burn’
 - d. *uput* ‘jump’ → *put-an* ‘UV.IRR.jump’

This seems to reflect the fact that pre-penultimate syllables are never composed of a vowel alone (SUBSECTION 2.3.3).

2.3.6 Summary

In summary, if we adopt Blust’s (1993, 2006) proposal that the voiced aspirates are phonemes, then Kelabit has 20 consonant phonemes and 6 vowel phonemes. The main phonemes that present difficulties for orthographic representation are the glottal stop and the schwa vowel. In this thesis, the glottal stop is represented using ⟨q⟩ word-finally and ⟨ʔ⟩ word-medially, the schwa is represented using ⟨e⟩ and [e] is represented using ⟨ey⟩, following Labang (2012). The phonemes are subject to allophonic variation in environments conditioned by co-articulation and syllable structure.

⁷¹ This also true of the *-in-* infix.

Kelabit syllables tend to be CV in non-final positions. However, there is a strong preference for heavy syllables in word-final position. This is achieved via syllables with long vowels (CV_{long}) or consonants in the syllable coda (CVC). Consonant clusters are highly disfavoured and occur mainly across morpheme and syllable boundaries. Lexical roots tend to be bi-syllabic, though functional roots can be mono-syllabic and stress is on the final syllable. Finally, there are a number of phonological processes that occur in Kelabit. These include nasal assimilation, diphthongisation, consonant germination, vowel reduction and vowel deletion.

2.4 Morphology

Like many other Western Austronesian languages, Kelabit is essentially an agglutinative language (Asmah 1983). This means that words often consist of more than one morpheme, but the boundary between morphemes is clear (Aikhenvald 2007). Nonetheless, voice morphemes show some fusional properties, combining voice and aspectual/modal features (SUBSECTION 2.4.1). For example, the *-in-* infix not only conveys undergoer voice but also realis mood/perfective aspect (SUBSECTION 2.4.1.2.3). Moreover, some forms are multifunctional. For example, *pe-* can have both a stem-forming function and a reciprocal interpretation, among others (SUBSECTION 2.4.1.1.7). Thus, morphemes differ in their degree of fusion, which is relatively common in agglutinative languages (Hagège 1990).

The section is structured as follows. SUBSECTION 2.4.1 describes word-formation processes including derivation and reduplication and SUBSECTION 2.4.2 discusses the major word-classes.

2.4.1 Word Formation

The most wide-spread word-formation strategies in Kelabit are derivation and reduplication, as is common in the languages of Borneo (cf. Soriente 2013).⁷² Derivational processes include prefixation, infixation and suffixation.⁷³ Some processes are highly productive and others are less so. Typically, words are formed of a root and a single affix and it is not common to find more than three affixes attached to a given root.⁷⁴ Most affixes can attach to roots of different word classes and derive different word classes depending on the root to which they attach.

2.4.1.1 Prefixation

Several prefixes, in particular *pe-*, *ke-* and *se-* seem to be multifunctional. It is not clear if these are polysemous or homophonous morphemes (Hemmings 2013). However, multifunctionality is found in many of the languages of Middle Borneo (Soriente 2013). Common prefixes are listed in TABLE 2.4 and will be discussed in turn.

⁷² Whether strategies such as compounding and incorporation are also used remains for further research. These do not appear to be as frequent as derivation and reduplication in any case.

⁷³ There is one case of a discontinuous affix, namely *pe- -en* (see SUBSECTION 2.4.1.3.1). It is sometimes argued that the voice affixes should be treated as inflection rather than derivation on account of their relative productivity and the fact that they form paradigms. However, the distinction between inflection and derivation is not clear-cut in Austronesian (see Hurlbut 1988 and Starosta 2009d for discussion).

⁷⁴ An example of a word with three affixes is *nenepu'un* (*ne-* + *N-* + *te-* + *pu'un*). Note that *neN-* could be treated as a single affix (SUBSECTION 2.4.1.1.5).

Table 2.4 Prefixes in Kelabit

Morpheme	Function	Root/Stem	Prefixed Form
<i>deN-</i>	kinship relationship	<i>anak</i> ‘child’	<i>denganak</i> ‘sibling’
<i>ke₁-</i>	abilitative	<i>kiding</i> ‘lift’	<i>kekiding</i> ‘be able to lift’
<i>ke₂-</i>	ordinal numbers	<i>limeh</i> ‘five’	<i>kelimeh</i> ‘fifth’
<i>N-</i>	AV irrealis	<i>bilaq</i> ‘broken’	<i>milaq</i> ‘AV.break sth’
	unergative intransitive	<i>dalan</i> ‘path’	<i>nalán</i> ‘walk’
<i>ne₁-</i>	perfective	<i>nerem</i> ‘AV.sink’	<i>nenerem</i> ‘sank’
<i>ne₂-</i>	accidental	<i>terem</i> ‘sink’	<i>neterem</i> ‘accidentally sink sth’
<i>pe-</i>	reciprocal	<i>keliq</i> ‘know’	<i>pekeliq</i> ‘know each other’
	causative	<i>rudap</i> ‘sleep’	<i>merudap</i> ‘AV.put to sleep’
<i>peN-</i>	plural actor	<i>lubid</i> ‘roll’	<i>pelubid</i> ‘all roll’
	IV irrealis	<i>tekul</i> ‘spoon’	<i>penekul</i> ‘use to spoon’
	instrumental	<i>tatek</i> ‘close’	<i>penatek</i> ‘door stop’
	nominalisation		
<i>peneN-</i>	IV perfective	<i>penekul</i> ‘IV.spoon’	<i>penenekul</i> ‘IV.PFV.spoon’
<i>pere-</i>	reflexive	<i>ngapung</i> ‘hide’	<i>perengapung</i> ‘hide oneself’
<i>se-</i>	non-serious action	<i>riruh</i> ‘laugh’	<i>seriruh</i> ‘pretend to laugh’
<i>seN-</i>	middle voice	<i>anuk</i> ‘dress’	<i>sanuk</i> ‘get dressed’
	UV perfective prefix (non-standard)	<i>ngelaak</i> ‘AV.cook’	<i>sengelaak</i> ‘UV.PFV.cook’
<i>te₁-</i>	stative	<i>ruyuh</i> ‘sway’	<i>teruyuh</i> ‘swaying’
<i>te₂-</i>	distributive numbers	<i>limeh</i> ‘five’	<i>telimeh</i> ‘five by five’

2.4.1.1.1 *deN-*

This prefix describes reciprocal relationships and is typically attached to kinship terms:

- (20) ***DeN-* Prefixation**
- anak* ‘child’ → *denganak* ‘siblings’
 - kanid* ‘cousin’ → *dengekanid* ‘cousins’
 - rumaq* ‘house’ → *dengerumaq* ‘spouses’
 - ruyung* ‘together’ → *dengeruyung* ‘family’

As shown in (20), *deN-* does not undergo the same process of nasal substitution as *N-* (see SUBSECTION 2.3.5.1 and 2.4.1.1.4). When attached to vowel-initial roots the allomorph *deng-* is used, when attached to consonant-initial roots the allomorph *denge-* is used regardless of whether the consonant is obstruent or approximant.

The terms are typically used with inclusory pronouns to specify the relationship between groups (SUBSECTION 2.4.2.8.4):

- (21) **Function of *deN-***
- a. Mulaq men kekamih, kekamih dengeruyung bah!
 many PT 1PL.EXCL.EMPH 1PL.EXCL.EMPH of.one.family PT
 ‘There are a lot of us you know!’
 (text, BAR22102013CH_04 00:03:41.530-00:03:44.940)

2.4.1.1.2 *ke₁-*

The primary function of *ke-* is to derive an abilitative interpretation when attached to bare verbal and nominal roots. It is fairly productive and could be grammaticalised from the pre-verbal auxiliary *kereb* ‘able’ (SUBSECTION 2.4.2.6):

- (22) ***Ke-* Abilitative**
- a. *kiding* ‘lift’ → *kekiding* ‘be able to lift’
 b. *terem* ‘sink’ → *keterem* ‘be able to sink’
 c. *atey* ‘liver/death’ → *kekatey* ‘be able to kill’
 d. *itun* ‘question’ → *kekitun* ‘be able to ask’
 e. *eseb* ‘burn’ → *kekeseb* ‘be able to burn’

As illustrated in (22), when the root begins with a vowel, the allomorph *kek-* is used.

Abilitative *ke-* verbs express their actors using FORM 2 pronouns, which are elsewhere used for actor non-subjects (SUBSECTION 2.4.2.8). This is also true of accidental predicates (SUBSECTION 2.4.1.1.6) and predicates with an experiencer subject (see CHAPTER 4):

- (23) **Function of *ke-***
- a. Am kek-itun kuh kapeh taruq dih ngeneh.⁷⁵
 NEG ABIL-question 1SG.2 how do DEM from.3SG.2
 ‘I wasn’t able to ask him how to do it.’ (elicitation, fieldnotes)
- b. Ken ke-terem muh ieh?
 Q ABIL-sink 2SG.2 3SG.1
 ‘Are you able to sink him?’ (elicitation, fieldnotes)

There are also *ke-* verbs which have a lexicalised meaning:

- (24) ***Ke-* with experiential predicates**
- a. *keliq* ‘know/thought’ → *kekeliq* ‘understand’
 b. *iti* ‘identification’ → *kekiti* ‘recognise’
 c. *sekenan* ‘remember/memory’ → *kesikenan* ‘remember’
 d. *daluh* ‘anger’ → *kedaluh* ‘quarrel’

In (24), the prefixed forms convey that the subject experiences or possesses the quality of the root.

2.4.1.1.3 *ke*₂-

In a number of Austronesian languages, *ke-* and cognate prefixes are multifunctional (cf. Gil 2014). In Kelabit, *ke-* also derives ordinal numbers from the set of cardinal numbers. The allomorph *k-* is used when numerals begin with a schwa. Other vowel initial roots add an epenthetic glottal stop. This is illustrated in TABLE 2.5.

⁷⁵ Kelabit has two sets of pronouns, FORM 1 and FORM 2 (see SUBSECTION 2.4.2.8). The glossing conventions are discussed in the section on abbreviations and conventions.

Table 2.5 Ordinal Numerals in Kelabit

	Cardinal	Ordinal	Meaning
1	edteh	pu'un-pu'un	first
2	duweh	keduweh	second
3	teluh	keteluh	third
4	epat	kepat	fourth
5	limeh	kelimeh	fifth
6	enem	kenem	sixth
7	tuduq	ketuduq	seventh
8	waluh	kewaluh	eighth
9	iwak	ke'iwak	ninth
10	puluq	kepuluq	tenth

Apart from *pu'un-pu'un* 'first' and *peped* 'last', all ordinal numbers are derived via *ke*₂- prefixation. This is treated as a separate affix from *ke*₁-, since abilitative/ordinal polysemy is uncommon.

2.4.1.1.4 N-

Probably the most productive of all prefixes in Kelabit is the nasal prefix. The main function is to mark actor voice (AV). In addition, *N-* also derives intransitive verbal predicates and participates in a causative alternation. The nasal prefix can attach to roots of any word-class:

- (25) **Prefixation to Various Word-Classes**
- a. noun root *abet* 'tie (n)' → *ngabet* 'AV.tie'
 - b. adjective root *rayeh* 'big (adj)' → *ngerayeh* 'AV.celebrate'
 - c. verb root *terem* 'sink (v)' → *nerem* 'AV.sink sth/sb'

There are three allomorphs of the *N-* prefix, depending on the initial sound of the root that it attaches to: nasal substitution, *nge-* prefixation and *ng-* prefixation (cf. Blust 1977).

Nasal substitution occurs for roots beginning with an obstruent consonant. For verb stems beginning with a stop consonant (both voiced and voiceless), homorganic

nasal substitution occurs. For roots beginning with /s/, the nearest nasal consonant /n/ is substituted. This was illustrated in SUBSECTION 2.3.5.1 and is repeated in (26):

- (26) **AV homorganic nasal substitution with obstruent-initial roots**
- a. *N-* + *puwer* → *muwer* ‘AV.butcher’
 - b. *N-* + *bilaq* → *milaq* ‘AV.break’
 - c. *N-* + *terad* → *nerad* ‘AV.cut’
 - d. *N-* + *dinger* → *ninger* ‘AV.hear’
 - e. *N-* + *si’er* → *ni’er* ‘AV.see’
 - f. *N-* + *kiding* → *ngiding* ‘AV.lift’
 - g. *N-* + *gegkang* → *ngegkang* ‘AV.lift.up/fire’

In all cases, nasal substitution prevents consonants clusters word-initially.

For roots beginning with approximant consonants, the prefix *nge-* is added. This can be considered *ng-* with an epenthetic schwa, added due to the constraint against consonant clusters word-initially (SUBSECTION 2.3.2):

- (27) **AV nasal prefixation with approximant-initial roots**
- a. *N-* + *linuh* → *ngelinuh* ‘AV.think’
 - b. *N-* + *raruh* → *ngeraruh* ‘AV.lose’

For roots beginning with a vowel, the prefix *ng-* is added without the epenthetic schwa:

- (28) **AV nasal prefixation with vowel-initial roots⁷⁶**
- a. *N-* + *itun* → *ngitun* ‘AV.question’
 - b. *N-* + *emung* → *ngemung* ‘AV.collect’
 - c. *N-* + *aweh* → *ngaweh* ‘AV.marry’
 - d. *N-* + *udud* → *ngudud* ‘AV.comb’

It is possible to analyse the prefix as underlying *nge-* with vowel elision when attached to vowel-initial roots. However, Blust (1977) argues that an analysis of the prefix as

⁷⁶ As discussed in SUBSECTION 2.3.1, I do not know of any roots beginning with [e] or [ɔ] – I assume that they would follow the same pattern if such roots existed.

underlying *ng-* is simpler, since it would be difficult to explain patterns of nasal substitution, which also prevent consonant clusters initially, if the prefix were *nge-*. Indeed, where *nge-* is used as a linker, nasal substitution does not occur with obstruent-initial roots (see SUBSECTION 2.4.2.12 on numerals). Thus, it is simpler to assume that an epenthetic schwa is used as an alternative strategy to avoid clusters, wherever nasal substitution is not possible.

In some cases, the nasal prefix can also derive unergative intransitive predicates from non-verbal roots:

- (29) **Deriving intransitive predicates with *N-***
- a. *dalan* ‘path’ → *nalan* ‘walk’
 - b. *arang* ‘dance’ → *ngarang* ‘to dance’
 - c. *utaq* ‘vomit’ → *ngutaq* ‘to vomit’

This is common in many Western Austronesian languages, including Tagalog (Aldridge 2012), Madurese (Davies 2005: 203), Balinese (Arka 1998), Indonesian (Sneddon 1996) and Javanese (Hemmings 2012).

Finally, for unaccusative intransitive predicates, and roots that take the *-em-* infix (SUBSECTION 2.4.1.2.1), the nasal prefix can have a causative function:

- (30) **Causative Alternation with *-em-* verbs**
- a. *matey* ‘die.INTR’ → *ngatey* ‘kill’
 - b. *meseb* ‘burn.INTR’ → *ngeseb* ‘burn.TR’
- (31) **Causative Alternation with unaccusatives**
- a. *tudo* ‘sit.INTR’ → *nudo* ‘seat.TR’
 - b. *terem* ‘sink.INTR’ → *nerem* ‘sink.TR’
 - c. *bilaq* ‘break.INTR’ → *milaq* ‘break.TR’

The morphosyntax of these alternations is discussed in SUBSECTION 2.4.2.2. The nasal prefix combines with stem-forming prefixes such as *pe-* (SUBSECTION 2.4.1.1.7) and *te₁-* (SUBSECTION 2.4.1.1.13).

2.4.1.1.5 *ne₁-*

The prefix *ne₁-* attaches to both AV verb forms and bare intransitive/transitive roots to indicate realis mood/perfective aspect:⁷⁷

- (32) **Deriving Perfective with *ne₁-***
- a. transitive root *nerad* ‘AV.cut.TR’ → *nenerad* ‘PFV.AV.cut.TR’
 - b. intransitive root *bilaq* ‘break.INTR’ → *nebilaq* ‘PFV.break.INTR’
 - c. bare transitive root *keliq* ‘know/see’ → *nekeliq* ‘PFV.know/see’

This is a highly productive process and typically corresponds to a past tense interpretation:

- (33) **Function of *ne₁-***
- a. Uih nelaq edteh utung kayuh.
1SG.1 AV.throw one stick wood
‘I throw a stick.’

 - b. Uih ne-nelaq edteh utung kayuh.
1SG.1 PFV-AV.throw one stick wood
‘I threw a stick.’
- (elicitation, fieldnotes)

It combines with the AV verb form but not with the UV infix *-in-*, which is a portmanteau morpheme representing voice and perfective aspect (SUBSECTION 2.4.1.2.3). Hence, the combination *neN-* (*ne-* + *N-*) could be considered to have a paradigmatic relationship with other voice markers (see SUBSECTION 3.5).

⁷⁷ *ne-* was traditionally treated as an auxiliary and written like the particle *neh* (see Asmah 1983: 563). However, unlike other auxiliaries in SUBSECTION 2.4.2.6, it always appears directly before the verb and no other material can intervene. Hence, it is treated as a prefix in this thesis.

2.4.1.1.6 *ne*₂-

The prefix *ne*₂- has an accidental interpretation and attaches to bare roots. In some cases, such as *nebilaq* in (34), this creates ambiguity between a perfective intransitive predicate and an accidental transitive predicate:

- (34) **Accidental Interpretations**
- a. *bilAQ* ‘break.INTR’ → *nebilaq* ‘accidentally break.TR’
 - b. *terAD* ‘cut.INTR’ → *neterAD* ‘accidentally cut.TR’
 - c. *tatek* ‘close.INTR’ → *netatek* ‘accidentally close.TR’

The function of *ne*₂- contrasts with *ne*₁-, which tends to be used in contexts where the action was completed on purpose:

- (35) **Function of Accidental *ne*₂-**
- a. Ne-terAD kuh berek ih.
ACCID-cut 1SG.2 pig PT
‘I accidentally cut the pig.’
(elicitation, BAR21102013CH_01 01:19:39.377-01:19:42.040)
 - b. Uih ne-nerAD berek ih.
1SG.1 PFV.AV.cut pig PT
‘I cut up the pig (on purpose).’
(elicitation, BAR21102013CH_01 01:33:38.319-01:33:40.819)

Moreover, the accidental construction is syntactically transitive, whereas perfective *ne*₁- can derive intransitive predicates when attached to intransitive verbal roots (SUBSECTION 2.4.1.1.5). Similarly, as the actor of (35a) lacks volition, the FORM 2 pronoun *kuh* is used (SUBSECTION 2.4.2.8). Finally, both perfective *ne*₁- and accidental *ne*₂- can co-occur:

- (36) **Co-occurrence of *ne*- prefixes**
- a. Ne-ne-bilaq kuh bigan ih.
PFV-ACCID-break 1SG.2 plate PT
‘I accidentally broke the plate.’ (elicitation, fieldnotes)

For this reason, *ne*₁- and *ne*₂- are treated as separate prefixes. Accidental prefixes are found in other languages in Borneo, including Lundayeh (Clayre 2002), Kimaragang (Kroeger 1990) and Timugon (Prentice 1995: 390-391).

2.4.1.1.7 *pe*-

The prefix *pe*- is highly multifunctional. The most productive of these functions is to mark a reciprocal construction when attached to a bare root:

- (37) **Reciprocal Interpretation**
- a. *rengat* ‘scream’ → *perengat* ‘scream at each other’
 - b. *rekem* ‘claw’ → *perekem* ‘claw at each other’
 - c. *repet* ‘hope’ → *perepet* ‘place hope in each other’
 - d. *kedaluh* ‘fight’ → *pekedaluh* ‘fight each other’
 - e. *bu’uh* ‘be angry’ → *pebu’uh* ‘be angry at each other’
 - f. *tabiq* ‘greeting’ → *petabiq* ‘greet each other/shake hands’
 - g. *imet* ‘hold/grasp’ → *pimet* ‘hold each other/hold hands’
 - h. *apuuq* ‘meeting point/act of meeting’ → *papuuq* ‘meet each other’⁷⁸

Like abilitative *ke*- and accidental *ne*-, reciprocal *pe*- verbs can be syntactically transitive:

- (38) **Function of reciprocal *pe*-**
- a. Pe-pering diweh ebpuk diweh dih.
 RECP-dry 3DU hair 3DU DEM
 ‘They dried each other’s hair.’ (elicitation, fieldnotes)

A second function of *pe*- is as a stem-forming prefix, much like the Philippine formative *pa*- (Rubino 2005: 342). It tends to derive verbs from nominal roots (cf. Asmah 1983):

⁷⁸ The forms *pepapuuq* and *pepimet* are also heard, particularly among younger speakers. This may be in analogy with *ke*- and *kek*- (see SUBSECTION 2.4.1.1.2).

- (39) **Deriving verbal stems from nouns**
- a. *idang* ‘suns rays’ → *pidang* ‘dry in sun (root)’
 - b. *karuh* ‘language’ → *pekaruh* ‘talk’⁷⁹
 - c. *deket* ‘act of sticking’ → *pedeket* ‘be stuck’

Typically, the resulting verbs are intransitive with inchoative aspect. The *pe-* prefix also combines with voice markers, such as *N-* (SUBSECTION 2.4.1.1.4) and *-in-* (SUBSECTION 2.4.1.2.3) to derive transitive AV and UV verbs:

- (40) **Causatives with *pe-***
- a. *peta’ut* ‘scared’ → *meta’ut* ‘AV.scare’ → *pinta’ut* ‘UV.PFV.scare’
 - b. *petulu* ‘meet’ → *metulu* ‘AV.introduce’ → *pintulu* ‘UV.PFV.introduce’
 - c. *pedeket* ‘stuck’ → *medeket* ‘AV.stick’ → *pindeket* ‘UV.PFV.stick’

These have a causative interpretation and hence *me-* is sometimes listed as a causative prefix (cf. Asmah 1983). Typical causatives with *pe-* and voice morphology are illustrated in (41):

- (41) *Function of pe- causative*
- a. **Actor Voice**

Nih	tesineh nedih	me-rudap	anak nedih.
DEM	mother 3SG.POSS	AV.CAUS-sleep	child 3SG.POSS

‘The mother is putting her child to sleep.’
 (elicitation, BAR17102013CH_01 00:37:54.959-00:37:58.797)
 - b. **Undergoer Voice**

P<in>taso	kuh	anak sidih ngen nuk	belaan kuh.
CAUS<UV.PFV>distract	1SG.2	child DEM with REL	say.UV.IRR 1SG.2

‘I entertained that child with my words.’ (elicitation, fieldnotes)

This is not a productive process, and mainly applies to stative roots, or roots encoding unbounded events. Forming causatives with *pa-* is attested in a large number of Austronesian languages (Himmelman 2005a: 170, Blust 2003).

⁷⁹ This could also be thought to have a reciprocal meaning.

A third function of *pe-* is marking locatives, when attached to certain roots:

- (42) **Deriving a position/location**
- | | | | |
|----|-------------------------------------|---|--|
| a. | <i>mudtih</i> ‘last’ | → | <i>pemudtih</i> ‘behind’ |
| b. | <i>ma’un</i> ‘old/first’ | → | <i>pema’un</i> ‘in front’ |
| c. | <i>senu’eh</i> ‘right’ | → | <i>pesenu’eh</i> ‘right side’ |
| d. | <i>kabing</i> ‘left’ | → | <i>pekabing</i> ‘left side’ |
| e. | <i>iring</i> ‘next to (prep)’ | → | <i>pesiring</i> ‘at the side’ |
| f. | <i>ditaq</i> ‘high/tall’ | → | <i>peditaq</i> ‘above’ |
| g. | <i>liang</i> ‘underneath’ | → | <i>peliang</i> ‘below’ |
| h. | <i>lai</i> ‘outside’ | → | <i>pelai</i> ‘outside of’ |
| i. | <i>dingi</i> ‘over there’ | → | <i>pedingi</i> ‘over there’ |
| j. | <i>luun</i> ‘on top’ | → | <i>peluun</i> ‘on the top of’ |
| k. | <i>lalad</i> ‘abreast/side by side’ | → | <i>pelalad</i> ‘side by side/abreast of’ |

This may constitute a separate prefix from the other instances of *pe-*, as it is unclear how the meanings are related. Moreover, unlike the *p-* allomorph attached to vowel-initial roots above, in (42e) an epenthetic consonant is added between the *pe-* prefix and the vowel-initial root *iring* ‘next to’. It is not clear why this is an [s].

The function of these derived forms is to specify a location or position:

- (43) **Function of *pe-* marking position/location**
- | | | | | | | | |
|----|------|-----|-------|----------|-----|--------------|-------|
| a. | Let | lem | edteh | patiq | nuk | sidteh | deh |
| | from | in | one | suitcase | REL | UV.PFV.leave | 3PL.2 |
- [*pema’un* hotel nuk inan deh m-udeng].
 in.front.of hotel REL exist 3PL.2 INTR-stay
 ‘In a suitcase that they left in front of the hotel where they were staying.’
- (text, BAR21082014CH_01 00:04:55.452-00:05:01.885)

In (43), the form *pema’un* ‘in front of’ could be interpreted either as a preposition with an NP complement or as a possessed noun.

Finally, *pe-* can indicate that the subject of a predicate is plural. This interpretation is typically derived when attached to stative roots:

- (44) **Plural Subject/Collective Interpretation**
- a. *lubid* ‘roll’ → *pelubid* ‘all roll’
 - b. *da’at iat* ‘sad’ → *peda’at iat* ‘all sad’
 - c. *teneb* ‘cold’ → *peteneb* ‘all cold’
 - d. *mulun* ‘live/shine’ → *pemulun* ‘all shine’

The interpretation is illustrated in (45):

- (45) **Function of *pe-* as plural subject marker**
- a. Pe-m-ulun lapung ih.
PL-INTR-live light PT
‘All the lights were burning.’ (elicitation, fieldnotes)
 - b. Pe-teneb deh lem takung ih.
PL-cold 3PL.2 in pond PT
‘They are all freezing in the pond.’ (elicitation, fieldnotes)

It is likely that the collective interpretation in (45) is linked to the reciprocal function, as is true in many Formosan languages (Zeitoun, p. c.). Indeed, reciprocity is often associated with dynamic verbs and collectivity with stative verbs (Lichtenberk 1985).

The other functions of *pe-* may not be related.

2.4.1.1.8 *peN-*

The main function of *peN-* is to derive the instrumental voice (IV). The nasal element has the same allomorphs as the AV nasal prefix:⁸⁰

- (46) **Instrumental Voice**
- a. *badaq* ‘instruction’ → *madaq* ‘AV.show’ → *pemadaq* ‘IV.show/teach’
 - b. *tatek* ‘closed’ → *natek* ‘AV.close’ → *penatek* ‘IV.close’
 - c. *raruh* ‘lost’ → *ngeraruh* ‘AV.lose’ → *pengeraruh* ‘IV.lose’
 - d. *lukaq* ‘fall over’ → *ngelukaq* ‘AV.push.over’ → *pengelukaq*
‘IV.push.over’
 - e. *abet* ‘a tie’ → *ngabet* ‘AV.tie’ → *pengabet* ‘IV.tie’
 - f. *upan* ‘bait’ → *ngupan* ‘AV.bait’ → *pengupan* ‘IV.bait’

⁸⁰ Indeed, it would be possible to analyse IV predicates as a combination of *pe-* + *N-* (see below for discussion).

The *peN-* IV prefix can also be attached to derived stems:

- (47) **IV with Derived Stems**
- a. *megatum* ‘AV.knot’ → *pemegatum* ‘IV.knot’
N- + pe- + gatum *peN- + pe- + gatum*
- b. *meru’it* ‘AV.spread’ → *pemeru’it* ‘IV.spread’
N- + pe- + ru’it *peN- + pe- + ru’it*

Since IV *peN-* can co-occur with causative *pe-* it would suggest that they two are separate prefixes, rather than treating *peN-* as a combination of *pe-* + *N-*.

Unlike the AV prefix, *peN-* does not apply productively to all roots and stems.

In some cases, the prefix serves a nominalisation function. Forms are typically interpreted as instrumental nouns, associative nouns or actor nominalisations:

- (48) **Instrumental Nouns**
- a. *naag* ‘AV.chop’ → *penaag* ‘instrument of chopping’
b. *natek* ‘AV.close’ → *penatek* ‘door stop’
c. *ngekeb* ‘AV.cover’ → *pengekeb* ‘lid’
- (49) **Associative Nouns**
- a. *kuman* ‘AV.eat’ → *penguman* ‘things eaten with rice’
b. *ngelamud* ‘AV.mix.TR’ → *pengelamud* ‘ingredients mixed in’
- (50) **Actor nominalisation**
- a. *nawar* ‘AV.call’ → *penawar* ‘the person who calls’
b. *ngitun* ‘AV.question’ → *pengitun* ‘matchmaker’

There are cognates of *peN-* with this function in many of the languages of Indonesia, such as the *paŋ-* prefix in Madurese (Davies 2010) and Sasak (Austin, p.c.). Some forms, such as *penatek* ‘IV.close, door stop’ in (48b), can have either verbal or nominal interpretations.

2.4.1.1.9 *peneN-*

Much like AV and UV, IV also has a realis/perfective form that is derived using the *peneN-* prefix:

- (51) **Perfective IV**
- a. *pengupan* ‘IV.bait’ → *penengupan* ‘IV.PFV.bait’
 - b. *pengukab* ‘IV.open’ → *penengukab* ‘IV.PFV.open’
 - c. *pengabet* ‘IV.tie’ → *penengabet* ‘IV.PFV.tie’

It is used in perfective contexts, and often has a past tense interpretation:

- (52) **Function of *peneN-***
- a. Enun *penengeluit* muh?
what IV.PFV.fish 2SG.2
‘What did you use to fish with?’
(text, BAR17082014CH_03 00:01:15.780-00:01:17.770)

Note that FORM 2 pronouns are used for actors in IV constructions, as is further discussed in CHAPTER 4. It may be possible to further subdivide the prefix into *peN-* and the perfective *ne₁-* (SUBSECTION 2.4.1.1.5). However, it is not clear why *ne₁-* would be infixes within the IV prefix as it is otherwise prefixed to derived stems (SUBSECTION 2.4.1.1.5).⁸¹ Moreover, it is in a paradigmatic relationship with *neN-* and *-in-*. Hence, it is treated as a single prefix in this thesis.

2.4.1.1.10 *pere-*

The main function of *pere-*, which is also pronounced *peri-*, is to derive a reflexive meaning. This is not a particularly productive process, but some examples are given in (53):

⁸¹ Alternatively, this might support an analysis that subdivides *peN-* into *pe-* and *N-*. In any case, both elements are required for IV, as *pe-* has other functions when attached to non-nasalised roots (SUBSECTION 2.4.1.1.7).

- (53) **Reflexive Interpretation**
- a. *ngatey* ‘AV.kill’ → *perepatey* ‘kill oneself’
 - b. *ngapung* ‘AV.hide’ → *perengapung* ‘hide in wait’
 - c. *nekitoi* ‘AV.hang’ → *perekitoi* ‘hang oneself from something’
 - d. *ngimet* ‘AV.hold’ → *perepimet* ‘hold oneself onto another’

In addition, *pere-* can imply insincerity, much like *se-* (SUBSECTION 2.4.1.1.11):

- (54) **Non-serious interpretation**
- a. Peri-da’at iat tupu tieh.
NON.SER-bad breath only PT=3SG.1
‘She’s just pretending to be in a bad mood.’ (elicitation, fieldnotes)

2.4.1.1.11 *se-*

The prefix *se-* attaches to bare roots and can be used with a range of functions. Firstly, it can indicate faked or non-serious action:

- (55) **Non-serious Interpretation**
- a. *riruh* ‘laugh’ → *seriruh* ‘pretend to laugh’
 - b. *nangey* ‘cry’ → *setangey* ‘pretend to cry’
 - c. *dooq* ‘good’ → *sedooq* ‘pretend to be good’
 - d. *atey* ‘death’ → *satey* ‘pretend to be dead’

- (56) **Function of *se-***
- a. Am tebeyq ieh dooq tu’uh liat kadiq ieh se-riruh teh.
NEG PT 3SG.1 good really mood because 3SG.1 NON.SER-laugh PT
‘He’s not actually that happy because he’s just pretending to laugh.’
(elicitation, fieldnotes)

In addition, *se-* can mark middle voice functions, since it derives grooming verbs, change of position and naturally reciprocal events (see Kemmer 1994):

- (57) **Middle Voice**
- a. *sanuk* ‘get dressed’ (from *anuk* ‘dress’)
 - b. *sengiduh* ‘lie down’
 - c. *selinuh* ‘thoughtful’ (from *linuh* ‘thought’)
 - d. *semulud* ‘converse’⁸² (from *ulud* ‘story’)

Unlike Malay/Indonesian, *se-* does not productively indicate ‘one’ with numbers. However, it is found in the form *sebuleng* ‘alone’ and in singular demonstratives (SUBSECTION 2.4.2.7).

2.4.1.1.12 *seN-*

In some instances, *seN-* is used as a variant UV perfective prefix. This is illustrated in (58) alongside traditional UV *-in-* infixation (see SUBSECTION 2.4.1.2.3):

(58)		Older Generations	Younger Generations
a.	stem	<i>laak</i>	<i>laak</i>
	AV	<i>nge-laak</i>	<i>nge-laak</i>
	UV	<i>l<in>aak</i>	<i>senge-laak</i>
b.	stem	<i>pidang</i>	<i>pidang</i>
	AV	<i>midang</i>	<i>m-idang</i>
	UV	<i>p<in>idang</i>	<i>sem-idang</i>

Like IV *peN-*, the prefix could be considered a combination of *se-* + *N-*. Usage is generally considered non-standard, incorrect and associated with the younger generations.

The development of *seN-* may have begun with the reanalysis of UV forms like *seninger* ‘UV.PFV.hear’ and *senaruq* ‘UV.PFV.do’. These are formed via *-in-* infixation from the roots *dinger* and *taruq*. However, the derivational process is made opaque by two phonological processes. The first is the reduction of the infix vowel to schwa in

⁸² It is not clear why there is an [m] in *semulud*. This may be a combination of *se-* + *-em-*.

pre-penultimate position (see SUBSECTION 2.3.5.4). The second is the fact that in the perfective the stem allomorphs *singer* and *saruq* are used.⁸³ Hence, the original derivation is obscured and reanalysed as the addition of a prefix to the nasalised stems *ninger* ‘AV.hear’ and *naruq* ‘AV.do’. This is subsequently extended to other roots, such as *laak*, which would not traditionally have formed the UV perfective in this manner (see SUBSECTION 2.4.1.2.3).

2.4.1.1.13 *te*₁-

The prefix *te*₁- is also an intransitive stem-forming suffix, like *pe*- (SUBSECTION 2.4.1.1.7):

- (59) **Deriving predicates with *te*-**
- | | | | |
|----|-------------------------------|---|-----------------------------------|
| a. | <i>pu'un</i> ‘first/start’ | → | <i>tepu'un</i> ‘be begun’ |
| b. | <i>anak</i> ‘child’ | → | <i>tenganak</i> ‘be born’ |
| c. | <i>iring</i> ‘near’ | → | <i>tebiring</i> ‘be on the side’ |
| d. | <i>kitoi</i> ‘act of hanging’ | → | <i>tekitoi</i> ‘hanging’ |
| e. | <i>ebpaq</i> ‘water’ | → | <i>tebpaq</i> ‘be full of liquid’ |

Like *ke*₂-, *te*₁- has several allomorphs, depending on whether it attaches to a consonant-initial root, a root that begins with schwa, or a root that begins with any other vowel (see SUBSECTION 2.4.1.1.3). For roots beginning with schwa, the allomorph *t*- is used, as in (59e). For roots beginning with other vowels, an epenthetic consonant is added. It is not clear why /ng/ is added in (59b) and /b/ in (59c). The choice may be lexically determined or *te*- may attach to derived stems. This remains to be further explored.

⁸³ Blust (2001) argues that the sibilant of /t/ to [s] was originally conditioned by the high front vowel of the infix. However, this conditioning factor is lost when all pre-penultimate vowels are reduced to schwa.

Typically, *te-* predicates have a stative interpretation:⁸⁴

(60) **Function of *te-***

- a. Te-kitoi kayuq na'ah ih teh pa'up peped buluq nih.
 STAT-hang like before PT PT both ends bamboo DEM
 'Both ends of the bamboo are hanging like before.'
 (text, BAR27102013CH_01 00:05:54.577-00:06:00.166)

This may be cognate with *te-* as accidental, unintentional or uncontrolled movement in the Punan Tubu' language of Kalimantan (Soriente 2013: 184). It may also be related to Malay stative *ter-* or Sasak passive *te-* (Austin, p.c.).

Like *pe-* verbs, *te-* can be combined with *N-* to derive transitive AV constructions, or *-in-* to derive transitive UV constructions:

(61) ***te-* in Actor Voice**

- a. Ne-ne-pu'un deh naruq Sarawak Rangers.
 PFV-AV.STAT-start 3PL.2 AV.do Sarawak Rangers
 'They started the Sarawak Rangers.'
 (text, BAR25102013CH_03 00:09:06.245-00:09:09.950)

b. ***te-* in Undergoer Voice**

- Senepu'un kuh edto ma'un ih neh baney sinih.
 UV.PFV.STAT.start 1SG.2 day old PT PT necklace DEM
 'I started this necklace the day before yesterday.'
 (elicitation, BAR28102013CH_01 01:02:46.461-01:02:51.661)

⁸⁴ In cases where a verbal predicate is derived via *te-* prefixation, there are often two variant perfective forms, derived via *-in-* affixation to either the root *epak*, or the stem *tepak*. The latter form (in this case *sipak*) is a verbal form and is used transitively, whilst the former (in this case *ipak*) is used adjectivally to modify nouns:

- (i) Sipak kuh neh kayuq dih (ii) Kayuq ipak
 UV.PFV.chop 1SG.2 PT wood DEM wood chopped
 'I chopped the wood' 'Chopped wood'
 (elicitation, BAR28102013CH_03 01:33:59.837-01:34:26.689)

2.4.1.1.14 *te*₂-

Finally, *te*₂- expresses distribution when it attaches to numerals (cf. Malay *berduah-duah*). As shown in TABLE 2.6, the allomorph *tet*- is used with vowel-initial roots. This contrasts with *te*₁-, where the allomorph used with vowel-initial roots is *t*- and hence *te*₂- can be considered a separate morpheme.

Table 2.6 *Distributive Numerals in Kelabit*

	Cardinal	Distributive Numeral	Meaning
1	edteh	tetedteh	one-by-one
2	duweh	teduweh	two-by-two
3	teluh	teteluh	three-by-three
4	epat	tetepat	four-by-four
5	limeh	telimeh	five-by-five
6	enem	tetenem	six-by-six
7	tuduq	tetuduq	seven-by-seven

A distributive numeral is illustrated in (62):

- (62) **Distributive Numerals**
 Tet-epat burur kamih mayaq lem edteh taksi.
 DISTR-four body 3PL.EXCL AV.follow in one taxi
 ‘The four of us will follow in one taxi.’
 (elicitation, BAR21102013CH_02 00:56:51.855-00:56:55.531)

2.4.1.2 Infixation

Although less common than prefixes, Kelabit has two relatively productive infixes, *-em-* and *-in-* (SUBSECTION 2.4.1.2.1 and SUBSECTION 2.4.1.2.3). There is also a remnant form of the PAn **-um-* infix used to mark AV in the verb *kuman* ‘eat’ (SUBSECTION 2.4.1.2.2). This is probably related to *-em-* historically but has a distinct synchronic function and is therefore handled separately in this thesis (see SUBSECTION 4.2.1.2.2). They tend to attach after the first consonant and before the first vowel and are discussed in turn below.

Table 2.7 Infixes in Kelabit

Morpheme	Function	Root/Stem	Infixed Form
-em-	intransitive verb	<i>turun</i> ‘down’	<i>temurun</i> ‘go down’
-um-	remnant AV marker	<i>kan</i> ‘eat’	<i>kuman</i> ‘AV.eat’
-in-	UV realis/perfective	<i>bilag</i> ‘broken’	<i>binilag</i> ‘UV.PFV.break’

2.4.1.2.1 -em-

The basic function of *-em-* is to derive intransitive verbs. When attached to a verbal root, the infix adds an entailment of intention, volition or agency to the actor:

(63) Intransitive Verbs

- a. *tuked* ‘go up’ → *temuked* ‘climb up’
- b. *turun* ‘go down’/‘fall down’ → *temurun* ‘intentionally go down’
- c. *tutuq* ‘drop down’ → *temutuq* ‘drop down (purposefully)’
- e. *libung* ‘in a circle’ → *lemibung* ‘to circle’
- f. *liget* ‘looking behind’ → *lemiget* ‘turn around to look’
- g. *terem* ‘be sunk’ → *temerem* ‘to sink/dunk oneself’
- h. *raruh* ‘lost’ → *remaruh* ‘to get oneself lost/run away’

The contrast between non-volitional bare predicates and *-em-* verbs is illustrated in

(64):

(64) Function of -em-

- a. Ne-terem ieh lem ebpaq ih.
 PFV.sunk 3SG.1 in river PT
 ‘He sank in the river.’
- b. Ieh ne-terem lem takung ih.
 3SG.1 PFV<INTR>sunk in pond PT
 ‘He dunked himself in the pond.’

(elicitation, BAR28102013CH_03 00:00:58.868-00:01:44.364)

These constructions could also be analysed as resultatives, bringing about the base predicate as a result of an action, without introducing a distinct causative agent. The predicates are intransitive and cannot generally be used transitively:⁸⁵

(65) **Intransitive**

- a. M-eseb neh uduh nuk ngi iring dalam sineh.
 INTR-burn PT grass REL at near road DEM
 ‘The grass next to the road is burning.’
 (elicitation, BAR17102013CH_01 00:16:49.050-00:16:56.040)

Transitive

- b. *Uih ne-m-eseb arep ih.
 1SG.1 PFV-INTR-burn rubbish PT
 FOR: ‘I burnt the rubbish.’
 (elicitation, BAR17102013CH_01 00:46:00.793-00:46:04.618)

Transitive versions of *-em-* predicates are formed via *N-* prefixation to the bare root, as discussed in SUBSECTION 2.4.1.1.4.

When attached to noun roots, *-em-* infixation is category-changing and derives an intransitive verb:

(66) **Deriving Verbs from Nouns**

- a. *ulun* ‘life’ → *mulun* ‘live’
 b. *udan* ‘rain’ → *mudan* ‘to rain’
 c. *laput* ‘cloud’ → *lemaput* ‘to be cloudy’
 d. *erur* ‘tiredness’ → *merur* ‘to be tired’
 e. *tulud* ‘plane’ → *temulud* ‘to fly’

The allomorph *m-* is used when *-em-* attaches to vowel-initial roots. The *-em-* infix could be thought of as a reflex of PAN **-um-* (see SUBSECTION 2.4.1.2.2).

⁸⁵ However, in certain environments they can be used in a way that appears syntactically transitive, at least in the sense that there are two nominal arguments, both expressed as NPs:

- (i) Peter na'am kereb temuked puun sineh
 Peter NEG can climb mountain DEM
 ‘Peter can’t climb that mountain’
 (elicitation, BAR21102013CH_01 01:37:07.005-01:37:10.343)

2.4.1.2.2 *-um-*

Kelabit has one predicate that marks AV with an *-um-* infix, namely *kuman* ‘AV.eat’. This is widespread as the verb ‘eat’ in Borneo and the Philippines. Unlike other lexical roots, *kan* is monosyllabic. The function of AV is fulfilled by the nasal prefix for all other predicates (see SUBSECTION 2.4.1.1.4). The main voice forms of the verb ‘eat’ are shown in (67):

(67)	ROOT	AV	UV.PFV	UV.IRR
	kan	k<um>an	k<in>an	ken-en

Hence, *kuman* maintains the more conservative AV infix, which is reconstructed to Proto-Austronesian (PAn) and has cognates in a number of Western Austronesian languages, particularly in Philippine-type languages (see SUBSECTION 3.4).

Elsewhere, the PAn infix **-um-* has probably developed into the intransitive *-em-* infix. This follows the reduction of vowels to schwa in pre-penultimate syllables (SUBSECTION 2.3.5.4). It is possible to think of *-em-* and *-um-* as allomorphs of the same morpheme, where *-um-* is infixes in monosyllabic roots beginning with consonants, which are rare in contemporary Kelabit. However, the functions of *-um-* and *-em-* are quite different and for this reason they are handled separately in this thesis, even if there is a historical relationship.

2.4.1.2.3 *-in-*

Unlike the PAn **-um-* infix, the *-in-* infix (a reflex of PAn **-in-*) is highly productive in Kelabit, as well as in the Philippines, and marks both perfective aspect/realis mood and undergoer voice.⁸⁶ The *-in-* infix has several allomorphs depending on whether

⁸⁶ It is unclear whether this marks aspect or mood. The most typical interpretation seems to be past tense/perfective aspect. However, since *-in-* is in a paradigmatic relationship with *-en*, which marks

the root is consonant-initial or vowel-initial, and whether the first vowel of the root is schwa.

If the first vowel is schwa, it is replaced with *-i-* in the UV perfective form:

(68) **Ablaut**

- a. *perek* → *merek* ‘AV.squeeze’ → *pirek* ‘UV.PFV.squeeze’
- b. *beppet* → *mebpet* ‘AV.hit’ → *bibpet* ‘UV.PFV.hit’
- c. *telaq* → *nelaq* ‘AV.throw’ → *silaq* ‘UV.PFV.throw’
- d. *dedtar* → *nedtar* ‘AV.throw down’ → *sidtar* ‘UV.PFV.throw down’
- e. *semin* → *nemin* ‘AV.cement’ → *simin* ‘UV.PFV.cement’
- f. *kedta* → *ngedta* ‘AV.withstand’ → *kidta* ‘UV.PFV.withstand’
- g. *gegkang* → *ngegkang* ‘AV.lift up’ → *gigkang* ‘UV.PFV.lift up’
- h. *letuq* → *ngelutuq* ‘AV.pluck’ → *lituq* ‘UV.PFV.pluck’
- i. *redtuq* → *ngeredtuq* ‘AV.fold’ → *ridtuq* ‘UV.PFV.fold’
- j. *emung* → *ngemung* ‘AV.collect’ → *imung* ‘UV.PFV.collect’

This process is described as ablaut (Blust 1997). In addition to the vowel gradation, when the root begins with an alveolar/dental consonant, /t/ or /d/, the initial consonant undergoes sibilation to [s].⁸⁷ This may have been conditioned by the high front vowel of the infix (see SUBSECTION 2.4.1.1.12).

If the first vowel is anything other than schwa, the infix *-in-* is used. It is typically infixed between the first consonant and first vowel of the root. The *-in-* infix is subject to vowel reduction in pre-penultimate syllables and is consequently often pronounced [ən] though [ɪn] is also heard (SUBSECTION 2.3.5.4).

irrealis mood (see SUBSECTION 2.4.1.3.1), it could be argued to represent a mood contrast. In this thesis, it is glossed UV.PFV with the understanding that this could also be understood as realis mood.

⁸⁷ In certain dialects, particularly in Pa’ Umur, roots beginning with /d/ do not have a perfective stem allomorph beginning with /s/. Forms such as *dinawar* and *senawar* ‘UV.PFV.call’ are used in free variation.

- (69) **-in- Infixation**
- | | | | | | | | |
|----|---------------|---|-----------------|------------|---|-----------------|----------------|
| a. | <i>pa'id</i> | → | <i>ma'id</i> | 'AV.wipe' | → | <i>pena'id</i> | 'UV.PFV.wipe' |
| b. | <i>babeh</i> | → | <i>mabeh</i> | 'AV.carry' | → | <i>benabeh</i> | 'UV.PFV.carry' |
| c. | <i>tatek</i> | → | <i>natek</i> | 'AV.close' | → | <i>senatek</i> | 'UV.PFV.close' |
| d. | <i>dinger</i> | → | <i>ninger</i> | 'AV.hear' | → | <i>seninger</i> | 'UV.PFV.hear' |
| e. | <i>si'er</i> | → | <i>ni'er</i> | 'AV.see' | → | <i>seni'er</i> | 'UV.PFV.see' |
| h. | <i>lubid</i> | → | <i>ngelubid</i> | 'AV.roll' | → | <i>lenubid</i> | 'UV.PFV.pluck' |
| i. | <i>ruyuh</i> | → | <i>ngeruyuh</i> | 'AV.sway' | → | <i>renuyuh</i> | 'UV.PFV.sway' |

Again, roots beginning with /t/ and /d/ have a stem allomorph beginning with /s/ in the perfective. It is possible that [ən] is being reanalysed as the UV.PFV infix rather than a phonologically conditioned variation of *-in-*, since it is sometimes extended to situations where ablaut would be expected. In these cases, multiple forms are in free variation:

- (70)
- | | AV Form | | Ablaut UV Form | | Infixd UV Form | |
|----|----------------|---------|-----------------------|------------|-----------------------|------------|
| a. | <i>mebpet</i> | 'hit' | <i>bibpet</i> | 'UV.hit' | <i>benebpet</i> | 'UV.hit' |
| b. | <i>nelaq</i> | 'throw' | <i>silaq</i> | 'UV.throw' | <i>senelaq</i> | 'UV.throw' |
| c. | <i>rerep</i> | 'lower' | <i>rirep</i> | 'UV.lower' | <i>renerep</i> | 'UV.lower' |

When roots begin with a vowel that is not schwa, the prefix *n-* is used:

- (71) ***n-* Prefixation**
- | | | | | | | | |
|----|--------------|---|----------------|------------|---|---------------|----------------|
| a. | <i>imet</i> | → | <i>ngimet</i> | 'AV.hold' | → | <i>nimet</i> | 'UV.PFV.hold' |
| b. | <i>aweh</i> | → | <i>ngaweh</i> | 'AV.marry' | → | <i>naweh</i> | 'UV.PFV.marry' |
| c. | <i>udtuq</i> | → | <i>ngudtuq</i> | 'AV.stop' | → | <i>nudtuq</i> | 'UV.PFV.stop' |

This could be thought of as *in-* with vowel deletion of the initial /i/ (see SUBSECTION 2.3.5.5). In this way, *-in-* is similar to *-em-*, which has a variant *m-* that attaches to vowel-initial roots (SUBSECTION 2.4.1.2.1).

Finally, when the root begins with /k/ or /g/ there are two possibilities. Either the infix *-in-* is used in analogy with other consonant-initial roots, or the prefix *n-* is used in analogy with other UV forms for AV verbs beginning with /ŋ/:

- (72) **Infixation or Prefixation**
- a. *kiding* → *ngiding* ‘AV.lift’ → *niding* ‘UV.PFV.lift’
kiniding ‘UV.PFV.lift’
 - b. *guraq* → *nguraq* ‘AV.shake salt’ → *nuraq* ‘UV.PFV.shake salt’
 - c. *gisek* → *ngisek* ‘AV.produce timber’ → *ginisek* ‘UV.PFV.produce timber’

It is not clear if the choice of allomorph makes a meaningful difference. In some cases, the choice between *n-* and *-in-* is lexicalised and in others the *n-* prefix is preferred.

Like the AV prefix in SUBSECTION 2.4.1.1.4, the *-in-* infix can co-occur with other predicate-forming prefixes, such as *pe-* and *te-*. In these cases, the final nasal is subject to nasal assimilation as it is infixed between the prefix and the root (SUBSECTION 2.3.5.1):

(73)	a. <i>-in-</i> → [im]	AV form <i>nebukuh</i> <i>N- + te- + bukuh</i> ‘AV. knot’	UV.PFV form <i>simbukuh</i> <i>te- + -in- + bukuh</i> ‘UV. knot’
	b. <i>-in-</i> → [in]	<i>metelaq</i> <i>N- + pe- + telaq</i> ‘AV.throw.away’	<i>pintelaq</i> <i>pe- + -in- + telaq</i> ‘UV.throw.away’
	c. <i>-in-</i> → [iŋ]	<i>nekaruh</i> <i>N- + te- + karuh</i> ‘AV.talk’	<i>singkaruh</i> <i>te- + -in- + karuh</i> ‘UV.talk’

Typically, the *-in-* infix marks high transitivity (see SUBSECTION 3.5.2). However, it can also be used to signal past tense/perfective action with a few intransitive predicates, such as *mey* ‘to go’:⁸⁸

⁸⁸ This construction could be thought of as transitive, since there are two nominal arguments. However, it differs from other uses of *-in-* in that the actor and not the undergoer is subject. This can be seen from the fact that the particle *neh* precedes the actor (see SUBSECTION 2.5.1.1).

- (74) **-in- with Intransitives**
- a. M<in>ey rumaq John neh kamih ngimalem.
 <PFV>go house John PT 1PL.EXCL yesterday
 ‘We went to John’s house yesterday.’
 (elicitation, BAR14102013CH_02 00:33:55.800-00:34:01.830)

There is a clear difference in tense-aspect-mood between the use of the UV *-in-* infix and the *-en* suffix. This is discussed in SUBSECTION 2.4.1.3.1 and is largely parallel to the use of *N-* or *neN-* in AV.

2.4.1.3 Suffixation

Like infixes, suffixes in Kelabit are much less common than prefixes. They are limited to the irrealis UV suffix, the UV imperative, fossilised possessive suffixes on a subset of inalienable nouns and a locative nominalising suffix.

Table 2.8 Suffixes in Kelabit

Morpheme	Function	Root/Stem	Suffixed Form
<i>-en</i>	UV irrealis	<i>laak</i> ‘cook’	<i>laaken</i> ‘UV.IRR.cook’
<i>-um</i>	UV imperative	<i>bala</i> ‘say’	<i>bela’um</i> ‘UV.IMP.say’
<i>-q/-m/-n</i>	inalienable possession	<i>tesineh</i> ‘mother’	<i>sinaq</i> ‘my mother’ <i>sinam</i> ‘your mother’ <i>sinan</i> ‘its mother’
<i>-an</i>	locative nominalisation	<i>telen</i> ‘swallow’	<i>telenan</i> ‘throat’

2.4.1.3.1 *-en*

The primary function of *-en* is to mark UV irrealis verbs. These are used in typical irrealis contexts, such as imperatives, conditionals and generic statements:

- (75) **Imperatives**
- a. Laak-en narih si'it nubaq na'an.
 cook-UV.IRR IMPERS little rice later
 'Could you cook a little rice later.' (elicitation, fieldnotes)

- Conditionals**
- b. Rengaq narih la'uh laak-en narih nubaq.
 if IMPERS hungry cook-UV.IRR IMPERS rice
 'If you're hungry, cook some rice.' (elicitation, fieldnotes)

- Generic Statements**
- c. Kinih teh dooq laak-en kuh nubaq ih.
 like.this PT good cook-UV.IRR 1SG.2 rice PT
 'This is the best way I can cook rice.' (elicitation, fieldnotes)

As a result, the *-en* suffix never co-occurs with the perfective *-in-* infix. Accordingly, *-in-* can be used with adjuncts representing past time reference, whilst *-en* cannot, and *-en* can be used with adjuncts representing future time reference, whilst *-in-* cannot:

- (76) **Undergoer Voice -in-**
- a. Kinan John neh buaq kabér nedih ngimalem.
 UV.PFV.eat John PT fruit pineapple 3SG.POSS yesterday
 'John ate his pineapple yesterday.'
 (elicitation, BAR21102013CH_01 00:27:31.945-00:27:37.756)

- b. *Kinan John neh buaq kabér nedih edto riak.
 UV.PFV.eat John PT fruit pineapple 3SG.POSS tomorrow
 For: 'John will eat his pineapple tomorrow.' (elicitation, fieldnotes)

- (77) **Undergoer Voice -en**
- a. Ken-en Peter teh buaq kabér nedih edto riak.
 eat-UV.IRR Peter PT fruit pineapple 3SG.POSS tomorrow
 'Peter will eat his pineapple tomorrow.'
 (elicitation, BAR21102013CH_01 00:25:12.486-00:25:15.526)

- b. *Ken-en Peter neh buaq kabér nedih ngimalem.
 eat-UV.IRR Peter PT fruit pineapple 3SG.POSS yesterday
 For: 'Peter ate his pineapple yesterday.' (elicitation, fieldnotes)

Thus, the two UV affixes are in complementary distribution.⁸⁹

The same function is fulfilled by the suffix *-en* in some cases, and the suffix *-an* in other cases:⁹⁰

(78) **-en Suffixation**

- a. *pupuq* → *pepu'en* 'UV.IRR.hit'
- b. *kan* → *kenen* 'UV.IRR.eat'
- c. *tekap* → *tekapen* 'UV.IRR.search for'
- d. *laak* → *laaken* 'UV.IRR.cook'

-an Suffixation

- e. *ulud* → *ludan* 'UV.IRR.talk about sth'
- f. *badaq* → *beda'an* 'UV.IRR.show/invite'
- g. *belad* → *beladan* 'UV.IRR.open again'

There does not seem to be a meaningful difference between the use of *-an* and *-en* in (78). However, *-an* is also used with a few remnant forms to mark a locative voice construction and as a nominalising suffix (SUBSECTION 2.4.1.3.4).

Suffixation with *-en* and *-an* triggers a series of phonological processes, as discussed in SUBSECTION 2.3.5. Firstly, pre-penultimate vowels are reduced to schwa (SUBSECTION 2.3.5.1):

(79) **Vowel Reduction under *-en* suffixation**

- a. *badaq* → *beda'an* 'UV.IRR.show/invite'
- b. *pupuq* → *pepu'en* 'UV.IRR.hit'
- c. *piliq* → *pe'li'en* 'UV.IRR.choose'

⁸⁹ Similar distributional facts obtain for the *N-* vs *neN-* (or *-um-* vs *ne-* + *-um-*). Hence these can be treated as a single paradigm (see CHAPTER 3):

- (i) Edto riak teh Peter kuman buaq kaber nedih.
tomorrow PT Peter AV.eat fruit pineapple 3SG.POSS
'Tomorrow Peter will eat pineapple.'
(elicitation, BAR21102013CH_01 00:22:56.540-00:23:00.540)
- (ii) Peter ne-kuman buaq kaber nedih ngimalem.
Peter PFV-AV.eat fruit pineapple 3SG.POSS yesterday
'Peter ate his pineapple yesterday.'
(elicitation, fieldnotes)

⁹⁰ The apostrophe represents a word-medial glottal stop (see TABLE 2.3).

Similarly, voiced aspirates are changed to voiced plosives when they do not occur in the onset of a final syllable:

- (80) **Consonant Reduction under *-en* suffixation**
- | | | | | |
|----|---------------|---|----------------|-----------------------------|
| a. | <i>tebpaq</i> | → | <i>teba'an</i> | 'UV.IRR.fill' |
| b. | <i>tudtuq</i> | → | <i>tedu'en</i> | 'UV.IRR.salt' ⁹¹ |
| c. | <i>pigkuq</i> | → | <i>pegu'en</i> | 'UV.IRR.touch (cause pain)' |

This reflects the fact that pre-penultimate syllables are unstressed.

Secondly, root-final consonants and vowels undergo a series of changes. If the root ends in a non-schwa vowel, the vowel is lengthened and spread across both syllables. Where possible, the vowel is realised as a semi-vowel or glide in the onset of the final syllable. For roots ending in diphthongs, the vowels are simply divided across the two syllables:

- (81) **Vowel lengthening across syllables**
- | | | | | | |
|----|-------------------------|--------------|---|----------------|-------------------|
| a. | /i/ → /iy/ | <i>abi</i> | → | <i>biyen</i> | 'UV.IRR.finish' |
| | | <i>belih</i> | → | <i>beliyen</i> | 'UV.IRR.buy' |
| b. | /e/ → /ay/ | <i>patey</i> | → | <i>petayen</i> | 'UV.IRR.kill' |
| c. | /o/ → /aw/ | <i>peno</i> | → | <i>penawen</i> | 'UV.IRR.steal' |
| | | <i>alo</i> | → | <i>lawen</i> | 'UV.IRR.chase' |
| | | <i>uto</i> | → | <i>tawen</i> | 'UV.IRR.tease' |
| d. | /u/ → /uw/ | <i>putuh</i> | → | <i>petuwen</i> | 'UV.IRR.request' |
| | | <i>linuh</i> | → | <i>lenuwen</i> | 'UV.IRR.think' |
| e. | /a/ → /a: ⁹² | <i>sipa</i> | → | <i>sepaan</i> | 'UV.IRR.pack' |
| | | <i>bala</i> | → | <i>belaan</i> | 'UV.IRR.say' |
| | | <i>siwa</i> | → | <i>sewaan</i> | 'UV.IRR.exchange' |
| | | <i>laba</i> | → | <i>lebaan</i> | 'UV.IRR.pass' |
| f. | /iu/ → /iw/ | <i>biliu</i> | → | <i>beliwen</i> | 'UV.IRR.let.go' |
| | | <i>pasiu</i> | → | <i>pesiwen</i> | 'UV.IRR.sell' |

⁹¹ This form sounds possible but may not be in existence (Florance Apu, p.c.)

⁹² It remains to be explored whether these are simply long vowels or geminate vowels. The latter hypothesis may be likely, given that consonant germination also occurs in this context (see below).

For roots ending in schwa, the vowel of the root is deleted:

(82) **Vowel deletion and lengthening in final syllable**

- a. /ə/ → Ø *babeh* → *baben* ‘UV.IRR.carry (on back)’
 tedteh → *tedtan* ‘UV.IRR.leave’

These processes reflect the preference for the onset of final syllables to be filled (SUBSECTION 2.3.3). The same processes occur when roots end in a vowel + /h/, as /h/ is deleted intervocalically.

When the stem ends in a schwa + a voiced consonant, the consonant is lengthened and realised as a voiced aspirate. It forms part of both the following and preceding syllables, just like the vowels in (81):

(83) **Consonant Lengthening across boundaries**

- a. /b/ → /bp/ *rereb* → *rerebpen* ‘UV.IRR.baptise’
 eseb → *sebpen* ‘UV.IRR.burn’
 kekeb → *kekebpen* ‘UV.IRR.cover’
- b. /d/ → /dt/ *tuked* → *tekedten* ‘UV.IRR.put.at.angle’
 lened → *lenedten* ‘UV.IRR.cook (vegetables)’
- c. /g/ → /gk/ *eleg* → *legken* ‘UV.IRR.divorce’

As discussed in SUBSECTION 2.3.5.3, this does not occur if the final vowel is not a schwa, as the consonant is simply resyllabified as part of the final syllable:

(84) **Resyllabification following non-schwa vowels**

- a. *ukab* → *kaben* [ka:.bən] ‘UV.IRR.open’
 b. *terad* → *teraden* [tə.ra:.dən] ‘UV.IRR.cut’
 c. *palug* → *pelugan* [pə.lu:.gan] ‘UV.IRR.trick’

Hence, lengthening appears to occur to satisfy the need for a filled onset in final-syllables and maintain stress on the penultimate syllable. Where roots end in non-schwa vowel + consonant, the root-final consonant is simply resyllabified. Where

roots are vowel-final, the vowel is lengthened across both syllables. Where roots end in schwa + consonant, the consonant is lengthened across both syllables.

Finally, some roots drop syllables under suffixation. It is possible that the AV forms are derived from conservative mono-syllabic roots (in the same manner as *kan* ‘eat’) and involve additional stem-forming prefixation. Alternatively, there may just be an overriding preference for bisyllabic words.⁹³ Some examples are listed in (85):

(85)	Syllable elision under suffixation		
a.	<i>neput</i> ‘AV.use.blow.pipe’	→	<i>putan</i>
b.	<i>nuruq</i> ‘AV.instruct’	→	<i>ru’en</i>
c.	<i>nukat</i> ‘AV.dig’	→	<i>katen</i>
d.	<i>naruq</i> ‘AV.do’	→	<i>tu’en</i>
e.	<i>ngalap</i> ‘AV.fetch’	→	<i>apen</i>
f.	<i>ngi’ir</i> ‘AV.sharpen’	→	<i>iran</i>
g.	<i>ngitun</i> ‘AV.question’	→	<i>tunen</i>
h.	<i>muit</i> ‘AV.take’	→	<i>iten</i>
i.	<i>memug</i> ‘AV.remove’	→	<i>pugen</i>
j.	<i>ngenep</i> ‘AV.catch’	→	<i>depen</i>
k.	<i>nganud</i> ‘AV.float’	→	<i>nuden</i>
l.	<i>ma’en</i> ‘AV.carry(on shoulders)’	→	<i>panen</i>
m.	<i>mepaq</i> ‘AV.chew’	→	<i>pa’en</i>
n.	<i>na’uk</i> ‘AV.scoop’	→	<i>uken</i>

In cases like *tunen* from *itun*, forms are subject to initial vowel deletion, as described in SUBSECTION 2.3.5.5. However, in other cases it is medial syllables that are dropped. It may be that the original roots have dropped out of use and the UV irrealis forms are simply preserved.

In at least two cases, the UV irrealis form of the verb occurs with the additional prefix *pe-*. It is not clear whether these are functionally different from other *-en* verbs. However, UV irrealis forms with *-en* alone do not exist for these predicates:

⁹³ In some cases, both bisyllabic and trisyllabic forms are in co-existence. For example, both *keruwen* ‘UV.IRR.talk’ and *kuwan* were elicited. It is not clear, if *kuwan* was once a locative voice form or if one is used more frequently than the other. Similarly, *lenuwen* and *liwen* from *linuh* ‘think’.

- (86) **Undergoer Voice Circumfix**
 a. *dinger* → *pedingeren* (**dingeren*)
 b. *ranih* → *pereniyen* (**reniyen*)

Thus, the *-en* suffix triggers a number of phonological processes, and has two allomorphs that seem to be lexically triggered, namely *-an* and *pe-* *-en*. Perhaps because of the phonological complexity, only the more frequent vocabulary items are in widespread use and the form is not especially productive. Consequently, the morphological irrealis is often replaced with a periphrastic *tu'en* construction, described in SUBSECTION 2.5.2.2.

2.4.1.3.2 *-um*

The main function of *-um* is to mark a UV imperative. This is considered very polite and is associated with the older generations and ‘deep’ Kelabit:

- (87) **Function of *-um***
 a. Bela-um muh dih ngedeh nangey.
 say-UV.IMP 2SG.2 DEM to.3PL there
 ‘Would you please tell them over there.’ (elicitation, fieldnotes)

It is even less productive than the UV irrealis suffix and has only been elicited in the following forms, though it is found in old songs (cf. Talla 1979: 192):

- | | | | | |
|------|----------|---------------|---------------|---------------|
| (88) | | AV | UV.IRR | UV.IMP |
| a. | eat | <i>kuman</i> | <i>kenen</i> | <i>kenum</i> |
| b. | take | <i>ngalap</i> | <i>apen</i> | <i>apum</i> |
| c. | see | <i>ni'er</i> | <i>siren</i> | <i>sirum</i> |
| d. | question | <i>muit</i> | <i>iten</i> | <i>itum</i> |
| e. | drink | <i>mirup</i> | <i>rupen</i> | <i>rupum</i> |
| f. | say | <i>mala</i> | <i>belaan</i> | <i>belaum</i> |
| g. | do | <i>naruq</i> | <i>tu'en</i> | <i>tu'um</i> |

It is possible that the *-um* and *-en* suffixes derive from second and third person possessive suffixes (see SUBSECTION 2.4.1.3.3) and that they originally represented third and second person non-subject actors.⁹⁴ However, in elicited examples they always co-occur with overt pronouns that realise the actor of the irrealis or imperative construction, as in (87). Hence, the function of the suffixes in modern Kelabit is to mark undergoer voice together with the irrealis interpretations. UV imperatives are also found in Lundayeh, which is said to have three suffixes *-a*, *-u* and *-i*. These are used only rarely in discourse (Clayre 2002).

2.4.1.3.3. *-q*, *-m*, *-n*

There are three non-productive pronominal suffixes or enclitics that mark possession of select inalienable kinship terms:

Table 2.9 Bound Possessive Pronouns

Person	Suffix
1SG	<i>-q</i>
2SG	<i>-m</i>
3SG	<i>-n</i>

These are only found today on kinship terms for mother, father and grandparent:

- (89) **Inalienable Possessive Pronouns**
- a. 1SG.POSS → *sinaq* ‘my mother’
tamaq ‘my father’
tepuq ‘my grandmother/grandfather’
- b. 2SG.POSS → *sinam* ‘your mother’
tamam ‘your father’
tepum ‘your grandmother/grandfather’

⁹⁴ Moreover, a few speakers recalled verbal forms ending in *-aq*, which were used in the past as commands, i.e. *apaq* from *ngalap* ‘fetch’ (Lucy Bulan, p.c.)

- c. 3SG.POSS → *sinan* ‘mother.3SG.POSS’
taman ‘father.3SG.POSS’
tepun ‘grandparent.3SG.POSS’

The first person forms are typically used when directly addressing a family member. The second person forms are typically used to refer to a family member when addressing a third party who is younger than the speaker, as in (90a). The third person forms are very rarely used, but do occur in names and in reference to animals, as in (90b).

(90) **Inalienable Possessive Pronouns**

- a. Peh Tama-m?
 where father-2SG.POSS
 ‘Where’s your father/uncle?’ (elicitation, fieldnotes)
- b. Duweh neh, sina-n mey anak ih.
 two DEM mother-3SG.POSS and child PT
 ‘So that’s two, the (wild boar) sow and the piglet.’
 (text, PDA10112013CH_02 00:01:17.030-00:01:19.260)

The more common way to refer to relations of a third person is using the forms *tesineh*, *tetameh* and *tetepuh*. The function of the *te-* prefix is unknown but may be used for politeness in a similar manner to *te* in Sasak (Peter Austin, p.c.) or honorific particles in other Western Austronesian languages (see also *si* in Riau Indonesian in SUBSECTION 5.4.2):

(91) **Possession of kinship terms**

- a. te-sineh nedih
 HON-mother 3SG.POSS
 ‘his/her mother’
- b. te-tameh nedih
 HON-father 3SG.POSS
 ‘his/her father’

- | | | | |
|----|--|-------------------|---------------------------|
| c. | te-tepuh
HON-grandparent
'his/her grandparent' | nedih
3SG.POSS | (elicitation, fieldnotes) |
|----|--|-------------------|---------------------------|

The inalienable possessive suffixes may be related to the UV suffixes *-en* and *-um*, as discussed in SUBSECTION 2.4.1.3.2.

2.4.1.3.4 *-an*

In addition to functioning as an allomorph of *-en*, *-an* also functions as a non-productive nominalising suffix. Most nouns that end in *-an* have a locative interpretation, though this appears lexicalised:

- (92) **Nominalisation with *-an***
- | | | | |
|----|-------------------------|---|--|
| a. | <i>telen</i> 'swallow' | → | <i>telenan</i> 'throat' |
| b. | <i>guta</i> 'cross' | → | <i>getaan</i> 'crossing' |
| c. | <i>naag</i> 'chop' | → | <i>penagan</i> 'chopping board' |
| d. | <i>irup</i> 'drink' | → | <i>rupan</i> 'watering hole' |
| e. | <i>laak</i> 'cook/ripe' | → | <i>laakan</i> 'season for rice planting' |
| f. | <i>gatum</i> 'knot' | → | <i>getuman</i> 'connection' |

It is possible that *-an* is descended from an earlier locative voice suffix, which is reconstructed for PAn (cf. Ross 2002, Adelaar 2005). It seems to survive in this function in Kelabit in the form *tu'an*. This marks a construction where the locative is subject, as demonstrated in SUBSECTION 2.5.3.1:

- (93) **Remnant Locative Voice**
- | | | | | | |
|----|---------------------------------|-----------|-------|------|---------------------------|
| a. | Lidung | tu'an | neh | babe | nedih. |
| | corner | do/put-LV | 3SG.2 | bag | 3SG.POSS |
| | 'He put his bag in the corner.' | | | | (elicitation, fieldnotes) |

Locative and dative voice constructions are common in the languages of Sabah but typically not found in the languages of Sarawak (Clayre 1991). This construction applies only to *tu'an* and is not found for any other predicates.

2.4.1.4 Reduplication

The second major word-formation process in Kelabit is reduplication. This is common in Western Austronesian and can have a number of distinct functions, depending on the type of root being reduplicated (Blust 2013). Reduplication can have both derivational and inflectional functions. In other words, reduplication can derive a new lexical item of a different word-class or mark grammatical information. The most frequent form of reduplication is full reduplication. However, partial reduplication is also found in casual speech in forms like *ki-kineh* vs. *kineh-kineh* ‘in that manner’. In the following sections, I discuss reduplication of nominal roots, adjectival roots and verbal roots.

2.4.1.4.1 Reduplication of Nominal Roots

Nominal roots can be reduplicated to derive verbal predicates or to indicate plurality. The first function of nominal reduplication is to create stative predicates. This is particularly common with weather-related words:

- (94) **Deriving Verbs from Nouns**
- | | | | |
|----|---------------------------------|---|---|
| a. | <i>laput</i> ‘cloud’ | → | <i>laput-laput</i> ‘to be cloudy’ |
| b. | <i>legkuq</i> ‘thunder’ | → | <i>legkuq-legkuq</i> ‘to thunder’ |
| c. | <i>belal</i> ‘sheet lightening’ | → | <i>belal-belal</i> ‘to have lightening’ |
| d. | <i>bariu</i> ‘wind’ | → | <i>bariu-bariu</i> ‘to be windy’ |

In (94), reduplication plays a similar role to *-em-* infixation (SUBSECTION 2.4.1.2.1):

- (95) **Function of reduplicated nouns**
- a. Laput~laput tieh edto kinih.
 REDUP~cloud PT=3SG.1 day now
 ‘It’s cloudy today.’
 (elicitation, BAR28102013CH_01 00:04:35.227-00:04:38.970)
- b. Laput tieh edto kinih.
 <INTR>cloud PT=3SG.1 day now
 ‘It’s cloudy today.’ (elicitation, fieldnotes)

It is not clear if there is a difference between (95a) and (95b). Both processes regularly derive predicates.

Secondly, like in other Western Austronesian languages, reduplication can be used to signal plurality (Blust 2013). Non-reduplicated nouns can also have plural referents. However, reduplicated nouns are specified or emphasised as plural. Examples of reduplicated plurals from the text corpus are given in (96):

- (96) **Reduplicated Plurals**
- a. Tak narih ni’er [...] gerai~gerai nuk senaruq ih keyh...
 if IMPERS AV.see REDUP~stall REL UV.PFV.do PT PT
 ‘If I look at the stalls they put up...’
 (text, BAR02082014CH_02 00:00:56.570-00:01:02.150)
- b. Mulaq na’an~na’an buaq lem kebun nedih.
 many REDUP~type fruit in garden 3SG.POSS
 ‘There were many different types of fruit in her garden.’
 (text, PDA10112013CH_01 00:00:44.470-00:00:47.590)

2.4.1.4.2 Reduplication of Adjectival Roots

Adjectives (and some quantifiers) can be reduplicated to derive adverbs:

- (97) **Deriving Adverbs from Adjectives**
- | | | | |
|----|-----------------------|---|--------------------------------------|
| a. | <i>dooq</i> ‘good’ | → | <i>dooq-dooq</i> ‘well’ |
| b. | <i>saget</i> ‘fast’ | → | <i>saget-saget</i> ‘quickly’ |
| c. | <i>beruh</i> ‘new’ | → | <i>beruh-beruh</i> ‘recently’ |
| d. | <i>tu’uh</i> ‘real’ | → | <i>tu’uh-tu’uh</i> ‘really/properly’ |
| e. | <i>pu’un</i> ‘first’ | → | <i>pu’un-pu’un</i> ‘firstly’ |
| f. | <i>mulaq</i> ‘many’ | → | <i>mulaq-mulaq</i> ‘a lot’ |
| g. | <i>si’it</i> ‘few’ | → | <i>si’it-si’it</i> ‘a little’ |
| h. | <i>muneng</i> ‘close’ | → | <i>muneng-muneng</i> ‘close-by’ |
| i. | <i>mado</i> ‘far’ | → | <i>mado-mado</i> ‘far away’ |

As adverbs, the reduplicated forms modify verbs, as shown in (98b) (SUBSECTION 2.4.2.4). In contrast, as adjectives, the non-reduplicated form either modifies a noun or functions as an adjectival predicate, as in (98a):

- (98) **Function of Reduplicated Adjectives**
- a. Tu’uh tineh.
 true PT=DEM
 ‘That’s true.’
 (text, BAR08092014CH_05 00:08:03.330-00:08:04.170)
- b. Naruq tu’uh~tu’uh narih keyh.
 AV.do REDUP~true IMPERS PT
 ‘You have to work hard.’
 (text, BAR08092014CH_04 00:01:22.190-00:01:23.510)

Secondly, reduplicated adjectives can have an emphatic or intensifying function:

- (99) **Intensification**
- a. Kuman nuk kenen nuk dooq~dooq ih.
 AV.eat REL UV.IRR.eat REL REDUP~good PT
 ‘Eat food that is very good.’
 (text, BAR08092014CH_04 00:09:58.005-00:10:00.425)

In (99), the reduplicated form remains an adjective, modifying the head noun *nuk kenen* ‘food’ rather than the predicate *kuman* ‘eat’.

2.4.1.4.3 Reduplication of Verbal Roots/Stems

Finally, reduplication of verbal roots and stems typically has an inflectional function, indicating progressive or imperfective aspect. The exact interpretation differs, depending on whether a bare root is reduplicated, or a nasalised stem.

In some cases, reduplication of bound bare roots is the only way in which they can be used predicatively. Typically, the reduplicated predicate implies a sense of non-seriousness or lack of success/completion of the action:⁹⁵

(100) **Non-serious Action**

- a. Si'er~si'er tupu tuih ngedeh.
 REDUP~see only PT=1SG.1 to.3PL
 'I was just looking over at them (I didn't get a proper look).'
- (elicitation, fieldnotes)
- b. Kiding~kiding tieh ngen nuk midih let ngineh.
 REDUP~lift PT=3SG.1 with things from there
 'He's just lifting things from there (with no particular aim).'
- (elicitation, fieldnotes)

Reduplication is also used for inherently reciprocal events:

(101) **Inherent Reciprocity**

- a. Siwa~siwa teh diweh ngen sapaq diweh ih.⁹⁶
 REDUP~exchange PT 3DU with shirt 3DU PT
 'They exchanged shirts.'
- (elicitation, fieldnotes)

⁹⁵ In some cases, this is also true of reduplicated nasalised stems:

- (i) Mey ngelulut~ngelulut kinéh kekuh adiq dooq teh mängey ngidih.
 go REDUP~AV.beat like.that say.1SG.2 but good PT fun with.DEM
 'We were just beating in time like that, but we had fun with it.'
- (text, BAR02082014CH_02 00:10:07.030-00:10:11.050)

Note that Formosan languages typically only allow reduplication of two syllables and consequently forms like Kelabit *ngelulut-ngelulut* 'beat (repeatedly/idly)' would not be possible (cf. Zeitoun & Wu 2006). However, similar patterns are found in more innovative Austronesian languages like Indonesian, where morphologically complex stems such as *menarik* (*meN-* + *tarik*) 'AV.pull' can be reduplicated to form *menarik-menarik* 'pulling (iteratively)' (see Mistica et al 2009). This may be one way in which Kelabit is similar to Indonesian-type languages (see CHAPTER 3-5 for further comparison).

⁹⁶ *siwa* 'swap' does occur in its un-reduplicated form. However, this could be analysed as a noun, since it occurs as the complement of the preposition *koq* 'for'.

Finally, reduplication of verbal roots can signal an ongoing or progressive event:

- (102) **Progressive aspect**
 a. Keteng kuel~kuel teh u'et dih lem container.
 still REDUP~move PT worm DEM in container
 'The worms are still wriggling around in the container.'
 (text, BAR02082014CH_02 00:04:44.915-00:04:48.025)

The progressive interpretation is supported by the fact that the reduplicated predicate appears with the progressive auxiliary *keteng* 'still' (SUBSECTION 2.4.2.6). However, much like plurality of nouns, reduplication is not required to signal progressive aspect, as this can also be indicated using irrealis voice markers and auxiliaries (see SUBSECTION 2.4.1.3.1 and SUBSECTION 2.4.2.6).

When a nasalised stem is reduplicated, this may signal an ongoing or intensive event:

- (103) **Present Tense/Ongoing Action**
 a. Lit tuih ni'er ngegkang~ngegkang selapang
 suddenly PT=1SG.1 AV.see REDUP~AV.lift gun

 nedih koq.
 3SG.POSS PT
 'All of a sudden I saw he was lifting up his rifle.'
 (text, PDA10112013CH_02 00:00:53.030-00:00:56.440)

Similarly, it can represent habitual action:

- (104) **Habitual Action**
 a. Edteh la'ih edteh nuk masiu~masiu luang.
 one man one REL REDUP~AV.sell fish
 'A man who sells fish/a fisherman.'
 (text, BAR21082014CH_01 00:09:00.707-00:09:03.061)

Hence, though the effect of reduplication can be loosely defined as marking progressive aspect, the aspectual interpretation differs, depending on whether the

nasalised or non-nasalised root is reduplicated. All of these forms are lower in transitivity than non-reduplicated, perfective forms (see CHAPTER 3).

2.4.2 Major Word Classes

In the final section on morphology, the major word classes in Kelabit are described and distinguished from one another. As shown in SUBSECTION 2.4.1, Kelabit roots can often derive either verbal or nominal forms under productive derivational processes. This is common in Western Austronesian languages, including Seediq (Tsukida 2005), Tagalog (Himmelman 2005b), Iloko (Rubino 2005), Kimaragang (Kroeger 2005) and Javanese (Hemmings 2012). Moreover, adjectives are similar to stative verbs, as shown in SUBSECTION 2.4.2.3. In Philippine-type languages, these properties have led to the proposal that roots are pre-categorial and that there is no distinction between nouns and verbs in the lexicon (Foley 2008). However, I follow Kroeger (1998b) and Himmelman (2008) in suggesting that word classes can be distinguished in Kelabit on the basis of functional and distributional factors down to the level of the root.⁹⁷

The major lexical word classes in Kelabit are nouns, verbs, adjectives and adverbs. In addition, there are closed functional classes, including prepositions, auxiliaries, deictic terms, pronouns, interrogatives, relativisers, conjunctions, numerals, quantifiers and particles. As discussed in SUBSECTION 2.3.3, lexical classes

⁹⁷ The debate can be understood as follows. Since morphology appears to be able to attach to roots of different kinds and derive stems of different kinds, it follows that either there is no syntactic distinction between roots, or the morphology is simply multifunctional and can have both inflectional and derivational uses depending on the root that it attaches to (see Crouch 2009 for further discussion of the precategoriality debate and SUBSECTION 2.4.1 on the inflection vs derivation debate). It is theoretically possible that roots could be precategorial and that inflected or derived words could nonetheless have different classes on the basis of syntactic distribution. Hence, the central debate is whether roots of different classes can also be distinguished. Kroeger's (1998) main argument for assuming this in Tagalog is that voice morphology obligatorily attaches to verbal roots with systematic interpretations, whereas they only optionally attach to nominal roots with highly variable semantics. Similar arguments can be made for Kelabit, as can be seen from the varied uses of affixes in SUBSECTION 2.4.1. Full discussion of precategoriality is beyond the scope of this thesis and remains for future research.

are mostly bisyllabic, whereas functional classes are often monosyllabic. Defining characteristics of each are discussed in turn.

2.4.2.1 Nouns

Nouns in Kelabit cannot be identified on the basis of grammatical categories, such as number, gender or case. These are not expressed through morphological inflection but determined through context.⁹⁸ Equally, nominalising suffixes like *pe-* and *-an* also form verbal predicates, as described in SUBSECTION 2.4.1. However, nouns can be identified through their function and distribution.

In terms of function, nouns and NPs are typically the subject or non-subject core arguments of a clause (SUBSECTION 2.5.1):

(105) a. **Subject**

[Edteh anak] ne-ni'er uih.
 one child PFV-AV.see 1SG.1
 'A child saw me.'

b. **Non-Subject Core**

Uih ne-ni'er [edteh anak].
 1SG.1 PFV-AV.see one child
 'I saw a child.'

(elicitation, BAR30072014CH_03 00:01:27.090-00:01:33.713)

Nouns can also function as prepositional objects and nominal predicates:

(106) **Object of a PP**

a. La'ih sineh ne-merey nubaq [ngen [anak nedih]].
 man DEM PFV-AV.give rice to child 3SG.POSS
 'That man gave rice to his child.'

(elicitation, BAR30072014CH_03 00:02:25.520-00:02:31.350)

⁹⁸ One exception is the word *anak* 'child', which does have a dedicated plural form *anak-adiq*, lit. 'child-small'.

b. **Nominal Predicate**

[Anak iih] sineh?
child who DEM
'Whose child is that?'

(elicitation, BAR30072014CH_03 00:03:40.870-00:03:42.400)

As complements, they typically follow the head verb or preposition. As predicates, they typically occur clause-initially. This suggests that Kelabit is head-initial (Polinsky 2012).

Nouns can be modified by numerals, quantifiers, adjectives, relative clauses, demonstratives and possessive pronouns. They can also be modified by the particle *ih*, a reduced form of the medial demonstrative *idih* (see SUBSECTION 2.4.2.7). This functions as a marker of definiteness or specificity. Numerals and quantifiers precede the noun, whilst adjectives, relative clauses, possessors and determiners follow:

(107) a. **Numerals**

edteh anak
one child
'one child'

(elicitation, BAR30072014CH_03 00:03:44.680-00:03:45.740)

b. **Quantifiers**

mulaq anak-adiq
many child-PL
'many children'

(elicitation, BAR30072014CH_03 00:03:52.240-00:03:53.730)

c. **Adjectives**

anak i'it
child small
'small child'

(elicitation, BAR30072014CH_03 00:03:49.890-00:03:51.050)

d. **Relative Clauses**

anak [suk muit bakul]
child REL AV.carry basket
'the child who took the basket'

(text, BAR03082014CH_02 00:00:45.330-00:00:47.940)

e. **Demonstratives**

anak sinih
child DEM
'this child'

(elicitation, BAR30072014CH_03 00:04:19.940-00:04:20.830)

f. **Possessive Pronouns**

anak kudih
child 1SG.POSS
'my child'

(elicitation, BAR30072014CH_03 00:04:21.810-00:04:22.870)

Although numerals and quantifiers typically precede the noun, they can also be used referentially. In these cases, they follow the noun like adjectives. The contrast is illustrated in (108):

(108) **Indefinite Referent**

- a. Kadiq ieh ne-ngatey [**duweh** anak let England].
because 3SG.1 PFV-AV.kill two child from England
'He killed two boys from England.'
(text, BAR21082014CH_01 00:09:04.689-00:09:10.796)

b. **Definite Referent**

[lun **duweh** nuk natey neh] ...
people two REL UV.PFV.kill 3SG.2
'the two people that he killed...'
(text, BAR21082014CH_01 00:09:14.473-00:09:15.951)

c. **Definite Referent**

[anak-adiq **mulaq** nuk ineh]
child-PL many REL DEM
'those many children'
(text, BAR30072014CH-03 00:04:01.150-00:04:02.950)

In (108a), the numeral precedes the head noun and serves to introduce an indefinite NP into the discourse. In (108b) and (108c), however, the referents are given in discourse and the numerals and quantifiers follow the head noun.

The preferred word order within the NP is N – Adj – Poss – Rel Clause – Dem:

- (109) **Word Order in the NP**
- | | | | | | | | |
|----|-----------|------------|-------|-------------|------|-------------------|------------|
| a. | buah | kaber | birar | kudih | [nuk | mulaq] | nih |
| | pineapple | yellow | | 1SG.POSS | REL | many | DEM |
| | N | Adj | | Poss | | Rel Clause | Dem |
- ‘these many yellow pineapples of mine’
 (elicitation, BAR21102013CH_01 01:43:00.692-01:43:05.230)

Demonstratives also follow relative clauses in Western Austronesian languages like Javanese (Hemmings 2012).

2.4.2.2 Verbs

Of all word classes, verbs are the most morphologically complex and can often be identified on the basis of verbal morphology, as discussed in SUBSECTION 2.4.1. Kelabit verbs can be subdivided into intransitive, transitive and ditransitive verbs, based on the number of arguments that they take. Many verbs can be used both transitively and intransitively, as in (110), and pro-drop is common in discourse:

- (110) **Intransitive**
- | | | | |
|----|-------|-------|--------|
| a. | Na'am | uih | keliq. |
| | NEG | 1SG.1 | know |
- ‘I don't know.’ (one argument)
- Transitive**
- | | | | | |
|----|-------|-------|-------|-------|
| b. | Na'am | uih | keliq | ieh. |
| | NEG | 1SG.1 | know | 3SG.1 |
- ‘I don't know her.’ (two arguments) (elicitation, fieldnotes)

2.4.2.2.1 Intransitive verbs

Many intransitive predicates are bare roots. This includes unergative predicates, whose single argument acts volitionally and actively, and unaccusative predicates, whose single argument undergoes the action/state of the predicate (cf. Perlmutter 1978).⁹⁹

(111)	Unergative	Unaccusative
a.	<i>rudap</i> ‘sleep’	<i>terem</i> ‘sink’
b.	<i>lawey</i> ‘walk’	<i>tutuq</i> ‘fall down’
c.	<i>upun</i> ‘run’	<i>lubid</i> ‘roll’
d.	<i>uput</i> ‘jump’	<i>lukaq</i> ‘fall over’
e.	<i>riek</i> ‘cough’	<i>bilaaq</i> ‘break’
f.	<i>riruh</i> ‘laugh’	<i>teneb</i> ‘cold’
g.	<i>bebpaq</i> ‘urinate’	<i>pering</i> ‘dry’

Intransitive predicates can also be derived from non-verbal roots by the addition of the *-em-* infix, *N-* prefix or *te-* and *pe-* stem-forming prefixes, as outlined in SUBSECTION 2.4.1.

Many intransitive predicates undergo a causative/transitive alternation through the addition of the nasal *N-* prefix. This applies to unaccusative bare roots and stems derived through *-em-* infixation:

(112)	Causative Alternation with Bare Roots
a.	Ne-terem ieh lem ebpaq ih. PFV-sink 3SG.1 in river PT ‘He sunk in the river.’ (elicitation, BAR28102013CH_03 00:00:58.868-00:01:04.084)
b.	Uih ne-nerem ieh lem ebpaq ih. 1SG.1 PFV-AV.sink 3SG.1 in river PT ‘I dunked him in the river.’ (elicitation, BAR28102013CH_03 00:03:34.542-00:03:38.447)

⁹⁹ Some predicates seem to be understandable as either unergative or unaccusative – e.g. *tudo* ‘sit’, which means both ‘sit down’ and ‘be seated’.

(113) **Causative Alternation with *-em-* Verbs**

- a. M-eseb neh uduh nuk ngi iring dalam sineh.
INTR-burn PT grass REL at near road DEM
'The grass next to the road is burning.'

(elicitation, BAR17102013CH_01 00:16:49.050-00:16:56.040)

- b. La'ih suk lem latiq ih ne-ng-eseb uduh.
man REL in farm PT PFV-AV-burn grass
'The farmer burnt the grass.'

(elicitation, BAR17102013CH_01 00:20:11.783-00:20:18.716)

The alternations are similar to inchoative/causative alternations (cf. Haspelmath 1993). The equivalent transitive form of each of the verbs expresses an event causing either a change-of-state or a change-in-location.

Unergative predicates typically do not undergo the causative alternation with *N-*:

(114) **No Causative Alternation**

- a. **ngupun* 'run'
b. **ngelawey* 'walk'
c. **nguput* 'jump'

In fact, a small subset of unergative predicates are formed through nasal prefixation

(SUBSECTION 2.4.1.1.4):

(115) **Intransitives with *N-* morphology**

- a. *dalan* 'road' → *nalan* 'to walk'
b. *tangey* 'cry' → *nangey* 'to cry'
c. *arang* 'dance' → *ngarang* 'to dance'
d. *utaq* 'vomit' → *ngutaq* 'to vomit'

Some unergative predicates undergo a causative alternation with the prefix *pe-* plus voice marking (see SUBSECTION 2.4.1.1.7). This differs from the other causative alternations in the sense that the change is not irreversible or completed:¹⁰⁰

(116) **Causative Alternation with *pe-***

- a. Ne-rudap uih medto na'ah.
 PFV-sleep 1SG.1 afternoon earlier
 'I slept earlier this afternoon.'
- b. Nih tesineh nedih me-rudap anak nedih.
 DEM mother 3SG.POSS AV.CAUS-sleep child 3SG.POSS
 'The mother is putting her child to sleep.'
 (elicitation, BAR30072014CH_03 01:09:00.344-01:09:07.933)

There are several verbs that can take either the nasalised causative or the *pe-* causative:

(117) **Causatives with *pe-* and *N-***

- a. *tudo* 'sit' → *metudo* → *nudo*
 b. *tu'i* 'get up' → *metu'i* → *nu'i*
 c. *turun* 'down' → *meturun* → *nurun*

The *pe-* causative seems to imply indirect causation, whilst the nasal causative implies direct causation:

(118) **Direct vs Indirect Causation**

- a. Me-tudo diweh ngineh.
 AV.CAUS-sit 2DU there
 'Show them to their seats over there.'
- b. Nudo diweh ngineh.
 AV.sit 2DU there
 'Seat them there.'
 (elicitation, BAR30072014CH_03 00:29:33.960-00:29:37.216)

¹⁰⁰ A similar contrast was seen in the behaviour of causative predicates in Javanese (cf. Hemmings 2012). The meanings are not necessarily predicatable, i.e. *menalan* 'manage/cause to run'.

Hence, intransitive predicates have a single argument. Depending on whether they are unergative or unaccusative, they may also be identified by verbal morphology and whether they undergo causative alternations.

2.4.2.2.2 Transitive verbs

Transitive verbs take two core arguments and are distinguished by their voice-marking: *N-* or *neN-* for AV, *-in-* or *-en* for UV. This can be illustrated below:

(119) **Actor Voice**

- a. Ne-kuman nubaq uih.
 PFV-AV.eat rice 1SG.1
 ‘I ate rice.’

(elicitation, BAR30072014CH_03 01:10:31.920-01:10:33.880)

Undergoer Voice

- b. Kinan kuh nubaq ih.
 UV.PFV.eat 1SG.2 rice PT
 ‘I ate rice.’

(elicitation, BAR30072014CH_03 01:12:11.900-01:12:13.570)

In addition, transitive predicates may be marked with the abilitative prefix *ke-*, reciprocal *pe-*, reflexive *pere-* and the accidental prefix *ne-*, described in SUBSECTION 2.4.1.

There are a few bare predicates that do not take morphological marking but are used (ambi-)transitively. These include:

(120) **Bare Transitive Predicates**

- a. *keliq* ‘know’
 b. *kelupan* ‘forget’
 c. *sekenan* ‘remember’
 d. *raut* ‘play’
 e. *uwan* ‘have/own’

As shown in SUBSECTION 2.5.3, these are not subject to the same restrictions as voice-marked transitive predicates.¹⁰¹

2.4.2.2.3 Ditransitive verbs

Ditransitive verbs require three arguments: typically an actor, an undergoer and a benefactive/goal. In Kelabit, predicates such as ‘give’ and ‘show’ encode the benefactive/goal argument as a prepositional phrase in both AV and UV constructions:

(121) **Actor Voice**

a. *Uih merey anak nubaq.
 1SG.1 AV.give child rice
 For: ‘I give the child rice.’

b. Uih merey nubaq ngen anak.
 1SG.1 AV.give rice to child
 ‘I give rice to the child.’

(elicitation, BAR30072014CH_04 00:05:33.390-00:05:45.790)

Undergoer Voice

c. *Bilih kuh ieh nubaq.
 UV.PFV.buy 1SG.2 3SG.1 rice
 For: ‘I bought her rice.’

d. Bilih kuh nubaq ngeneh.
 UV.PFV.buy 1SG.2 rice for.3SG.2
 ‘I bought rice for her.’

(elicitation, BAR30072014CH_04 00:09:24.741-00:09:33.432)

However, the instrumental voice appears to involve a ditransitive construction with three nominal arguments expressed without prepositional phrases:

¹⁰¹ It is possible that *kelupan* and *sekenan* are affixed forms. It is not clear if and how frequently potential roots are used.

- (122) **Instrumental Voice**
- a. [Seduk] penekul [la'ih sineh] [nubaq nedih].
 spoon IV-spoon up man DEM rice 3SG.POSS
 'That man used a spoon to spoon up his rice.'
 (elicitation, BAR30072014CH_04 00:10:44.351-00:10:47.013)

Hence, the instrumental voice, like equivalent voices in Philippine-type languages, is applicative-like as it triggers a change in valency (see SUBSECTION 1.3.1).

This also suggests that IV is not a nominalised construction, since there are two nominal arguments following the predicate. Moreover, the actor cannot be replaced with a possessive pronoun:

- (123) **Against a Nominal Analysis of IV**
- a. *Seduk penekul nedih nubaq nedih.
 spoon IV-spoon up 3SG.POSS rice 3SG.POSS
 For: 'the spoon was his implement for the spooning up of rice.'
 (elicitation, fieldnotes)

Hence, the nominalisation hypothesis does not extend to Kelabit instrumental voice (see Kaufman 2009).

2.4.2.2.4 Distributional Characteristics

All verbs can be identified on the basis of distributional and functional criteria. The typical function of a verb is as a predicate and, as such, verbs often appear in initial position. However, word order differs according to the voice construction, as discussed in CHAPTER 5.

Verbs typically take nominal arguments, though zero anaphora is possible. They can be optionally modified by pre-verbal auxiliaries and adverbs:

- (124) **Preverbal Auxiliaries**
- a. Laq ngiup apui uih.
 DESID AV.blow fire 1SG.1
 ‘I’d like to blow the fire.’
 (elicitation, BAR14102013CH_01 01:36:57.770-01:37:00.030)
- b. **Adverbs**
 Senu’i neh bicycle nedih [dooq~dooq].
 UV.PFV.get up 3SG.2 bicycle 3SG.POSS REDUP~good
 ‘He propped his bicycle up nicely.’
 (pear story, BAR31072014CH_06 00:04:50.540-00:04:53.950)

More information on auxiliaries and adverbs can be found in SUBSECTIONS 2.4.2.6 and 2.4.2.4. As discussed in SUBSECTION 2.4.1.4.2, *dooq-dooq* functions as an adverb in (124b), since it modifies the verb and appears following the possessive suffix, unlike adjectives (see SUBSECTION 2.4.2.1).

2.4.2.3. Adjectives

Adjectives in Kelabit are similar to intransitive verbs, particularly when used predicatively. Nonetheless, there are some differences in distribution and function that identify a class of adjectives. For example, adjectives can be modified by (superlative) degree modifiers such as *leng-leng* and *pelaba*, which precede the adjective, and *tungen-tungen*, *tebuut* and *ketuh* which follow it:

- (125) **Degree Modifiers**
- a. leng~leng dooq
 REDUP~very good
 ‘very very good’
 (elicitation, BAR15102013CH_01 00:09:46.320-00:09:47-510)
- b. pelaba¹⁰² dooq
 very good
 ‘very good’
 (elicitation, BAR30072014CH_04 00:28:44.524-00:28:45.110)

¹⁰² From *laba* ‘to pass’ and *pe-* indicating position/perfectivity.

- c. dooq tungen~tungen
 good REDUP~very
 ‘very very good’
 (text, BAR21082014CH_07 00:05:13.440-00:05:14.420)
- d. dooq ketuh
 good most
 ‘extremely good’
 (text, BAR04092014CH_02 00:09:22.990-00:09:23.610)

There is no morphological comparative. Instead, the comparative is formed by expressing the contrast using *let ngen* ‘from’ and optionally the adverbial *ked i’t*.¹⁰³

- (126) **Comparative**
- a. Anak sinih dooq (ked i’t) let ngen anak sinih.
 child DEM good (more) from to child DEM
 ‘This child is better than that one.’ (elicitation, fieldnotes)

A superlative reading is created through adjectival modification with *pelaba* or *ketuh* etc.:

- (127) **Superlative**
- a. [Pelaba dooq] teh anak sinih.
 really good PT child DEM
 ‘This child is the best/really good.’
 (elicitation, BAR30072014CH_04 00:28:44.520-00:28:46.150)

Thus, typical grammatical categories associated with adjectives are not grammaticalised as morphological inflection, but established through context or through the construction.

There are a few adjectives that have different variants depending on whether they modify singular or plural nouns:

¹⁰³ *ked i’t* can mean ‘a little’ or ‘a little while’ but emphasises the comparison in a comparative construction.

(128)	Singular Referent	Plural Referent	
a.	<i>i'it</i>	<i>madiq</i>	‘small’
b.	<i>rayeh</i>	<i>merar</i>	‘big’

In terms of function, adjectives can either modify nouns or form adjectival predicates. As modifiers, adjectives follow the head noun and precede demonstratives (SUBSECTION 2.4.2.1). As predicates, the adjective is typically initial and often separated from the subject by the particle *teh*, as seen in (127).

It is sometimes difficult to distinguish between adjectives and intransitive verbs. For example, some intransitive predicates can also be modified by degree modifiers:¹⁰⁴

- (129) **Degree modifiers with *-em-* verbs**
- a. pelaba m-udan
 very INTR-rain
 ‘very rainy’
 (elicitation, BAR28102013CH_01 00:40:08.440-00:40:09.690)
- b. m-ileh ketuh
 INTR-knowledge very
 ‘very good’ (at something)
 (text, BAR21082014CH_09 00:09:34.310-00:09:34.840)

However, when we look at adjectives in attributive function, we see that verbs cannot always fill the same position:

- (130) **Adjective**
- a. Dooq pian kuh ngen [rumaq **rayeh** sineh].
 good want 1SG.2 to house big DEM
 ‘I like that big house.’
 (elicitation, BAR30072014CH_04 00:38:18.870-00:38:22.730)

¹⁰⁴ But not others:

- (i) *pelaba ngarang
 very AV.dance
 For: ‘very dancy’ (elicitation, BAR30072014CH_04 00:35:14.209-00:35:19.076)

Intransitive Verb

- b. *Dooq pian kuh ngen [la'ih **ngarang** sineh].
good want 1SG.2 to man dance DEM
For: 'I like that dancing man.'
(elicitation, BAR30072014CH_04 00:38:52.830-00:38:57.120)

Consequently, though distinguishing between intransitive verbs and adjectives is sometimes difficult, a class of adjectives can be distinguished from both transitive and intransitive verbs on account of distribution, function and modification patterns.

Finally, nouns can be derived from adjectives using *ken* or the reduced prefix *ke-*:

(131) Deriving Nouns from Adjectives

- a. *rayeh* 'big' → *ken rayeh* 'size'
b. *mado* 'far' → *ken mado/kemado* 'distance'
c. *ditaq* 'high' → *ken ditaq* 'height'
d. *beneh* 'low' → *ken beneh* 'depth'

2.4.2.4 Adverbs

The final open word-class in Kelabit is adverbs. They are often formed by reduplicating adjectival roots (SUBSECTION 2.4.1.4.2) and function to modify verbs:

(132) Adverbs

- a. Kuman dooq~dooq!
AV.eat REDUP~good
'Eat well/eat properly/eat a lot!'
(elicitation, BAR30072014CH_04 00:40:08.350-00:40:09.760)
- b. Kuman saget~saget!
AV.eat REDUP~fast
'Eat quickly!'
(elicitation, BAR30072014CH_04 00:40:12.350-00:40:13.790)

The position of adverbs is not fixed. They can occur after the verb, as in (132), or clause-initially, as in (133):

- (133) a. **Adverb Position**
 [Mawer~mawer] narih nalan!
 REDUP~quick IMPERS walk
 ‘Walk quickly!’
 (elicitation, BAR30072014CH_04 00:41:42.120-00:41:44.131)

Other, non-reduplicated adverbs include:

- (134) **Non-reduplicated adverbs**
- a. *dadan* ‘for a long time’
 - b. *setengen/temengen* ‘on purpose’
 - c. *lit* ‘suddenly’
 - d. *na’an* ‘later’
 - e. *na’ah* ‘earlier’
 - f. *kedu’it* ‘for a short while’
 - g. *terun* ‘perhaps’¹⁰⁵

2.4.2.5 Prepositions

It is sometimes difficult to differentiate between prepositions and verbs or determiners in Austronesian languages (Starosta 2009e: 288). Nonetheless, a class of prepositions can be identified in Kelabit on the basis of form, function and distribution. The main prepositions are summarised in TABLE 2.10:

¹⁰⁵ This can be used to express uncertainty in a proposition and often occurs clause-finally. Its use is particularly associated with the dialect of Kelabit spoken in Long Lellang.

Table 2.10 Prepositions in Kelabit

Form	Meaning
ngi	at
ngen	to (recipient)/for (benefit of)/with (instrument)
lem	in
let	from
luun	on top of
koq and kayuq	into/like
kenuwan	for (someone to have)
ruyung	with (comitative)
mey	to (location)
iring	beside
lem erang	between
liang	underneath
mayaq	with (following)/ by (transport)

Prepositions can be combined. For example, *let ngen* expresses the idea ‘from someone’ as opposed to *let* ‘from somewhere’.¹⁰⁶ They are not inflected and are typically, though not exclusively, monosyllabic. In this way, they differ from nouns, verbs and other lexical classes.

In terms of distribution and function, they head prepositional phrases and take nominal complements:

(130) **Function of prepositions**

- a. Lem edteh edto ieh mala [ngen [lemulun]]...
on one day 3SG.1 AV.say to people
‘One day he said to the people...’

(text, BAR25102013CH_01 00:00:23.740-00:00:31.390)

PPs usually encode obliques or adjuncts of time and place, as well as causes, sources and instruments etc. The preposition precedes its nominal complement and is another indication that Kelabit is head-initial.

¹⁰⁶ Similarly, *let lem* ‘from inside’, *let luun* ‘from on top’, *let liang* ‘from underneath’ etc.

Finally, prepositions can be modified by adverbs, such as the following:

(135) **PP Modifiers**

- a. *siri-siri* ‘straight’
- b. *su* ‘directly’

- (136) a. [su lem batek nedih]_{PP}
 directly in stomach 3SG.POSS
 ‘straight into its stomach’

(text, PDA10112013CH_01 00:07:55.850-00:07:57.180)

Although a class of prepositions can be identified, it is possible that they are derived from verbs and determiners. For example, *kenuwan* ‘for’ appears to take the verbal infix *-in-* (see SUBSECTION 2.4.1.2.3). Similarly, *mayaq* ‘with/by’ can be used as a main verb meaning ‘to follow’. On the other hand, *ngi* ‘at’ can function in a similar manner to other deictic terms (see SUBSECTION 2.4.2.7). Detailed analysis of prepositions remains for future research.

2.4.2.6 Auxiliaries

In Kelabit, there are verbal auxiliaries that express aspectual and modal information, as is common in the languages of Sarawak (Clayre 2002). They can be identified according to their function and pre-verbal position.¹⁰⁷ They include:

¹⁰⁷ It remains to be analysed in more detail whether these so-called auxiliaries really constitute a separate word class or are simply control predicates that take clausal complements (see also Starosta 2009e: 279 for similar discussion in relation to Formosan languages). The analysis of auxiliaries as functional phrase structure heads is supported by the fact that they can stand alone as sentence fragments:

- | | | | | | | | |
|-----|-----|------------------|-------|-------|-----|---------------|---------------------------|
| (Q) | Ken | kereb | Peter | tudo? | (A) | kereb | |
| | Q | can | Peter | sit | | can | |
| | | ‘Can Peter sit?’ | | | | ‘Yes, he can’ | (elicitation, fieldnotes) |

However, it is not clear if this prohibits an analysis of auxiliaries as verbal. Moreover, the typical word-order pattern following auxiliaries is Aux Subj Verb, which is also true of control predicates (SUBSECTION 2.5.3.3.). As discussed below, it is possible that the forms discussed in this section are in the process of grammaticalising from main verb to functional item. Auxiliaries can seemingly combine, as in example (142a). Exact patterns of combination remain to be explored.

(137)	Auxiliary	Function
a.	<i>(e)laq/pian</i>	desiderative/future
b.	<i>daraq/di'eyq</i>	negative desiderative
c.	<i>kereb</i>	potential
d.	<i>mileh</i>	ability
e.	<i>keteng</i>	durative
f.	<i>pengeh</i>	completive
g.	<i>lem puket</i>	progressive
h.	<i>murih</i>	habitual
i.	<i>mey</i>	future

Both *laq* and *pian* imply desiderative mood. They take verbal and prepositional complements, which suggests they might be in the process of grammaticalising from main verb to auxiliary.¹⁰⁸

- (138) **Desiderative Auxiliaries**
- a. Dooq pian uih [mirup kupa]_{VP}.
good want 1SG.1 AV.drink coffee
‘I like to drink coffee.’
(elicitation, BAR14102013CH_02 00:11:49.775-00:11:53.845)
- b. Dooq pian uih [ngen kupa]_{PP}.
good want 1SG.1 to coffee
‘I like coffee.’
(elicitation, BAR14102013CH_02 00:11:54.700-00:11:58.460)
- c. Laq teh keduih [mey m-uliq]_{VP}.
DESID PT 1SG.EMPH go INTR-back
‘I’d like to go home.’
(elicitation, BAR14102013CH_02 00:10:06.150-00:10:08.220)
- d. Laq uih [ngen idih]_{PP}.
DESID 1SG.1 to DEM
‘I’d like that.’
(text, PDA06112013CH_04 00:00:22.740-00:00:23.680)

¹⁰⁸ It is unclear if forms like *dooq pian* and *laq* are verbs. As discussed in SUBSECTION 2.4.2.6, verbs typically take intransitive or transitive voice marking. These forms appear nominal, or perhaps adjectival in the case of *dooq pian*, and could be understood as ‘my wish’ or ‘my desire’ rather than ‘I want’.

In (138b) and (138d), *laq* and *dooq pian* are the main predicates and take a subject argument and an oblique, expressed as a PP. In (138a) and (138c), however, *laq* and *dooq pian* take VP complements: *mey muliq* ‘go home’ and *mirup kopi* ‘drink coffee’. The subject appears between the auxiliary and the verb (see CHAPTER 5).

In some cases, *laq* and *pian* also take a complement clause (SUBSECTION 2.5.3.3):

- (139) **Auxiliaries as Control Predicates**
 Uih pian [ngeneh nekul nubaq].
 1SG.1 want to.3SG.2 AV.spoon rice
 ‘I want him to spoon up the rice.’ (elicitation, fieldnotes)

The structure in (139) appears biclausal, since the lower clause begins with the preposition *ngen* ‘with’, which acts as a complementiser (SUBSECTION 2.5.3.3). In contrast, the clauses in (138) do not appear biclausal, and the auxiliary and verb can form a single constituent:

- (140) **Auxiliary + VP**
 Neh tebeyq Peter [laq kuman buaq kaber ih].
 DEM PT Peter DESID AV.eat fruit pineapple PT
 ‘Peter would like to eat pineapple.’
 (elicitation, BAR19082014CH_03 00:41:39.310-00:41:41.540)

Hence, the auxiliaries appear to be used as predicates that take both PP and clausal complements, as well as auxiliaries that take VP complements.

As well as expressing desiderative mood, *laq* can denote other irrealis functions, such as future action or possibility. The exact semantics remains to be analysed in more detail, but some examples are given in (141) (see also SUBSECTION 2.5.3.3):

- (141) **Irrealis with *laq***
- a. Leng~leng elaq teh narih ngen anak~anak nuk
 REDUP~very DESID PT IMPERS to REDUP~child REL
 mudtih riak elaq ninger cerita nuk kayuq inih lah.
 last future ? AV.hear story REL like DEM PT
 ‘I really hope that the children of the future will listen to stories like
 this.’ (text, BAR22102013CH_05 00:09:21.380-00:09:28.260)
- b. Perintah nekap dalam laq ngalap lun kerja
 government AV.search road ? AV.get people work
 let negara beken.
 from nation other
 ‘The government is looking for a way to attract more workers from
 abroad.’
 (text, BAR29112013CH_01 0058:26.888-00:58:32.166)
- c. Kayuq laq m-udan teh edto ih edto kinih.
 like ? INTR-rain PT day PT day now
 ‘It looks like rain today.’
 (elicitation, BAR28102013CH_01 00:13:47.445-00:13:54.580)

The auxiliaries *daraq* and *di’eyq* have the opposite function, suggesting a desire not to do something or to avoid an action:¹⁰⁹

- (142) **Negative Desiderative**
- a. Laq di’eyq dereh~dereh latiq.
 DESID NEG.DESID REDUP~suffering farm
 ‘To avoid the hard life on the farm.’
 (text, BAR22102013CH_04 00:06:12.260-00:06:14.710)
- b. Kadiq di’eyq uih mey ruyung deh.
 but NEG.DESID 1SG.1 go together 3PL.2
 ‘But I don’t want to go with them.’
 (elicitation, BAR12082014CH_03 00:00:13.780-00:00:15.810)

Kereb and *mileh* express possibility and ability. *Kereb* possibly derives from the noun *kereb* ‘time’ and suggests being allowed to do something, or having the

¹⁰⁹ In Southern Kelabit, *daraq* is sometimes *araq*. This is also found in Lundayeh (Clayre 2002).

possibility of doing something. In contrast, *mileh* implies having the knowledge or ability to do something and is derived from the noun *ileh* ‘knowledge/ability/aptitude’ via the *-em-* infix:

(143) **Possibility and Ability**

- a. Na’am iko kereb naruq dih kineh.
 NEG 2SG.1 can AV.do DEM like.that
 ‘You can’t do it like that.’ (it won’t be possible/you aren’t allowed)
 (elicitation, fieldnotes)
- b. M-ileh ketuh tiko masaq.¹¹⁰
 INTR-able very PT=2SG.1 AV.read
 ‘You can read very well.’ (elicitation, fieldnotes)

The auxiliaries *keteng*, *pengeh*, *lem puket* and *murih* all convey aspectual meanings. *Keteng* denotes durative aspect, *pengeh* denotes completive aspect (and can be used as a main verb meaning ‘finished’), *lem puket* denotes in the process of, and *murih* conveys habitual action:

(144) **Aspectual Auxiliaries**

- a. Keteng ngekal~ngekal buaq nedih.
 still REDUP~AV.pick fruit 3SG.POSS
 ‘He is still picking his fruit.’
 (elicitation, BAR30102013CH_01 00:01:55.680-00:01:58.436)
- b. Pengeh nuh naruq dih.
 finish PT=1SG.1 AV.do DEM
 ‘I’ve already done it.’
 (elicitation, BAR28102013CH_03 01:08:53.150-01:08:54.893)
- c. Nih lem puket kamih kuman.
 DEM in process 1PL.EXCL AV.eat
 ‘We are currently eating.’
 (elicitation, BAR21102013CH_01 00:54:06.200-00:54:08.511)

¹¹⁰ Elsewhere in the thesis, morpheme boundaries are not represented for *mileh* ‘INTR.able’ as the analysis is not central to the analysis. It can be understood to be further decomposable nonetheless.

- d. Murih metoq kedieh ne-muliq~muliq udung migu.
 often PT 3SG.EMPH PFV-REDUP~return end week
 ‘She often came back at weekends.’
 (text, BAR08092014CH_05 00:03:08.830-00:03:11.310)

Finally, the verb *mey* ‘to go’ can also be used as an auxiliary, denoting future tense and habitual action. It can also form a hortative construction:

(145) **Future Tense**

- a. Dooq tebeyq narih mey, mey nalan, mey ni’er...
 good PT IMPERS go go walk go AV.see
 ‘It would be good actually if we went walking, went and had a look...’¹¹¹
 (text, BAR01082014CH_04 00:00:40.700-00:00:43.750)

Habitual Action

- b. M-uliq narih udung migu dih, mey neh kamih
 INTR-return IMPERS end week DEM go PT 1PL.EXCL
 naruq latiq ruyung tetameh, ruyung tesineh kamih.
 AV.do farm with father with mother 1PL.EXCL
 ‘When you got home at the end of the week, we went to farm with our mothers and fathers.’
 (text, BAR08092014CH_03 00:02:01.210-00:02:06.720)

c. **Hortative**

- Mey titeh kuman!
 go PT=DU.INCL AV.eat
 ‘Let’s go eat!’
 (elicitation, BAR14102013CH_02 00:35:27.810-00:35:30.040)

Some auxiliaries are borrowed from Malay, including *mesti* ‘must’ to indicate necessity:

(146) **Necessity**

- a. Mesti kamih kail~kail tupeh padey lah.
 need 1PL.EXCL REDUP~strong pound rice PT
 ‘We needed to pound the rice hard.’
 (text, BAR22102013CH_04 00:02:41.620-00:02:45.820)

¹¹¹ This appears to be future tense in that the discussion is centred around what to do at the weekend.

2.4.2.7 Deictic Terms

In Kelabit, there are three levels of deixis: proximal, medial and distal. These are typically encoded through the forms *inih*, *idih* and *ineh*, which can be shortened to *nih*, *dih* and *neh*.¹¹² These forms combine with prepositions to form demonstratives, spatial expressions and temporal/manner expressions:

- (147) a. **Demonstratives**
Proximal: *sinih* ‘this one’
Medial: *sidih* ‘that one’
Distal: *sineh* ‘that one’
- b. **Plural Demonstratives**
Proximal: *nuk inih* ‘these’
Medial: *nuk idih* ‘those’
Distal: *nuk ineh* ‘those’
- c. **Spatial Terms**
Proximal: *nginih* ‘here’
Medial: *ngidih* ‘there’
Distal: *ngineh* ‘there’
- d. **Manner Terms**
Proximal: *kinih* ‘like this’
Medial: *kidih* ‘like that’
Distal: *kineh* ‘like that’
- e. **Temporal Terms**
Proximal: *inih* ‘now’
Medial: *idih* ‘then’
Distal: *ineh* ‘then’

In addition to (147), there is a two-way contrast between *tungey* ‘here’ and *nangey* ‘there’. In some cases, the forms *sinih nih*, *sineh neh* and *sidih dih* are used:

¹¹² The preposition *ngi* is sometimes also used as a demonstrative and represents a further level of distance from speaker and hearer.

(148) **Multiple Demonstratives**

- a. kumaq uih ngi sekolah sinih nih
while 1SG.1 at school DEM DEM
'while I was at this school'

(text, BAR08092014CH_05 00:09:51.800-00:09:53.910)

This can be understood as emphatic, in a similar manner to 'this one here' in English and may be another instance of reduplication (SUBSECTION 2.4.1.4).

Deictic terms have several functions. As demonstratives, they modify nouns, and occur at the end of the NP (SUBSECTION 2.4.2.1). They have also developed an aspectual function and can be used to mark progressive aspect. The proximal is used if the action is occurring in front of the speaker. The distal is used if the action is occurring at some other time or location, and the medial reflects an action that follows directly from the previous one.

(149) **Progressive Aspect**

- a. [Nih] uih mekuleng beruh.
DEM 1SG.1 AV.repeat new
'Now I'm repeating again.'

(experiment, BAR09092014CH_03 00:03:13.730-00:03:16.170)

- b. [Neh] ieh riruh.
DEM 3SG.1 laugh
'He was laughing.'

(elicitation, BAR20082014CH_02 00:01:28.000-00:01:32.000)

- c. [Dih] kamih mey ngidih dih.
DEM 1PL.EXCL go there DEM
'We're going over there.'

(elicitation, fieldnotes)

In this function, the deictic terms always occur clause-initially.

They also serve a clause linking function, suggesting consecutive action:

(150) **Clause Linking**

- a. [Dih] nieh nalan~nalan nieh edteh edto keyh.
DEM PT=3SG.1 REDUP~walk PT=3SG.1 one day PT
'So he set off one day.'

[Neh] nieh karuh~karuh ngen diweh turun.
DEM PT=3SG.1 REDUP~talk to 3DU perhaps
'Then he talked to the two of them perhaps.'

(text, BAR17082014CH_02 00:00:55.970-00:01:01.460)

Demonstatives can also be used as pronouns, typically for inanimates, but also for animate third persons:

(151) **Pronouns**

- a. Kenen ieh [dih].
UV.IRR.eat 3SG.1 DEM
'He will eat it.'

(elicitation, BAR19082014CH_03 00:07:22.185-00:07:23.395)

Finally, demonstratives can be used in locative clauses to indicate position in space:

(152) **Locative Clauses**

- a. [Dih] edteh emuq m-udur luun edteh kerusi.
DEM one girl INTR-stand on one chair
'There a girl is standing on a chair.'

(elicitation, BAR20082014CH_01 00:00:37.430-00:00:40.650)

- b. [Nih] edteh kayuh luun tanaq ih.
DEM one wood on ground PT
'Here's a stick on the ground.'

(elicitation, BAR20082014CH_01 00:08:58.560-00:09:01.830)

- c. [Neh] bukuh ih luun miji neh.
DEM book PT on table DEM
'There's the book on the table.'

(elicitation, BAR15102013CH_01 00:57:30.759-00:57:35.500)

2.4.2.8 Pronouns

In Kelabit there are two basic sets of pronouns. These are referred to as FORM 1 and FORM 2 and differ in 1SG, 2SG, 3SG and 3PL. As a rule, FORM 1 pronouns are used for subjects and FORM 2 pronouns are used for actor non-subjects. However, their distribution is somewhat more complicated than this, as discussed in CHAPTER 4.

Table 2.11 Kelabit FORM 1 Pronouns

	1.INCL	1.EXCL	2	3
SINGULAR	n/a	uih	iko	ieh
DUAL	kiteh ¹¹³	kediweh	meduweh	diweh
PAUCAL	teluh	keteluh	meteluh	deteluh
PLURAL	tauh	kamih	muyuh	ideh

Table 2.12 Kelabit FORM 2 Pronouns

	FORM 1	FORM 2
1SG	uih	kuh
2SG	iko	muh
3SG	ieh	neh
3PL	ideh	deh

The dual and paucal pronouns are formed via the prefixation of the numerals *duweh* ‘two’ and *teluh* ‘three’. In this context, *ke-* represents first person, *me-* second person and *de-* third person.¹¹⁴ The paucal pronouns often refer to groups of three, like trial pronouns, and have a morphological connection with the numeral ‘three’. However, they are also used to refer to small groups. This is considered polite and is common to the Apad Uat languages (Beatrice Clayre p.c.):

¹¹³ Both forms *kiteh* and *titeh* are attested. They do not seem to be dialect variants and speakers do not suggest any obvious semantic difference but there may be some variation in distribution. This remains to be further analysed.

¹¹⁴ *de-* is found non-productively in other word-formations that refer to others, e.g. *dulun* ‘other people’ from *lun* ‘person’, *dingi* ‘over there’ from *ngi* ‘at/there’. The forms *la’ih* and *dela’ih* ‘man’, *edtur* and *dedtur* ‘woman’ also co-occur. It may be that *deN-* is also a combination of *de-* plus a linker *nge-* (SUBSECTION 2.4.1.1.1).

- (153) **Paucal pronouns**
- a. Laq uih nubut meteluh anak~anak adiq ah.
 DESID 1SG.1 AV.encourage 2PAU REDUP~child small EXCL
 ‘I want to motivate you children (referring to group of more than three).’ (text, BAR08092014CH_03 00:00:31.280-00:00:33:860)

Hence, they are analysed as paucal rather than trial pronouns in this thesis. Himmelmann (2005a: 149) suggests that dual and trial pronouns are not common in Western Austronesian. However, they are found in other languages in Borneo (Soriente 2013).

The pronouns can be combined with the preposition *ngen* ‘to/with’ to represent obliques. The FORM 2 pronouns cliticise:

- (154) ***Ngen* + FORM 2 pronouns**
- a. *ngekuh* ‘to.1SG’
 b. *ngemuh* ‘to.2SG’
 c. *ngeneh* ‘to.3SG’
 d. *ngedeh* ‘to.3PL’

2.4.2.8.1 Possessive Pronouns

Possession is marked by placing the possessor after the possessed noun (SUBSECTION 2.4.2.1). This is another way in which Kelabit is head-initial:

- (155) **Nominal Possession**
- a. rumaq [la’ih sineh]
 house man DEM
 Possessed noun Possessor
 ‘that man’s house’ (elicitation, fieldnotes)

Neither the possessed head noun nor the possessor is marked using special morphology.

With pronouns, either FORM 1 or FORM 2 pronouns can mark possession:

(156) **Pronominal Possession**

- a. rumaq uih
house 1SG.1
'my house'
- b. rumaq kuh
house 1SG.2
'my house'

(elicitation, fieldnotes)

However, there are also two sets of dedicated possessive pronouns. The first set derives from the combination of FORM 2 pronouns and demonstratives:

(157) **Possessive Pronouns**

- a. *kuh* + *dih* → *kudih* '1SG.POSS'
b. *muh* + *dih* → *mudih* '2SG.POSS'
c. *neh* + *dih* → *nedih* '3SG.POSS'
d. *deh* + *dih* → *dedih* '3PL.POSS'

These possessive pronouns follow the head noun and are treated as single pronouns since they can be combined with other demonstratives:

(158) **Possessive Pronouns**

- a. *ngi* *liang* [[ri'er kudih] *nih*
at under neck 1SG.POSS DEM
'underneath my neck'

(text, BAR25102013CH_01 00:00:36.795-00:00:37.875)

The second series of dedicated possessive pronouns involves the prefixation of *de-* or its allomorph *d-* to the FORM 1 pronouns:

Table 2.13 *Kelabit Possessive Pronouns*

Form 1 pronoun	Possessive pronoun
uih	duih
kiteh/titeh	dekiteh/detiteh
kediweh	dekediweh
teluh	deteluh
keteluh	deketeluh
tauh	detauh
kamih	dekamih
iko	diko
meduweh	demeduweh
meteluh	demeteluh
muyuh	demuyuh
ieh	dieh
diweh	dediweh
deteluh	dedeteluh
ideh	dideh
narih	denarih

In contrast to all other possessive constructions, these precede the possessed noun:

(159) **Possessive Pronouns**

a. Peh duih bukuh?
 where 1SG.POSS book
 ‘Where’s my book?’

b. *Peh bukuh duih?
 where book 1SG.POSS
 For: ‘where’s my book?’

c. Peh bukuh kudih?
 where book 1SG.POSS
 ‘Where’s my book?’

d. *Peh kudih bukuh?
 Where 1SG.POSS book
 For: ‘where’s my book?’

(elicitation, fieldnotes)

The forms can also occur without the possessed noun. The meaning is roughly equivalent to ‘mine’, ‘yours’ and ‘his/hers’ etc.:

(160) **Possessive Pronouns**

a. Peh duih?
where 1SG.POSS
'Where's mine?'

b. Duih dih.
1SG.POSS DEM
'That's mine.'

(elicitation, fieldnotes)

It is said to be uncommon for symmetrical voice languages to have dedicated possessive pronouns (Himmelmann 2005a). It remains to be seen what determines the use of the different strategies for marking possession.

2.4.2.8.2 Impersonal Pronouns

Kelabit also has an impersonal pronoun *narih* that can be used to refer to speaker, addressee or a third party. It is typically used in irrealis contexts, such as imperatives, negatives, habituals, questions and future:

(161) **Imperatives and wishes**

a. Belajar tu'uh~tu'uh narih keyh [...] kedeh.
study REDUP~true IMPERS PT say.3PL.2
'Make sure you (the addressee) study hard, ok, they said.'

(text, BAR22102013CH_04 00:05:33.310-00:05:35.850)

b. Kuman dooq~dooq narih!
AV.eat REDUP~good IMPERS
'Bon appétit (to a single addressee or group)!'

(elicitation, BAR14102013CH_02 00:23:15.860-00:23:17.680)

c. **Habitual**

Naruq nuk kereb tu'en narih.
AV.do REL can UV.IRR.do IMPERS
'Do what I can do.'

(text, BAR22102013CH_02 00:00:22.920-00:00:25.510)

d. **Negative Habitual**

Buken narih tudo ngi rumaq...
NEG IMPERS sit at house

‘We (generic) didn’t just sit around at home...’

Napuh neh narih lem rumaq narih. [...]
sweep PT IMPERS in house IMPERS

‘We (generic) swept in our houses.’

(text, BAR22102013CH_04 00:01:39.380-00:01:44.420)

Mey meman berek neh narih.
go AV.feed pig PT IMPERS

‘We (generic) went to feed the pigs.’

(text, BAR22102013CH_04 00:01:49.890-00:01:51.690)

e. **Question**

Ngudeh narih na’am mey ngalap kayuh toq? [...]
why IMPERS NEG go AV.fetch wood first

‘Why didn’t you (addressee+contemporaries) go and get wood first?’

(text, PDA06112013CH_10 00:01:13.073-00:01:14.853)

f. **Future (request)**

Pan-en narih kayuh ih rengaq narih muliq na’an.
carry-UV.IRR IMPERS wood PT when IMPERS return later

‘Take the wood on your shoulders when you go back later.’

(elicitation, fieldnotes)

g. **Conditional**

Getebpen tukang narih, rengaq narih na’am liang kelaboq ih.
UV.IRR.bite mosquito IMPERS if IMPERS NEG under net PT

‘You (generic) will get bitten by mosquitoes if you don’t sleep under a net.’

(elicitation, BAR18082014CH_01 00:15:24.330-00:15:28.410)

They are common in procedural texts and in personal histories, detailing habitual as opposed to specific events. As seen in (161), *narih* can refer to a generic referent, a specific addressee, a group that includes the addressee, the speaker, or a group that includes the speaker.

In addition to *narih*, which is used for generic/unspecified animate referents, there is an inanimate pronoun *enaq*. This can replace any noun or noun phrase and is often used in contexts where the speaker is searching for the right word to say:

(162) **Function of *enaq***

- a. Mala enaq uih ruka sinih,
 AV.say PRO 1SG.1 time DEM
 kapeh lun tauh tupeh padey.
 how people 1PL.INCL pound rice
 ‘What I’m going to talk about this time is how we pound rice.’
 (text, BAR27102013CH_02 00:00:02.610-00:00:11.970)

2.4.2.8.3 Emphatic Pronouns

Finally, there is a set of emphatic pronouns, formed by prefixing *ke-* or *ked-* to FORM 1:¹¹⁵

Table 2.14 *Kelabit Emphatic Pronouns*

	FORM 1	Emphatic
1SG	uih	keduih
1DU.INCL	kiteh	kekiteh/ketiteh
1DU.EXCL	kediw eh	kekediw eh/kedikediw eh
1PAU.INCL	teluh	keteluh
1PAU.EXCL	keteluh	keketeluh
1PL.INCL	tau h	ketau h
1PL.EXCL	kami h	kekami h
2SG	iko	kiko
2DU	meduw eh	kemeduw eh
2PAU	meteluh	kemeteluh
2PL	muyuh	kemuyuh
3SG	ieh	kedieh
3DU	diw eh	kedediw eh/kedidiw eh
3PAU	deteluh	kedeteluh
3PL	ideh	kedideh
IMPERS	narih	kenarih/kedinarih

¹¹⁵ This includes the allomorphs *k-* and *ked-* which attach to vowel initial roots.

These can be used in place of the basic pronouns, in which case they give an emphatic reading:

(163) **Function of Emphatic Pronouns**

- a. Keteng ta'ut teh keduih ngen sineh.
 still scared PT 1SG.EMPH with DEM
 'I'm still scared of that (worms).'
 (text, BAR02082014CH_02 00:04:57.715-00:04:59.855)

Emphatic pronouns can also show contrastive points of view:

(164) **Function of Emphatic Pronouns**

- a. Kadiq nieh ideh naruq lubang sineh ken keduih.
 because PT=3SG.1 3PL.1 AV.do hole DEM say 1SG.EMPH

 Am keliq kapeh ken lun beken.
 NEG know how say people other
 'That's why they make the hole, I say. I've no idea what other people think.'
 (text, BAR27102013CH_02 00:04:11.941-00:04:17.305)

Additionally, they can be used as NP modifiers in juxtaposition with full nouns, proper names and pronouns:

(165) **Function of Emphatic Pronouns**

- a. Ngi [teh la'ih rayeh suk na'ah ih] kedieh
 DEM PT man big REL earlier PT 3SG.EMPH

 ngi udung buaq nedih.
 at top fruit.tree 3SG.POSS
 'Here's the old man from before up in the tree that he's climbing.'
 (elicitation, BAR30102013CH_01 00:01:52.536-00:01:55.636)
- b. lun ruyung [Sinaq] kedieh
 people together mother 3SG.EMPH
 'mother herself's family'
 (text, BAR22102013CH_04 00:09:22.030-00:09:23.440)

- c. Na'am [Margaret] kedieh laq ngarang.
 NEG Margaret 3SG.EMPH DESID dance
 'Margaret doesn't like to dance.'
 (elicitation, BAR21102013CH_01 00:55:55.140-00:55:58.086)
- d. Nih [sidih] kedieh madaq buaq nuk birey nedih dih.
 now DEM 3SG.EMPH AV.show fruit REL UV.PFV.give 3SG.POSS DEM
 'Now that one is showing the fruit he was given.'
 (elicitation, BAR30102013CH_01 00:05:00.450-00:05:03.230)

In (165), the emphatic pronouns function similarly to reflexive pronouns, such as 'me myself' in English.

2.4.2.8.4 Inklusory Pronouns

Finally, in a similar manner to many Austronesian languages, Kelabit non-singular pronouns can be used to express a construction meaning 'X & Y' (see SUBSECTION 2.5.3.5 for co-ordination of NPs and VPs). As described by Lichtenberk (2000), the pronoun gives the total set of participants and the following noun delimits the possible referents:

- (166) **Inklusory Pronouns**
- a. kediwéh John
 1DU.EXCL John
 'John and I'
- b. meduwéh John
 2DU John
 'you and John'
- c. Peter diwéh John
 Peter 3DU John
 'Peter and John' (elicitation, fieldnotes)

The same is possible with words formed using *deN-* (SUBSECTION 2.4.1.1.1):

- (167) **Groups of relations**
- a. kamih denge-ruyung
1PL.EXCL kin-together
'my family and I'
 - b. muyuh denge-ruyung
2PL kin-together
'you and your family'
 - c. John ideh deng-anak
John 3PL kin-child
'John and his brother' (elicitation, fieldnotes)
 - d. Batuh Lawih diweh denge-rumaq
Proper Name 2DU kin-house
'Batuh Lawih and his wife'
(text, BAR17082014CH_02 00:00:02.750-00:00:04.500)

2.4.2.9 Interrogatives

The main interrogatives in Kelabit are as follows:

- (168) **Interrogatives**
- a. *iih* 'who'
 - b. *enun* 'what'
 - c. *ngapeh* 'where'
 - d. *idan* 'when'
 - e. *kapeh* 'how'
 - f. *ngudeh* 'why'
 - g. *suk apah* 'which one'
 - h. *tudaq* 'how many'

Though they share interrogative semantics, the interrogatives in (168) may not constitute a single class. The forms *iih* 'who' and *enun* 'what' have a similar distribution to nouns. In contrast, the remaining question words are more similar to adverbs or verbs. Indeed, *ngudeh* 'why' has verbal properties. Firstly, it is formed via *N-* prefixation from the root *kudeh*. Secondly, it can be used to mean 'say/do something' like Malay *mengapa* 'why' and similar forms in other Austronesian languages (Maria Polinsky, p.c.):

This always occurs initially and cannot be followed by sentence particles (SUBSECTION 2.4.2.14.1):

- (172) **Position of Question Particle**
- a. Ken elaq tebeyq Peter kuman buaq kaber?
 Q DESID actually Peter AV.eat fruit pineapple
 ‘Would Peter like to eat pineapple’ (elicitation, fieldnotes)
- b. *Ken teh iko tudo?
 Q PT 2SG.1 sit
 For: ‘Are you sitting?’
 (elicitation, BAR18082014CH_01 00:47:17.680-00:47:19.770)

The clause in (172b) would be grammatical either without the particle *teh*, or if the particle was preceded by either the subject or the verb: *ken iko teh tudo* ‘are you the one who is sitting?’ or *ken tudo teh iko* ‘are you sitting?’

2.4.2.10 Relativisers

In Kelabit, there are two relativisers that occur at the beginning of a relative clause:

- (173) **Relativisers**
- a. *suk* singular, specific referents
 b. *nuk* plural and singular non-specific referents

The contrast between *suk* and *nuk* is illustrated in (174):

- (174) **Relativisers**
- a. dela’ih [suk ma’it aleb]
 man REL AV.hurt knee
 ‘the man who hurt his knee’
 (elicitation, BAR31072014CH_05 00:00:44.220-00:00:46.580)
- b. #Uih ne-ni’er edteh dela’ih [suk nalan].
 1SG.1 PFV-AV.see one man REL walk
 ‘I saw a man who was walking/a walker.’
 (elicitation, BAR31072014CH_05 00:00:56.640-00:01:00.060)

- c. *Uih ne-ni'er mulaq dela'ih [suk nalan].
 1SG.1 PFV-AV.see many man REL walk
 'I saw many men who were walking.' (elicitation, fieldnotes)
- d. Uih ne-ni'er mulaq dela'ih [nuk nalan mey Pa Remapoh].
 1SG.1 PFV-AV.see many man REL walk to Pa Remapoh
 'I saw many men who were walking to Pa Remapoh.'
 (elicitation, fieldnotes)

It is ungrammatical to use *suk* with plural referents, as in (174c). It is also semantically odd to use *suk* with indefinite and non-specific referents, such as *edteh dela'ih* 'a man' in (174b), which can be understood as discourse new since it is modified by the indefinite numeral *edteh* 'one'.¹¹⁶ Hence, the distribution is not *suk* for singular referents and *nuk* for plural referents, but rather *suk* is reserved for singular and specific referents and *nuk* used everywhere else.

Both *nuk* and *suk* also function as nominalisers, forming headless relative clauses. These can be combined with UV irrealis forms to create generic nouns:

- (175) **Generic Nouns**
- a. nuk ken-en
 REL eat-UV.IRR
 'food'
- b. nuk belaan
 REL UV.IRR.say
 'speech/song'

These function as nouns, as they can be modified by demonstratives, adjectives and relative clauses:

¹¹⁶ The referent need not necessarily be definite, since it is compatible with the indefinite numeral *edteh* 'one' so long as the referent is specific and 'anchored in discourse' (Lambrecht 1994, CHAPTER 5):

- (i) Uih ne-ni'er edteh dela'ih suk nalan mey Pa Remapoh.
 1SG.1 PFV-AV.see one man REL walk to Pa Remapoh
 'I saw a (specific) man who walked to Pa Remapoh.' (elicitation, fieldnotes)

- (176) **Generic Nouns**
 nuk kenen [nuk dooq~dooq ih]
 food REL REDUP~good PT
 ‘very tasty food’
 (text, BAR08092014CH_03 00:09:59.155-00:10:00.425)

2.4.2.11 Conjunctions

Conjunctions serve the function of introducing clausal adjuncts.¹¹⁷ They include:

- (177) **Conjunctions**
- a. *rengaq* ‘if/when’
 - b. *utak* ‘if’
 - c. *tulu* ‘if’
 - d. *kadiq* ‘because/that’s why’
 - e. *lem kumaq* ‘whilst’
 - f. *pingan* ‘after’
 - g. *asal* ‘as long as’ (from Malay)
 - h. *aban* ‘because of/only if’
 - i. *pengeh* ‘after’
 - j. *pu’un* ‘before’ (also *sebelum* from Malay)
 - k. *atau pun* ‘or’ (from Malay)

They can also be identified by their position at the beginning of a subordinate clause.

More information on adjunct clauses and co-ordination is given in SUBSECTION 2.5.3.4 and 2.3.5.3.

2.4.2.12 Numerals

Cardinal numerals in Kelabit are shown in TABLE 2.15:

¹¹⁷ Some can take both clausal and nominal complements. For example, *pingan* ‘before’ and *pengeh* ‘after’ also occur with demonstrative arguments, e.g. *pingan inih* ‘next time’ or *pengeh ineh* ‘after that’.

Table 2.15 Cardinal Numbers

	Kelabit
1	edteh
2	duweh
3	teluh
4	epat
5	limeh
6	enem
7	tuduq
8	waluh
9	iwak
10	puluq
11	puluq edteh
12	puluq duweh
13	puluq teluh
14	puluq epat
15	puluq limeh
16	puluq enem
17	puluq tuduq
18	puluq waluh
19	puluq iwak
20	duweh ngepuluq
21	duweh ngepuluq edteh
22	duweh ngepuluq duweh
23	duweh ngepuluq teluh
24	duweh ngepuluq epat
25	duweh ngepuluq limeh
30	teluh ngepuluq
40	epat ngepuluq
50	dimeh ngepuluq
100	ratu
200	duweh ngeratu
1000	ribuh
2000	duweh ngeribuh

The form *nge* that appears in numerals after twenty could be considered a ligature or linker, as is often found in Philippine languages (e.g. *nga* in Iloko, Rubino 2005).

Morpheme boundaries remain to be further analysed. Ordinal numbers are formed via *ke*₂- prefixation (SUBSECTION 2.4.1.1.3), distributive numerals with *te*₂- (SUBSECTION 2.4.1.1.14) and multiplicative numerals via *min*- (or *mi*-) prefixation:

- (178) **Multiplicative Numerals**
- a. *edteh* → *midteh* ‘once’
 - b. *duweh* → *minduweh* ‘twice’
 - c. *teluh* → *minteluh* ‘three times’

Apart from their meaning and morphological derivation, numerals can also be identified via their position. Like quantifiers, numerals typically precede the noun that they modify (see SUBSECTION 2.4.2.1 for discussion).

2.4.2.13 Quantifiers and Negators

Kelabit also has a closed class of quantifiers. These include:

- (179) **Quantifiers**
- a. *mulaq* ‘many’
 - b. *si’it* ‘a bit’
 - c. *tudaq* ‘a few/several’
 - d. *ibal* ‘some’
 - e. *ngabi/abi-abi* ‘all’
 - f. *kenep-kenep* ‘every’
 - g. *sukup* ‘enough’ (perhaps borrowed from Malay *cukup*)

Like numerals, quantifiers can be determined by their function of quantifying nouns and their position pre-nominally.¹¹⁸ Examples of quantificational structures are given in SUBSECTION 2.5.2.1.

The main negator in Kelabit is *na’am* and typically appears clause-initially, as shown in example (110). There is also a variant form *buken*, illustrated in example (161d). It is not clear what the difference between *na’am* and *buken* is, or whether *buken* was borrowed from Malay. Strategies for negation are discussed in SUBSECTION 2.5.2.1.

¹¹⁸ Asmah (1983) treats numerals and quantifiers as sub-types of noun.

2.4.2.14 Particles

The final closed class in Kelabit is particles. These are distinguished from auxiliaries in that they modify the whole clause rather than the verb, and have discourse rather than aspectual or modal semantics. They can be subdivided into two types on the basis of their distribution. One class of particles appears exclusively in the clause-final position and typically signals the attitude of the speaker towards the utterance or elicits a response from the addressee. These resemble similar particles in Indonesian (Ewing 2005). The other class of particles has a variable distribution. They often occur in the so-called ‘second-position’, following an initial word or phrase, like equivalent particles in Philippine-type languages (Himmelmann 2005a: 113, see CHAPTER 4). However, as discussed in SUBSECTION 2.4.2.14.1, these particles do not have the typical patterns of ‘second-position’ particles/clitics as they also occur clause-finally and sometimes appear in both the second position and final position of a given clause. Consequently, these particles are given the more neutral label of ‘sentence particles’. Particles never appear clause-initially, unlike auxiliaries (SUBSECTION 2.4.2.6) and deictic terms (SUBSECTION 2.4.2.7).

2.4.2.14.1 Sentence Particles

The most frequently used sentence particles in Kelabit are listed in (180) with preliminary glosses based on their usage in the corpus. The exact semantics of each particle remains to be further specified during future research. However, they appear to serve discourse functions, including clause chaining, politeness and emphasis.

(180) **Clause-Chaining**

- a. *betuq* ‘yet/first’
- b. *netuq* ‘anymore’
- c. *metuq* ‘and’
- d. *men* ‘and/emphasis’
- e. *meteh* ‘and’

Politeness

- f. *tebeyq* ‘actually/politeness’
- j. *nebeyq* ‘actually/politeness’

Emphasis

- g. *tun* ‘indeed’
- h. *eden* ‘only’
- i. *ayuq* ‘nature/emphasis’
- j. *burur* ‘body/emphasis’

Information Structure

- k. *neh* ‘discourse topic/focus’
- l. *teh* ‘identificational focus’
- m. *peh* ‘too/even/additive focus’

The particles *teh* and *neh* are particularly frequent and often precede the argument privileged by the verbal morphology (see SUBSECTION 2.5.1.1). However, unlike pre-nominal particles in other Western Austronesian languages, they are optional and appear to have an information structure function, such as marking a discourse topic or indicating the focus status of material to the right.¹¹⁹ This is supported by the fact that the pre-particular constituent can be modified with focus particles like *sebuleng* ‘themselves/alone’ and *tupu* ‘only/just’:

¹¹⁹ Like *betuq*, *netuq* and *metuq* etc. they can specify a relationship between the current clause and a preceding or following clause:

- (i) Tak iko teh uwan sineh neh, uih neh uwan sinih nih.
if 2SG.1 PT have DEM DEM 1SG.1 PT have DEM DEM
‘If you have that one, then I have this one.’
(elicitation, fieldnotes)

(181) **Information Structure and Particles**

- a. [Kediweh Poline tupu] neh inan visa.
2DU.EXCL Poline only PT have visa
'Only Poline and I had visas.'

(text, BAR21082014CH_07 00:06:40.220-00:06:42.950)

Consequently, these particles have more in common with focus particles, such as *do* in Toba Batak (cf. Silitonga 1973), than pre-nominal or case-marking particles in Tagalog (see SUBSECTION 1.3.1).

Like 'second-position' particles, the particles in (180) often occur following an important word or phrase clause-initially. The pre-particular constituent is not restricted to a particular word class and includes adverbs, auxiliaries, demonstratives, quantifiers, verbs or question words:

(182) a. **Adverb**

- [Edto riak] teh Peter umak alud nedih.
day future PT Peter board boat 3SG.POSS
'Peter will board his boat tomorrow.'

(elicitation, BAR21102013CH_01 00:11:40.736-00:11:43.696)

b. **Auxiliary**

- [Kereb] teh Peter ne-kuman buaq kaber nedih.
can PT Peter PFV-AV.eat fruit pineapple 3SG.POSS
'Peter might have eaten his pineapple.'

(elicitation, BAR21102013CH_01 00:22:22.486-00:22:27.648)

c. **Demonstrative**

- [Neh] teh Peter kuman buaq kaber ih.
DEM PT Peter AV.eat fruit pineapple PT
'Then Peter ate pineapple.'

(elicitation, BAR19082014CH_03 00:41:49.570-00:41:52.140)

d. **Quantifier**

- [Na'am] teh Peter kuman buaq kaber.
NEG PT Peter AV.eat fruit pineapple
'Peter doesn't eat pineapple.'/'It is not the case that Peter eats pineapple.'

(elicitation, BAR19082014CH_03 00:44:21.230-00:44:24.180)

- e. **Verb**
 [Kuman] teh Peter buaq kaber ih na'an.
 AV.eat PT Peter fruit pineapple PT later
 'Peter will eat his pineapple later.'
 (elicitation, BAR19082014CH_03 00:38:42.492-00:38:46.605)

- f. **Question Word**
 [Ngapeh] teh Peter kuman buaq kaber ih?
 where PT Peter AV.eat fruit pineapple PT
 'Where did Peter eat the pineapple?'
 (elicitation, BAR19082014CH_03 00:31:50.820-00:31:55.790)

It can also be a larger constituent, such as the verb and its non-subject core argument, or a constituent containing the VP, auxiliaries and adjuncts:

- g. **Verb + Non-subject Core Argument**
 [Kuman enun] teh Peter na'an neh?
 AV.eat what PT Peter later DEM
 'What will Peter eat later?'
 (elicitation, BAR19082014CH_03 00:13:05.600-00:13:08.310)

- h. **Auxiliary + VP**
 [Laq kuman buaq kaber ngapeh] tebeyq Peter?
 DESID AV.eat fruit pineapple where PT Peter
 'Where does Peter want to eat pineapple?'
 (elicitation, BAR19082014CH_03 01:06:04.277-01:06:07.122)

Finally, it is also possible to find the subject NP in initial-position, followed by the particle and then the verb.¹²⁰ In such constructions, the subject NP has identificational focus, in the sense of É-Kiss (1998):

- (183) **Subject NP**
 a. [Diwéh sebuleng] teh mala sineh.
 3DU alone PT AV.say DEM
 'It was just the two of them that sang that one.'
 (text, PUM18102013CH_02 00:00:11.170-00:00:12.770)

¹²⁰ It is not clear if this applies only to *neh* and *teh* or to all of the particles in (180).

The only words that cannot appear alone before the particles are the question particle *ken* (SUBSECTION 2.4.2.9) and the non-subject core argument (SUBSECTION 2.5.1.2).

However, as discussed in SUBSECTION 2.4.2.14, the particles do not have the typical patterns of ‘second-position’ particles as they can also occur clause-finally:

- (184) **Particles Clause-finally**
- a. Kapeh~kapeh peh, dooq teh.
 REDUP~how PT good PT
 ‘However it is, it will be good.’ (elicitation, fieldnotes)
- b. Telenen kuh tabat kudih betoq.
 swallow.UV.IRR 1SG.2 medicine 1SG.POSS PT
 ‘I’ll swallow my medicine first.’
 (elicitation, BAR28102013CH_03 01:27:53.662-01:27:56.162)

Moreover, in some sentences there are sentence particles in both the clause-final and the second position:

- (185) **Particles**
- [Neh] teh [Peter kuman buaq kaber] tebeyq.
 DEM PT Peter AV.eat fruit pineapple PT
 ‘Then Peter would eat pineapple.’
 (elicitation, BAR19082014CH_03 00:42:02.430-00:42:05.310)

Hence, they differ from true second-position phenomena, as discussed in CHAPTER 4.

However, they never occur clause-initially, which suggests they may be clitics and require a prosodic host, as discussed in SUBSECTION 4.6.1:

- (186) **Particles**
- a. *Men Peter kuman buaq kabere.
 PT Peter AV.eat fruit pineapple
 For: ‘But Peter eats pineapple.’
- b. *Tebeyq Peter kuman buaq kabere.
 PT Peter AV.eat fruit pineapple
 For: ‘Peter would eat pineapple.’
 (elicitation, BAR19082014CH_03 00:51:29.247-00:51:34.049)

2.4.2.14.2 Clause-final Particles

The second class of particles are exclamatory particles, including the following:

- (187) **Clause-final particles**
- a. *keyh* ‘excl’ (elicits agreement for hearer)
 - b. *bah* ‘excl’ (from Malay/marks uncertainty)
 - c. *lah* ‘excl’ (from Malay/asserts truth of utterance)
 - d. *koq* ‘excl’ (emphasis)
 - e. *kah* ‘excl’ (marks surprise/uncertainty)

Again, the exact semantics and pragmatics of each of the particles remains to be further studied. However, they are typically used to invite a response from the addressee or comment on the speaker’s attitude towards the utterance. Unlike the sentence particles in SUBSECTION 2.4.2.14.1, they only occur clause-finally, as in example (217).

In addition to the particles in (187), there are gendered particles that occur with or without the prefix *ke-* (perhaps derived from the form *ken* ‘to say/according to’):¹²¹

- (188) **Gendered Particles**
- | | | | | |
|--------------|---|----------------|---------------|----------------------|
| a. masculine | → | <i>leyh</i> | <i>keleyh</i> | |
| b. feminine | → | <i>(e)dtiq</i> | <i>kedtiq</i> | Bario Kelabit |
| | | <i>dtuh</i> | <i>kedtuh</i> | Long Lellang Kelabit |
| | | <i>suh</i> | <i>kesuh</i> | Pa Dalih Kelabit |

These are used when the speaker wishes to express a particular attitude towards the utterance, though this remains to be studied in more detail. They also fulfil a similar function to clauses ending with *ken* + a pronoun (or the clitic forms *kekuh*, *kemuh*, *keneh*, *kedeh*) in indicating indirectness. They may function as a marker of evidentiality or politeness and are common in the languages of Borneo (see Soriente 2014).

¹²¹ It is said that among the younger generations, and particularly, children both girls and boys used *leyh* though this is anecdotal. I am told that this is also true of Lun Bawang (Lucy Bulan, p.c.)

Thus, the exact function of the particles remains to be seen but they can be distinguished from other word classes on the basis of their position, discourse-based function and uninflected form.

2.4.3 Summary

In this section, I provided an overview of word formation and word classes in Kelabit. Kelabit is a head-initial language and mostly agglutinating, though some affixes appear to be fusional, particularly the voice markers. The main word-formation processes include prefixation, infixation, suffixation and reduplication and are used for both derivational and inflectional purposes. Kelabit has several lexical word classes, including nouns, verbs, adjectives and adverbs that can be distinguished according to morphology, distribution and function. There are also closed, functional classes, namely prepositions, auxiliaries, deictic terms, pronouns, interrogatives, relativisers, conjunctions, numerals, quantifiers and particles.

2.5 Syntax

In this section, I present an analysis of Kelabit syntax, including grammatical functions (SUBSECTION 2.5.1), periphrastic voices (SUBSECTION 2.5.2) and multi-clausal constructions (SUBSECTION 2.5.3). Throughout the section, I address two questions that arise from the discussion in CHAPTER 1, namely whether Kelabit has identifiable grammatical functions like ‘subject’ and to what extent Kelabit voice alternations can be described as ‘symmetrical’? In doing so, I discuss the implications of Kelabit for the ‘subject debate’ (SUBSECTION 1.4.1) and lay the foundations for an in-depth analysis of Kelabit voice in CHAPTER 3.

2.5.1 Grammatical Functions

In SUBSECTION 1.4.1, I introduced the idea that ‘subject’ has been a controversial notion in Western Austronesian languages, since traditional subject properties are split between different arguments. In SUBSECTION 2.5.1.1, I demonstrate that split subject properties are also found in Kelabit. Some properties relate to the actor, regardless of voice construction, and others relate to the argument privileged by the verbal morphology. Hence, much like Tagalog and Indonesian in CHAPTER 1, it is debatable whether Kelabit voice alternations involve an alternation in the mapping of arguments to subject or not. As illustrated in SUBSECTION 2.4.1, Kelabit has three morphological voice constructions: an AV construction, marked with the nasal prefixes *N-* and *neN-*; a UV construction, marked with the *-in-* infix or *-en* suffix; and an IV construction marked with *peN-* and *peneN-*:

(189) *Kelabit Voice*

a. **Actor Voice**

[La’ih sineh] ne-nekul nubaq nedih ngen seduk.
man DEM PFV-AV.spoon.up rice 3SG.POSS with spoon
‘That man spooned up his rice with a spoon.’

b. **Undergoer Voice**

Sikul lai’h sineh [nubaq nedih] ngen seduk.
UV.PFV.spoon.up man DEM rice 3SG.POSS with spoon
‘That man ate his rice with a spoon.’

c. **Instrumental Voice**

[Seduk] pe-nekul la’ih sineh nubaq nedih.
spoon IV-spoon up man DEM rice 3SG.POSS
‘That man used a spoon to spoon up his rice.’

(elicitation, fieldnotes)

In the following sections, I establish the functions of the different arguments in the voice constructions in (189) in order to identify whether grammatical functions are important in Kelabit and what this can tell us about the nature of the voice system.

2.5.1.1 Subject

As discussed in SUBSECTION 2.5.1, typical subject properties are split in Kelabit. Interestingly, much like Tagalog and Indonesian, patterns of reflexivisation are central to the debate (SUBSECTION 1.4.1). Reflexivity can be expressed using the *se-* and *pere-* verbal prefixes, described in SUBSECTION 2.4.1. However, reflexive constructions can also be formed using the term *burur* ‘body’. The two constructions are illustrated in (190):

(190) *Kelabit Reflexives*

a. **Morphological Reflexive**

Uih ne-*peri-badaq* ngen polis.
1SG.1 PFV-REFL-show to police
‘I surrendered myself to the police.’

b. **Body Reflexive**

Uih ne-*madaq burur kudih* ngen polis.
1SG.1 PFV-AV.show body 1SG to police
‘I surrendered myself to the police.’ (elicitation, fieldnotes)

For body reflexives, the actor always binds the reflexive, regardless of voice construction:

(191) *Actor Binds Reflexive*

a. **Actor Voice**

Uih ne-*madaq burur kudih* ngedeh.
1SG.1 PFV-AV.show body 1SG.POSS to.3PL.2
‘I surrendered myself to them.’

b. **Undergoer Voice**

Binadaq kuh burur kudih ngedeh.¹²²
UV.PFV.show 1SG.2 body 1SG.POSS to.3PL.2
‘I surrendered myself to them.’ (elicitation, fieldnotes)

¹²² The choice of actor pronoun typically varies according to the voice construction, as discussed in SUBSECTION 2.4.2.8 and further examined in CHAPTER 4. I take this to reflect a difference in grammatical function – i.e. subject vs. non-subject core argument status. The actor binds the reflexive regardless of whether the pronoun is FORM 1 or FORM 2.

(192) *Undergoer Binds Reflexive*

a. **Actor Voice**

*Burur kudih ne-madaq uih ngedeh.
body 1SG.POSS PFV-AV.show 1SG.1 to.3PL.2
For: ‘I surrendered myself to them.’

b. **Undergoer Voice**

*Binadaq burur kudih uih ngedeh.
UV.PFV.show body 1SG.POSS 1SG.1 to.3PL.2
For: ‘I surrendered myself to them.’ (elicitation, fieldnotes)

In (191), the actor binds reflexives in both AV and UV clauses. The examples in (192) demonstrate that the undergoer never binds reflexives – even in UV. This might suggest that the actor is the subject in Kelabit.¹²³

However, patterns of relativisation point to a different conclusion. In the same manner as Tagalog and Indonesian, the only argument that can be relativised from a Kelabit clause is the argument signalled in the verbal morphology:

(193) *Kelabit Relative Clauses*

a. **Actor Voice**

Seni’er kuh **la’ih** [suk ne-nekul nubaq ngen seduk].
UV.PFV.see 1SG.2 man REL PFV-AV.spoon rice with spoon
‘I saw the man who spooned up rice with a spoon.’

b. *Seni’er kuh **seduk** [suk nekul la’ih nubaq nedih].
UV.PFV.see 1SG.2 spoon REL AV.spoon man rice 3SG.POSS
For: ‘I saw the spoon that the man used to spoon up his rice.’

c. *Seni’er kuh **nubaq** [suk nekul la’ih sineh].
UV.PFV.see 1SG.2 rice REL AV.spoon man DEM
For: ‘I saw the rice that the man spooned up.’

(elicitation, fieldnotes)

¹²³ This could be further tested by looking for examples of constructions involving more than two arguments, such as ‘X introduced Y to self’. If it is truly the actor that binds the reflexive irrespective of grammatical function we would predict that X binds the reflexive and not Y. This data is not available in the current corpus.

(194) **Undergoer Voice**

- a. Seni'er kuh **nubaq** [suk sikul la'ih sineh ngen seduk].
UV.PFV.see 1SG.2 rice REL UV.PFV.spoon man DEM with spoon
'I saw the rice that the man spooned up with a spoon'
- b. *Seni'er kuh **la'ih** [suk sikul nubaq].
UV.PFV.see 1SG.2 man REL UV.PFV.spoon rice
For: 'I saw the man who spooned up rice.' (elicitation, fieldnotes)

(195) **Instrumental Voice**

- a. Seni'er kuh **seduk** [suk pe-nekul la'ih sineh nubaq nedih].
UV.PFV.see 1SG.2 spoon REL IV-spoon man DEM rice 3SG.POSS
'I saw the spoon that the man used to spoon up his rice.'
(elicitation, fieldnotes)

In an AV clause, it is only possible to relativise on the actor, and not the undergoer or instrument, as shown in (193). Similarly, in a UV clause, it is only possible to relativise on the undergoer, as shown in (194). Finally, in an IV clause, only the instrument can be relativised, as shown in (195). This would suggest that the argument privileged in the verbal morphology (henceforth ASV) is subject, following Keenan & Comrie's (1979) accessibility hierarchy.

Consequently, we are left with the same puzzle as outlined in SUBSECTION 1.4.1 for Philippine-type and Indonesian-type languages. Is the solution to abandon the notion of subject in Kelabit or to redefine subject in light of the split properties identified? I would argue that Kelabit provides additional support for Manning's (1996) inverse approach, which redefines subject according to 'reference-related' properties, since the same split can be identified in a wide range of languages. Moreover, if we assume that reflexivisation and other 'role-related' properties can be handled at argument structure, then most other subject tests identify the ASV as subject. This includes the fact that only the ASV can be questioned in initial-position or clefted (SUBSECTION 2.5.3.2), the fact that a controlled argument must be the ASV of the lower

clause (SUBSECTION 2.5.3.3) and the fact that a shared argument can be omitted in co-ordination only when it is the ASV of both clauses (SUBSECTION 2.5.3.5).

Moreover, coding and distributional properties provide additional support for the ASV as subject analysis in Kelabit. Although there is no overt case-marking of nominals in Kelabit, there are properties that suggest a privileged syntactic status for the ASV. Firstly, when the particles *teh* and *neh* precede a nominal argument, this argument is always the ASV:

(196) **Actor Voice**

a. Kuman teh [Peter] buaq kaber nedih keneh.
 AV.eat PT Peter pineapple 3SG.POSS he.said
 ‘Peter does eat his pineapple he said.’

b. *Kuman Peter teh [buaq kaber nedih] keneh.
 AV.eat Peter PT pineapple 3SG.POSS he.said
 For: ‘Peter does eat his pineapple he said.’

c. *Kuman teh [Peter] teh [buaq kaber nedih] keneh.
 AV.eat PT Peter PT pineapple 3SG.POSS he.said
 For: ‘Peter does eat his pineapple he said.’

(elicitation, fieldnotes)

(197) **Undergoer Voice**

a. Kenen Peter teh [buaq kaber nedih] keneh.
 UV.eat Peter PT pineapple 3SG.POSS he.said
 ‘Peter will eat his pineapple he said.’

b. *Kenen teh [Peter] buaq kaber nedih keneh.
 UV.eat PT Peter pineapple 3SG.POSS he.said
 For: ‘Peter will eat his pineapple he said.’

c. *Kenen teh [Peter] teh [buaq kaber nedih] keneh.
 UV.eat PT Peter PT pineapple 3SG.POSS he.said
 For: ‘Peter will eat his pineapple he said.’

(elicitation, fieldnotes)

- (198) **Instrumental Voice**
- a. Penekul Peter nubaq nedih teh [tekul inih] na'ah.
 IV.spoon Peter rice 3SG.POSS PT spoon DEM before
 'Peter used this spoon to spoon up his rice.'
- b. *Penekul teh [Peter] nubaq nedih tekul inih na'ah.
 IV.spoon PT Peter rice 3SG.POSS spoon DEM before
 For: 'Peter used this spoon to spoon up his rice.'
- c. *Penekul Peter teh [nubaq nedih] tekul inih na'ah.
 IV.spoon Peter PT rice 3SG.POSS spoon DEM before
 For: 'Peter used this spoon to spoon up his rice.'
- (elicitation, fieldnotes)

These are not case markers, like in Philippine-type languages (SUBSECTION 2.4.2.14.1). Nonetheless, the following generalisation can be made: *teh* and *neh* only precede the ASV and not any other argument. Thus, there is a correspondence between *ang*-marking in Tagalog and the Kelabit particles, since both support the ASV = subject analysis.

Finally, the ASV also has more freedom of word order than other core arguments (see SUBSECTION 5.5). The ASV is the only core argument that can appear before the verb:

- (199) *Kelabit Word Order*
- a. **Actor Voice**
 [Uih] ne-kuman buaq kaber ngimalem.
 1SG.1 PFV-AV.eat fruit pineapple yesterday
 'I ate pineapple yesterday.'
 (elicitation, BAR18082014CH_02 00:17:12.730-00:17:15.520)
- b. *Buaq kaber ne-kuman [uih].
 fruit pineapple PFV-AV.eat 1SG.1
 For: 'I ate pineapple.'
 (elicitation, BAR18082014CH_02 00:08:01.770-00:08:05.590)

- c. **Undergoer Voice**
 [Buaq kabər] kinan kuh.
 fruit pineapple UV.PFV.eat 1SG.2
 ‘I ate pineapple.’
 (elicitation, BAR18082014CH_02 00:07:09.560-00:07:11.080)
- d. *Uih kinan [buaq kabər].
 1SG.1 UV.PFV.eat fruit pineapple
 For: ‘I ate pineapple.’
 (elicitation, BAR18082014CH_02 00:04:46.970-00:04:51.680)

Similarly, only the ASV can appear between the verb and an element in initial-position, such as a negative or pre-verbal auxiliary (see SUBSECTION 4.12.1):

- (200) **Actor Voice**
 a. Na’am [Peter] ne-kuman buaq kabər nedih.
 NEG Peter PFV-AV.eat fruit pineapple 3SG.POSS
 ‘Peter didn’t eat his pineapple.’
 (elicitation, BAR 21102013CH_01 00:21:17.337-00:21:22.391)

- Undergoer Voice**
 b. Na’am [buaq kabər] kinan Peter.
 NEG fruit pineapple UV.PFV.eat Peter
 ‘Peter didn’t eat pineapple.’
 (elicitation, BAR 21102013CH_01 00:21:30.175-00:21:33.202)

- c. Na’am teh [telu’aq] nalap diweh.
 NEG PT crow UV.PFV.catch 3DU
 ‘The two of them didn’t catch a single crow.’
 (text, BAR27102013CH_03 00:01:46.770-00:01:50.240)

- Instrumental Voice**
 d. Pengeh neh [tekul ih] penekul kuh buaq kabər.
 finish PT spoon PT IV.spoon 1SG.2 fruit pineapple
 ‘I already used the spoon to spoon up pineapple.’
 (elicitation, BAR18082014CH_02 00:29:39.120-00:29:43.290)
- e. Na’am [kayuh ipak] pena’up koq utup neh.
 NEG wood chopped IV.partner for partner 3SG.2
 ‘There was no chopped wood to be used as a partner for him (the log on the fire).’
 (text, BAR04092014CH_04 00:01:50.800-00:01:55.040)

Consequently, I analyse the actor as subject in AV, the undergoer as subject in UV and the instrument as subject in IV based on shared behavioural, distributional and coding properties. Furthermore, I argue that this supports the Manning (1996) approach to grammatical functions in syntactically ergative and Philippine-type languages in that Kelabit is another language with the predicted property split. This suggests that ‘subject’ does not need to be abandoned in Western Austronesian but rather identified by its ‘reference-related’ or pivot properties.

2.5.1.2 Non-subject Core Arguments

Whilst there are a number of subject properties shared by the ASV, it is harder to find specific non-subject core argument properties in Kelabit. Nonetheless, there are properties shared by the undergoer of an AV construction, the actor of a UV construction and the actor and undergoer of an IV construction. These motivate the concept of a non-subject core function and support an analysis of the alternations as syntactically symmetrical (see SUBSECTION 1.4.2).

Non-subject core arguments can be distinguished from subjects in that they do not have the subject properties outlined in SUBSECTION 2.5.1.1. However, they also differ from obliques in a number of ways (see SUBSECTION 2.5.1.3). Firstly, core arguments are typically realised as NPs, whilst obliques are PPs:

(201) *Coding of Non-subject Core Arguments*

a. **Actor Voice**

[La'ih sineh] _{NP}	ne-merey	[nubaq] _{NP}	[ngen anak nedih] _{PP} .
man DEM	PFV-AV.give	rice	to child 3SG.POSS
Subject		Core	Oblique

‘The man gave rice to his child.’

(elicitation, BAR30072014CH_03 00:02:25.520-00:02:31.350)

b. **Undergoer Voice**

Birey	[neh] _{NP}	[nubaq] _{NP}	[ngen anak-adiq] _{PP} .
UV.PFV.give	3SG.2	rice	to child-PL
	Core	Subject	Oblique
‘He gave rice to the children.’			(elicitation, fieldnotes)

In (201), the oblique goal is a PP, headed by the preposition *ngen* ‘to/with’. Both the subject and the non-subject core argument are NPs. This is true irrespective of whether the main verb is AV or UV, in contrast to Lundayeh, where non-subject undergoers in AV are typically oblique (SUBSECTION 4.2.1.2). Indeed, pronominal non-subject core actors in UV can be expressed using FORM 2 pronouns, which are generally used for non-subject core functions in main clauses (see SUBSECTION 4.2). Hence, both AV and UV appear to contain two core arguments.

Secondly, non-subject core arguments typically occupy the immediately post-verbal position.¹²⁴ It is ungrammatical for an adjunct to intervene between the verb and its non-subject core argument:

(202) **Actor Voice**

- a. Uih [ne-kuman buaq kabər] ngimalem.
 1SG.1 PFV-AV.eat fruit pineapple yesterday
 ‘I ate pineapple yesterday.’
 (elicitation, BAR18082014CH_02 00:17:12.730-00:17:15.520)
- b. *Uih ne-kuman ngimalem buaq kabər.
 1SG.1 PFV-AV.eat yesterday fruit pineapple
 For: ‘I ate pineapple yesterday.’
 (elicitation, BAR18082014CH_02 00:17:28.440-00:17:32.210)

(203) **Undergoer Voice**

- a. [Kinan kuh] ngimalem neh buaq kabər ih.
 UV.PFV.eat 1SG.2 yesterday PT fruit pineapple PT
 ‘I ate the pineapple yesterday.’
 (elicitation, BAR18082014CH_02 00:21:11.370-00:21:20.540)

¹²⁴ The exception is VSO order in AV (see SUBSECTION 5.5.1.2 for discussion).

- b. *Kinan ngimalem kuh neh buaq kaber.
 UV.PFV.eat yesterday 1SG.2 PT fruit pineapple
 For: ‘I ate the pineapple yesterday.’
 (elicitation, BAR18082014CH_02 00:22:07.600-00:22:11.880)

(204) **Instrumental Voice**

- a. [Penekul kuh nubaq] ngimalem tekul ih.
 IV.spoon 1SG.2 rice yesterday spoon PT
 ‘I used a spoon to spoon up rice yesterday.’
 (elicitation, BAR18082014CH_02 00:32:42.000-00:32:45.950)

Finally, non-subject core arguments cannot appear in pre-verbal position, as shown in SUBSECTION 2.5.1.1. In contrast, adjunct PPs can appear initially:

(205) **Adjuncts in initial position**¹²⁵

- a. [Ngi bawang lun beken] kuman lemulun deley kinih.
 at place people other AV.eat people corn now
 ‘In other places, people eat corn today.’
 (text, PDA06112013CH_06 00:07:44.567-00:07:48.420)
- b. [Let ngineh] saget neh video dih senaruq mayaq social media.
 from there fast PT video DEM UV.PFV.do follow social media
 ‘From there, videos were quick to appear on social media.’
 (text, BAR02092014CH_03 00:06:00.362-00:06:07.674)

Hence, non-subject core-arguments share the property of being realised as NPs rather than PPs, appearing in the immediately post-verbal position and the constraint against appearing in initial position. Since this applies equally to the undergoer in AV and the actor in UV, the alternations can be considered symmetrical (see SUBSECTION 1.4.2).

¹²⁵ This appears to depend on the type of PP, since some PPs cannot appear pre-verbally:

- (i) *[Luun asuq] tudo uih.
 on stool sit 1SG.1
 For: ‘I sit on the stool.’ (elicitation, BAR18082014CH_01 00:56:58.520-00:57:03.530)

The PP in such cases could be considered a derived argument in the sense of Needham & Toivonen (2011). I suspect that it would also be ungrammatical for obliques to appear initially.

2.5.1.3 Obliques and Adjuncts

Finally, obliques and adjuncts can be distinguished from core arguments in that they are not realised as NPs but rather as PPs. Obliques typically follow the predicate and any core arguments:¹²⁶

(206) **Obliques**

- a. Tak betoq ideh bu'uh [ngen lun merar ih].
if PT 3PL.1 angry with people big PT
'If they are angry at the elders.'
(elicitation, BAR18082014CH_01 00:01:26.660-00:01:29.770)
- b. Ngabit ko [ibal bera] [ngen Sineh Raben] betoq.
AV.lend 2SG.1 some rice to Proper Name PT
'Lend some rice to Sineh Raben.'
(elicitation, BAR30072014CH_01 00:01:20.560-00:01:25.360)

These are distinguished from adjuncts in that they are subcategorised for by the predicate.

Adjuncts are optional and can occur in various positions, including clause-initially, clause-finally and inside a VP that appears in initial position:¹²⁷

(207) *Adjuncts*

- a. **Clause-initially**
[Ngimalem] ne-kuman buaq kaber uih.
yesterday PFV-AV.eat fruit pineapple 1SG.1
'I ate pineapple yesterday.'
(elicitation, BAR18082014CH_02 00:16:58.290-00:17:01.480)

¹²⁶ There are no naturally occurring examples in the corpus where an oblique precedes a non-subject core argument.

¹²⁷ It is possible that adjuncts form separate intonation units when they appear in initial position but the rest of the clause remains predicate-initial as above (cf. Lee & Billings 2005: 246). Kroeger (1993) suggests that there are three different constructions in which adjuncts occur before the verb in Tagalog. This remains to be further explored in Kelabit but it is certainly possible to find Kelabit clauses with initial adjuncts and particles; initial adjuncts, no particles and SVO order subsequently and initial adjuncts, no particles and predicate-initial order subsequently.

b. **Clause-finally**

Ne-kuman buaq kaber uih [ngimalem].
PFV-AV.eat fruit pineapple 1SG.1 yesterday
'I ate pineapple yesterday.'
(elicitation, BAR18082014CH_02 00:16:43.530-00:16:48.520)

c. **VP-internal**

[Ne-kuman buaq kaber [ngimalem]] uih.
PFV-AV.eat fruit pineapple yesterday 1SG.1
'I ate pineapple yesterday.'
(elicitation, BAR18082014CH_02 00:16:52.140-00:16:56.140)

Hence, adjuncts do not appear to be subject to the same word-order restrictions as non-subject core arguments and obliques. As illustrated in SUBSECTION 2.5.1.2, it is ungrammatical for adjuncts to appear in the immediately post-verbal position, regardless of the word order of the clause.

Similarly, there can be any number of adjuncts in a given clause, and they can occur in different orders with respect to each other:

(208) **Adjuncts**

- a. Kinan kuh neh buaq kaber [luun asuq] [ngimalem].
UV.PFV.eat 1SG.2 PT fruit pineapple on stool yesterday
'I ate pineapple on the stool yesterday.'
- b. Kinan kuh neh buaq kaber [ngimalem] [luun asuq].
UV.PFV.eat 1SG.2 PT fruit pineapple yesterday on stool
'I ate pineapple yesterday on the stool.'
(elicitation, BAR18082014CH_02 00:21:58.400-00:22:06.930)

Hence, obliques and adjuncts differ from core arguments in terms of their coding. Moreover, they differ from each other in terms of distribution. Obliques appear within the VP, following non-subject core arguments. Adjuncts are less restricted in their position, and any number may occur in a given clause.

Thus, it is possible to identify grammatical functions in Kelabit and these are determined by the voice construction in the following manner:

(209) **Grammatical Functions**

- a. Subject = ASV
- b. Non-subject core = undergoer in AV, actor in UV, both in IV
- c. Oblique = PP subcategorised for by verb
- d. Adjunct = Any other constituent

This suggests that the Kelabit voice system enables an alternation in the mapping of arguments to functions. Hence, Kelabit voice is similar to active/passive and ergative/antipassive alternations and differs only in that the alternations are symmetrical (see SUBSECTION 3.2.1).

2.5.2 Periphrastic Voices

In SUBSECTION 2.5.1, I argued that the voice system functions to map different semantic roles to subject in Kelabit. In AV, the actor is mapped to subject. In UV, the undergoer is mapped to subject and in IV, the instrument is mapped to subject. In order to map other semantic roles to subject, periphrastic constructions are used, namely clauses with *inan* ‘to have/to exist’ and clauses with *tu'en* ‘UV.IRR.do’. It remains for future research to explore whether such constructions should be analysed as mono-clausal or biclausal.

2.5.2.1 *Inan* clauses

The basic function of *inan* is to form existential clauses. In parallel with quantificational structures involving *mulaq* ‘many’ and *na’am* ‘negative’, *inan* occurs clause-initially.¹²⁸

(210) **Existential Clauses**

- a. *Inan buaq udung kayuh sineh.*
EXIST fruit top tree DEM
‘There is fruit at the top of the tree.’
(elicitation, BAR30102013CH_03 00:07:39.122-00:07:41.727)

(211) **Quantificational Structures**

- a. *Mulaq pirit lem latiq kamih malem.*
many sparrows in field 1PL.EXCL before
‘There were many sparrows in our fields in the past.’
(elicitation, BAR30072014CH_01 00:44:09.429-00:44:12.092)

Negative Clauses

- b. *Na’am teh luang dingi.*
NEG PT fish inside
‘There were no fish inside.’
(text, BAR17082014CH_08 00:01:33.050-00:01:34.280)

Inan can also be used to express possession and as a noun meaning ‘place’:

(212) a. **Possession**

- rengaq ko inan masa*
if 2SG.1 have time
‘if you have time’
(text, BAR29112013CH_01 00:06:12.340-00:06:14.150)

b. **Place**

- Uih mekaaq [inan lajang sineh]_{NP}.*
1SG.1 AV.change place pot DEM
‘I’ll change the position of that pot.’
(elicitation, BAR28102013CH_03 00:08:36.534-00:08:39.912)

¹²⁸ Many people now express negative existentials using the combination *na’am inan*, possibly in analogy with Malay *tidak ada* ‘NEG exist’. There is an additional negator in Kelabit, *buken*. This appears cognate with Malay *bukan* but can seemingly be used to negate verbs as well as nouns. Himmelmann (2005a) describes morphologically independent existential and negative existential particles as a Philippine-type characteristic.

Hence, *inan* has a number of functions in ‘simple’ clauses.

As a periphrastic voice, *inan* is used to map peripheral arguments to subject. This typically occurs in the context of relative clauses (SUBSECTION 2.5.3.1) and can also be used as a periphrastic IV construction. Generally, *inan* is followed by the non-subject actor and then a predicate marked with AV morphology.¹²⁹ The non-subject actor is expressed using the FORM 2 pronoun and all other arguments follow the second predicate:

(213) *Mapping Peripheral Arguments to Subject*

a. **Goal**

[Peter] inan John ne-merey buaq kaber ih.
 Peter have John PFV-AV.give fruit pineapple PT
 ‘John gave Peter the pineapple.’ (elicitation, fieldnotes)

b. **Recipient**

[Mulaq lun ineh] kereb inan narih masiu.
 many people DEM can have IMPERS AV.sell
 ‘There were lots of people to sell (beads) to.’
 (text, BAR08092014CH_02 00:07:05.530-00:07:08.660)

c. **Locative**

[Award ceremony] inan tauh merey prize.
 award ceremony have 1PL.INCL AV.give prize
 ‘An award ceremony where we give prizes.’
 (text, BAR21082014CH_08 00:03:52.228-00:03:56.315)

d. **Comitative**

[Kawan] nuk inan kuh pep-uto.
 friend REL have 1SG.2 RECP-tease
 ‘A friend that I used to tease and get teased by.’
 (text, PUM18102013CH_05 00:00:41.570-00:00:45.190)

¹²⁹ It is also possible for *inan* to be followed by the undergoer, in which case the lower predicate is marked with UV morphology:

(i) Na’am Peter inan buaq kaber birey John.
 NEG Peter have fruit pineapple UV.PFV.give John
 ‘John didn’t give Peter any pineapple.’ (elicitation, fieldnotes)

This suggests that the structure may be bi-clausal and that the argument directly following *inan* must also be the subject of the lower clause (see SUBSECTION 2.5.3.3 on complement clauses).

- e. **Theme**
 [Enun] inan dulun pelaba tu'uh dooq pian tebeyq ken narih koq?
 what have others very true good want PT say IMPERS PT
 'what is it that other people really like, I wonder?'
 (text, BAR21082014CH_06 00:06:41:630-00:06:45.640)
- f. **Instrument**
 [Seduk] suk inan neh ne-nekul nubaq nedih.
 spoon REL have 3SG.2 PFV-AV.spoon rice 3SG.POSS
 'The spoon that he used to spoon up his rice.'
 (elicitation, fieldnotes)

As shown in (213), this can be used to promote peripheral arguments of both transitive and intransitive predicates to subject (shown in brackets), which allows them to appear clause-initially.

In addition to *inan*, the borrowed form *pakai* can also be used to as a periphrastic IV construction:

- (214) **Periphrastic Construction with *pakai***
 a. Enun pakai neh ngeluit.
 what use 3SG.2 AV.fish
 'what he uses to fish.'
 (text, BAR17082014CH_03 00:01:26.440-00:01:27.960)

2.5.2.2 *Tu'en* clauses

There is also a periphrastic UV construction in Kelabit, using *tu'en*, the UV irrealis form of the verb 'to do/put'. Like other irrealis UV verbs, *tu'en* can be used as a main verb:

- (215) ***Tu'en* as a main verb**
 Mo, tu'en kuh idih.
 yes, UV.IRR.do 1SG.2 DEM
 'Yes, I'll do it.'
 (elicitation, BAR14102013CH_01 01:20:52.389-01:20:54.260)

As a periphrastic construction, *tu'en* is typically followed by a non-subject actor, expressed as a FORM 2 pronoun, and a predicate marked with AV morphology. The undergoer subject can occur clause-initially or clause-finally, much like in morphological UV clauses (SUBSECTION 5.5.1.3).

Tu'en clauses are often used instead of morphological UV irrealis forms. As such, they tend to fulfill irrealis functions, such as imperatives, and are common in procedural texts:

(216) *Function of tu'en clauses*

a. **Imperative**

Tu'en	narih	nge-lulun	[epin neh]	na'an.
UV.IRR.do	IMPERS	AV-roll	mat DEM	later

'Roll the mats up later.'

(elicitation, BAR15102013CH_01 01:17:30.172-01:17:33.593)

b. **Generic Statement/Procedure**

Tu'en	narih	milit ngen wey	[nidih].
UV.IRR.do	IMPERS	AV.tie with rattan	PT=DEM

'You tie it together with rattan.'

(text, BAR27102013CH_01 00:01:18.348-00:01:20.497)

In casual speech, *tu'en* is often shortened to *en*:

(217) **en clauses**

a. En	deh nawar	teretek	[ieh] keyh.
do	3PL.2 AV.call	on.purpose	3SG.1 PT

'And they call it (the spirit) on purpose.'

(text, PUM18102013CH_17 00:07:18.105-00:07:20.105)

Periphrastic constructions are common in the languages of Sarawak, including Lundayeh, Sa'ban and Kayan, but relatively infrequent in the languages of Sabah (cf. Clayre 2002).¹³⁰

¹³⁰ Nb. Boutin (1996) describes a 'periphrastic passive' in Bonggi.

2.5.3 Multi-clausal Constructions

In the final section, I discuss multi-clausal constructions, including relative clauses (SUBSECTION 2.5.3.1), cleft constructions (SUBSECTION 2.5.3.2), complement clauses (SUBSECTION 2.5.3.3), adjunct clauses (SUBSECTION 2.5.3.4) and co-ordination strategies (SUBSECTION 2.5.3.5). These provide additional support for analysing the ASV as subject (see SUBSECTION 2.5.1.1).

2.5.3.1 Relative Clauses

Relative clauses in Kelabit are post-nominal. However, relativisation involves a number of different strategies, depending on the syntactic status of the argument being relativised on. The primary strategy for relativisation is the gap strategy, which is used for the relativisation of subjects. Non-subject core arguments and peripheral arguments must first be mapped to subject via voice morphology or periphrastic constructions before they can be relativised (see SUBSECTION 2.5.1.1). A second strategy for relativisation is resumptive pronouns, which are used to relativise on possessors.

In SUBSECTION 2.5.1.1, I demonstrated that Kelabit shares the Western Austronesian restriction against relativisation of non-subject arguments. This is sometimes known as the Object Extraction Restriction (Aldridge 2004, 2008 etc.). The same restrictions do not apply for transitive clauses in which the verb is not overtly voice-marked (see SUBSECTION 2.4.2.2.2). In such clauses, either the actor or the undergoer can be relativised, creating ambiguity:

- (218) **Ambiguity with bare predicates**
- a. Seni'er kuh la'ih [suk keliq John].
UV.PFV.see 1SG.2 man REL know John
'I saw the man who knew John.'
OR: 'I saw the man who John knew.' (elicitation, fieldnotes)

Interestingly, this is also true of other Western Austronesian languages with the ‘extraction restriction’, including Indonesian (Cole, Hermon & Yanti 2008) and Tagalog (Jed Pizarro-Guevara p.c.) and remains to be further explored.

Relativisation constitutes good evidence for treating *tu’an* as a remnant locative voice form in Kelabit (SUBSECTION 2.4.1.3.4), since the locative can be relativised on in this construction:

(219) **Locative Voice**

- a. Seni’er kuh **lidung** [suk tu’an neh babeh nedih].
 UV.PFV.see 1SG.2 corner REL put.LV 3SG.2 bag 3SG.POSS
 ‘I saw the corner where he put his bag.’ (elicitation, fieldnotes)

In all other cases, peripheral arguments must be mapped to subject via periphrastic constructions before they can be relativised (see SUBSECTION 2.5.2):

(220) *Relativisation of Peripheral Arguments*

a. **Goal Subject**

- Seni’er kuh **anak** [suk inan neh ne-merey nubaq].
 UV.PFV.see 1SG.2 child REL have 3SG.2 PFV-AV.give rice
 ‘I saw the child that he gave rice to.’ (elicitation, fieldnotes)

b. **Locative Subject**

- Keliq kuh **kedai** [suk inan neh ne-belih nubaq] dih.
 know 1SG.2 shop REL have 3SG.2 PFV-buy rice DEM
 ‘I know the shop where he bought rice.’ (elicitation, fieldnotes)

c. **Theme Subject**

- Edteh** [nuk inan keduih sekenan dooq~dooq] bah.
 one REL have 1SG.EMPH remember REDUP~good EXCL
 ‘Something that I remember well.’
 (text, BAR22102013CH_05 00:07:59.200-00:08:02.330)

Similarly, the undergoer of an AV clause can be relativised as the subject of a *tu’en* clause:

- (221) *Relativising with tu'en*
 a. **Undergoer Subject**
 Seni'er kuh **nubaq** [suk tu'en neh kuman].
 UV.PFV.see 1SG.2 rice REL UV.IRR.do 3SG.2 AV.eat
 'I saw the rice that the man ate.' (elicitation, fieldnotes)

Thus, in order to be relativised using the gap strategy, an argument must be mapped to subject via a morphological voice construction or a periphrastic voice construction.

Exactly the same patterns hold of long-distance relativisation. The relativised argument must be the subject of its clause (shown in brackets):

- (222) *Long-distance Relativisation*
 a. **Actor Voice**
 Seni'er kuh **la'ih** [suk tu'en kuh ngelinuh
 UV.PFV.see 1SG.2 man REL UV.IRR.do 1SG.2 AV.think
 [masaq bukuh ih]].
 AV.read book PT
 'I saw the man that I thought was reading a book.'
- b. **Undergoer Voice**
 Seni'er kuh **bukuh** [suk tu'en kuh ngelinuh
 UV.PFV.see 1SG.2 book REL UV.IRR.do 1SG AV.think
 [tu'en la'ih sineh masaq ih]].
 UV.IRR.do man DEM AV.read PT
 'I saw the book that I thought the man was reading.'
- c. **Instrumental Voice**
 Seni'er kuh **tekul** [suk tu'en kuh ngelinuh
 UV.PFV.see 1SG.2 spoon REL UV.IRR.do 1SG.2 AV.think
 [penekul la'ih sineh nubaq nedih]].
 IV.spoon man DEM rice 3SG.POSS
 'I saw the spoon that I thought the man used to scoop up his rice.'
 (elicitation, fieldnotes)

Possessors are low on the Keenan & Comrie (1979) Accessibility Hierarchy and are relativised using an alternative strategy, namely resumptive pronouns:

(223) **Relativisation of Possessors**

- a. Seni'er kuh la'ih [suk tesineh nedih ma'it] ih.
UV.PFV.see 1SG.2 man REL mother 3SG.POSS INTR.ill PT
'I saw the man whose mother is ill.'
- b. Seni'er kuh la'ih [suk ukuq nedih ne-upun buro].
UV.PFV.see 1SG.2 man REL dog 3SG.POSS PFV-run away
'I saw the man whose dog ran away.' (elicitation, fieldnotes)

It is ungrammatical to omit the resumptive pronoun:

- (224) *Seni'er kuh la'ih [suk ukuq ne-upun buro].
UV.PFV.see 1SG.2 man REL dog PFV-run away
For: 'I saw the man whose dog ran away.'
(elicitation, fieldnotes)

Hence, relativisation and long-distance relativisation support analysing the ASV as subject, as other arguments cannot be relativised or are relativised using a different strategy.

2.5.3.2 Cleft Constructions

Relativisers are also used in cleft-constructions, which are subject to the same 'extraction' restrictions as relative clauses. Clefting is used as a strategy in focus constructions and question formation. The examples in (225) illustrate clefting as a strategy to focus material to the left of the cleft:

- (225) *Clefting*
- a. **Actor Voice**
[Dih ieh dih] suk laq kuman ih.
DEM 3SG.1 DEM REL DESID AV.eat PT
'It's him (pointing) who wants to eat.'
(elicitation, BAR19082014CH_03 00:06:34.150-00:06:35.940)

b. **Undergoer Voice**

[Buaq kaber] suk kenen Peter ih.
fruit pineapple REL UV.IRR.eat Peter PT
'It's pineapple that Peter will eat.'

(elicitation, BAR19082014CH_03 00:10:52.965-00:10:57.585)

The cleft in (225a) could answer the question 'who wants to eat?', whilst the cleft in (225b) could answer the question 'what did Peter eat?' Hence, the clefts are used to represent the focus information in the clause (cf. Lambrecht 1994, CHAPTER 5). Only the actor can be clefted in AV, and the undergoer in UV, which supports an analysis of ASV as subject.

Clefting is also used in question-formation. Much like relative clauses, only subjects (i.e. the ASV) can be questioned using a *wh*-cleft or pseudo-cleft:¹³¹

(226) *Clefting*

a. **Actor Voice**

[lih] suk kuman buaq kaber?
who REL AV.eat fruit pineapple
'Who is it that eats pineapple?'

(elicitation, BAR19082014CH_03 00:09:51:510-00:09:52.930)

b. **Undergoer Voice**

[Enun] suk kenen Peter ih?
what REL UV.IRR.eat Peter PT
'What is it that will Peter eat?'

(elicitation, BAR19082014CH_03 00:10:31.360-00:10:35.200)

Clefting is not the only strategy for question formation. In Kelabit, it is also possible for *wh*-words to appear in initial position without the relativiser. This is only grammatical for subjects, i.e. the ASV (SUBSECTION 2.4.2.9):

¹³¹As discussed in Potsdam & Polinsky (2012), it is difficult to tell whether (226) constitutes a cleft or a pseudo-cleft, given that Kelabit allows nominal predicates (SUBSECTION 2.4.2.1) and headless relative clauses (SUBSECTION 2.4.2.10) and does not have an overt expletive subject, e.g. in existential constructions (SUBSECTION 2.5.2.1). This would require further study.

- (227) *Wh-first*
- a. **Actor Voice**
 [lih] kuman buaq kabər?
 who AV.eat fruit pineapple
 ‘Who eats pineapple?’
 (elicitation, BAR19082014CH_03 00:28:46.035-00:28:48.300)
- b. **Undergoer Voice**
 [Enun] seni`er muh?
 what UV.IRR.see 2SG.2
 ‘What did you see?’
 (pear story, BAR31072014CH_06 00:00:09.630-00:00:11.330)

To question a non-subject core argument, a *wh*-in situ strategy is employed:

- (228) *Wh-in situ*
- a. **Actor Voice**
 Kuman [enun] teh Peter?
 AV.eat what PT Peter
 ‘What does Peter eat?’
 (elicitation, BAR19082014CH_03 00:12:17.170-00:12:18.460)
- b. **Undergoer Voice**
 Kenen [iih] buaq kabər sineh?
 UV.IRR.eat who fruit pineapple DEM
 ‘Who will eat this pineapple?’
 (elicitation, BAR19082014CH_03 00:11:22.955-00:11:24.975)

Hence, question formation strategies, both clefting and *wh*-first, support the distinction between subjects and other core arguments.

2.5.3.3 Complement Clauses

There are several verbs which appear to take complement clauses in Kelabit.¹³² These include:

¹³² Nb. since complement clauses are not typically marked with an overt complementiser and there is no overt marking of finiteness, it is sometimes difficult to say whether clauses are embedded or simply juxtaposed.

(229) **Verbs taking Clausal Complements**

- a. *mutuh* ‘request’
- b. *nuruq* ‘order’
- c. *merey* ‘allow/permit’
- d. *naruq* ‘cause’
- e. *nutun* ‘try’

The predicates in (229) are control predicates and trigger argument sharing between higher and lower clauses.¹³³ Typically, the argument immediately following the verb fulfils the function of non-subject core argument in the higher clause and subject in the lower clause:

(230) *Control Constructions*

a. **Actor Voice**

Uih ne-nuruq **ieh** [nge-laak ngen tauh].
1SG.1 PFV-AV.order 3SG.1 AV-cook for 1PL.INCL
‘I asked him to cook for us.’

(experiment, BAR19082014CH_02 00:01:14.611-00:01:17.118)

b. **Undergoer Voice**

Ieh merey **padey** [sebuwen kuh].
3SG.1 AV.give rice UV.IRR.plant 1SG.2
‘He allows rice to be planted by me.’ (elicitation, fieldnotes)

- c. *Ieh merey **padey** [nibu uih].¹³⁴
3SG.1 AV.give rice AV.plant 1SG.1
For: ‘He allows me to plant rice.’ (elicitation, fieldnotes)

As shown in (230), if the controlled argument is an actor, then the lower clause predicate must be AV. In contrast, if the controlled argument is an undergoer, then the lower clause predicate must be in UV.¹³⁵

¹³³ As discussed in SUBSECTION 2.5.2, *inan* and *tu'en* clauses have a similar structure. These could also be considered cases of pro-drop.

¹³⁴ Note that, as with many ungrammatical sentences in this thesis, this could have the semantically odd interpretation of rice planting the speaker.

¹³⁵ Similar patterns obtain in languages like Balinese and Indonesian (Riesberg 2014: 37-42). In Tagalog, the controlled element is the actor (cf. Schachter 1976: 504). However, Kroeger (1993) notes some exceptions. This can be explained by semantic restrictions (see CHAPTER 1 and Riesberg 2014 for discussion).

The shared argument may have the function of subject or non-subject core argument in the higher clause, so long as it is the subject of the lower clause:

(231) *Control Constructions*

a. **Ditransitive Subject**

Senuruq neh **ieh** [nibu padey].
 UV.PFV.order 3SG.2 3SG.1 AV.plant rice
 ‘He asked him to plant rice.’ (elicitation, fieldnotes)

b. **Transitive Subject**

Uih ne-nutun [nibu padey].
 1SG.1 PFV-AV.try AV.plant rice
 ‘I tried to plant rice.’ (elicitation, fieldnotes)

c. **Transitive Non-Subject**

Senutun **kuh** [nibu padey].
 UV.PFV.try 1SG.2 AV.plant rice
 ‘I tried to plant rice.’ (elicitation, fieldnotes)

d. ***Senutun kuh** [sebuwen padey].

UV.PFV.try 1SG.2 UV.IRR.plant rice
 For: ‘I tried to plant rice.’ (elicitation, fieldnotes)

Hence, control constructions also support the analysis of ASV as subject.

As for predicates like ‘think’ and ‘say’ that take closed clausal complements in other languages, most commonly the two clauses are simply in juxtaposition:

(232) **Complement Clauses with verbs of speaking**

a. John ne-mala [iko m-edting].

John PFV-AV.say 2SG.1 INTR-arrive
 ‘John said you arrived.’
 (elicitation, BAR21102013CH_02 00:34:24.918-00:34:28.553)

b. Am kekamih kekeliq [ideh nge-linuh
 NEG 1PL.EXCL.EMPH know 3PL.1 AV-think

[narih dooq intelligent]].
 IMPERS good intelligent
 ‘We didn’t know they thought we were intelligent.’
 (text, BAR21082014CH_06 00:07:18.910-00:07:23.490)

However, there are two forms that may function as complementisers: the particle *ken* and the preposition *ngen* ‘with’.¹³⁶ The question particle *ken* could be grammaticalising into a complementiser for verbs of speaking and thinking in non-factive contexts:

(233) ***Ken as Complementiser?***

- a. Am tuih keliq [ken birey deh mooq~mooq idih].
 NEG PT=1SG.1 know Q UV.PFV.give 3PL.2 REDUP~free DEM
 ‘I don’t know if they were giving them away for free.’
 (text, BAR21082014CH_05 00:08:15.340-00:08:17.380)

However, there are only a few occurrences in the corpus and it is not clear if these are used as complementisers or simply reflect direct speech (see SUBSECTION 2.4.2.9).

Finally, factive predicates such as *gagap* ‘surprised’ and *repet* ‘hope’ take a complement clause that begins with the preposition *ngen* ‘to/with’. In cases where the subject of the subordinate clause is a pronoun, the FORM 2 pronouns can be used:

(234) ***Ngen as Complementiser?***

- a. Gagap tuih [ngeneh mala anjing ngekuh].
 surprised PT=1SG.1 to.3SG.2 AV.say dog to.1SG.2
 ‘I was surprised that he said dog to me.’
 (text, BAR25102013CH_03 00:07:48.990-00:07:51.615)

¹³⁶ Sometimes *laq* is also used with the predicate *ngelinuh* ‘think’, such as in (i) and (ii):

- (i) Am tebut tuih ngelinuh laq ngitun department kamih.
 NEG EMPH PT=1SG.1 AV.think ? AV.ask department 1PL.EXCL.POSS
 ‘I didn’t even think of asking our department.’
 (text, BAR21082014CH_05 00:04:26.320-00:04:30.330)
- (ii) Am tuih ngelinuh laq muliq mey Bario
 Neg pt=1sg.1 av.think ? intr.return go Bario
 ‘I wasn’t thinking about coming back to Bario.’
 (text, BAR22102013CH_04 00:06:30.110-00:06:34.075)

This doesn’t seem to convey desiderative mood, but may indicate the irrealis status of the subordinate clause. There are very few examples in the corpus so it remains to be seen if *laq* ‘want’ may also be grammaticalising as a complementiser.

- b. Repet [ngemuh dooq nangey].
 hope to.2SG.2 good there
 ‘Hope you are well over there.’ (elicitation, fieldnotes)

Thus, there are at least two types of embedded clause in Kelabit: those whose subject is shared with an argument in the higher clause, and those with separate arguments. In the second instance, actor subjects may be expressed through FORM 2 pronouns and potential complementisers, such as the preposition *ngen*, may be used. These serve to distinguish main and subordinate clauses.

2.5.3.4 Adjunct Clauses

Kelabit also has a series of adjunct clauses that are not subcategorised for by the verb but add extra information about the state of affairs expressed in the main clause. Adjunct clauses are typically introduced by conjunctions (SUBSECTION 2.4.2.11). Much like phrasal adjuncts, they can occur before or after the main clause:

(235) **Adjunct Clauses**

- a. [**Rengaq** narih anak-adiq ngilad] mey narih mayaq lun uwan
 when IMPERS child-PL past go IMPERS follow parents

 narih mey lem pulung.
 IMPERS to in forest
 ‘When we were young, we followed our parents into the jungle.’
 (text, BAR25102013CH_04 00:00:00.820-00:00:05.940)
- b. [**Tulu** deh laq belajar terun] kereb teh narih madaq
 if 3PL.2 DESID learn maybe can PT IMPERS AV.show

 ih ngedeh.
 PT to.3PL.2
 ‘If they want to learn, I can show them.’
 (text, BAR10092014CH_02 00:06:47.250-00:06:51.010)

- c. Na'am neh lun nu'uh ieh, [**kadiq** ieh sido].
 NEG PT people AV.look.after 3SG.1 so 3SG.1 <INTR>grief.song
 'There's no-one to look after him, so he writes a grief song.'
 (text, BAR04092014CH_04 00:00:42.680-00:00:45.060)
- d. [**Tak** uih mey la'ud] beliyeu kuh ayuq
 if 1SG.1 go downstream UV.IRR.buy 1SG.2 PT
 teh ibal [iteu m-uliq].
 PT some UV.IRR.bring INTR-back
 'When I go to town, I buy a few to bring back.'
 (pear story, BAR02092014CH_01 00:01:18.365-00:01:22.425)

Much like in complement clauses, the FORM 2 pronoun can sometimes express subjects in adjunct subordinate clauses, such as (235b).

2.5.3.5 Co-ordination

Finally, Kelabit has a number of different strategies for co-ordination. The first is using the demonstrative *idih*:

- (236) **Co-ordination with *idih***
 a. Peter [[ne-tudo] **idih** [ne-kuman ba'ung nedih]].
 Peter PFV-sit and PFV-AV.eat banana 3SG.POSS
 'Peter sat and ate his banana.'
 (elicitation, BAR21102013CH_02 00:21:06.837-00:21:12.337)

Two VPs can only be co-ordinated if they have the same subject. They cannot be co-ordinated if they share the same argument, but it is mapped to subject in one conjunct and non-subject core argument in the other:

- (237) **Constraints on Co-ordination**
 a. *[[Kenen Peter edteh ba'ung] **idih** [mirup kopi]].
 UV.IRR.eat Peter one banana and AV.drink coffee
 For: 'Peter eats a banana and drinks coffee.'
 (elicitation, BAR21102013CH_02 00:29:14.864-00:29:23.391)

This is only grammatical if the actor is repeated in the second conjunct, and implies temporal ordering of the two events:

(238) **Constraints on Co-ordination**

- a. [[Kenen Peter edteh ba'ung] **idih** [tieh mirup kopi]].
 UV.IRR.eat Peter one banana and PT=3SG.1 AV.drink coffee
 'Peter will eat a banana and then he'll drink coffee.'
 (elicitation, BAR21102013CH_02 00:32:07.012-00:32:13.661)

Hence, co-ordination also supports an analysis of ASV, rather than actor, as subject.

The second method of co-ordinating is using *kineh teh* 'like that'. This is used for co-ordinating NPs and PPs as well as clauses:

(239) **PP Co-ordination with *kineh teh***

- a. [[ngi England] **kineh teh** [ngi Bario]]_{PP}
 at England and at Bario
 'in England and in Bario' (elicitation, fieldnotes)

NP Co-ordination with *kineh teh*

- b. lem erang [[Ukraine] **kineh teh** [Russia]]_{NP}
 in between Ukraine and Russia
 'between Ukraine and Russia'
 (text, BAR21102014CH_01 00:11:29.253-00:11:31.883)

Clausal Co-ordination with *kineh teh*

- c. [[Kekamih sediaq laq kerja paad~paad ngen FAS
 1PL.EXCL.EMPH ready DESID work REDUP~equal with FAS
 pingan inih] **kineh teh** [kekamih repet ngen nuk
 after DEM and 1PL.EXCL.EMPH hope that REL
 ko'ayuq inih dih na'am tu'en dulun beruh]]_{Clause}.
 like DEM DEM NEG UV.IRR.do other.people again
 'We are ready to work together with the FAS (Sarawak Football
 Association) from now on and hope that things like this do not
 happen again.'
 (text, BAR02092014CH_03 00:05:28.258-00:05:41.143)

Clauses and phrases can also be co-ordinated using the preposition/verb *mey*

‘go’.¹³⁷

(240) **NP Co-ordination with *mey***

- a. Ngarang neh [[John]**mey** [kenanak nedih]]_{NP}.¹³⁸
 dance PT John and sibling 3SG.POSS
 ‘John and his brother are dancing.’

(elicitation, BAR21102013CH_01 00:35:11.813-00:35:16.948)

Clausal Co-ordination with *mey*

- b. [Masiu~masiu neh kamih **mey** menad kayuh]_{Clause}.
 REDUP~AV.sell PT 1PL.EXCL and AV.climb tree
 ‘We played at selling and climbed trees.’

(text, BAR08092014CH_05 00:03:45.840-00:03:49.280)

Co-ordination as afterthought

- c. Kurang-lebih rinat kamih nih tupu teh turun
 less-more generation 1PL.EXCL DEM only PT perhaps

mey ibal anak-adiq nuk ngi sekolah ngi Bario ih.
 and some child-PL REL at school at Bario pt
 ‘More or less it is just our generation, and a few children who go to school in Bario.’

(text, BAR21082014CH_09 00:04:26.230-00:04:33.980)

Finally, particles like *men*, *metoq* and *meteh* can be used to connect clauses that both happen simultaneously:

(241) **Co-ordination with Particles**

[Laq buro neh dieh adaq ih] meteh [laq matey ieh].
 DESID away PT 3SG.POSS spirit PT and DESID INTR-die 3SG.1
 ‘Her spirit would go away and should would die.’

(text, PUM18102013CH_17 00:06:59.190-00:07:01.740)

¹³⁷ Much like subordination, it is sometimes difficult to tell if clauses have been co-ordinated or simply juxtaposed. If the latter is the case, then *mey* could be used as an auxiliary indicating motion rather than a co-ordinator in (240b).

¹³⁸ It is possible that this would normally be expressed through an inclusory pronoun, i.e. *John diweh kinanak nedih* ‘John and his brother’

Hence, there are no dedicated co-ordinators in Kelabit but words and clauses can be co-ordinated using the demonstrative *idih*, the form *kineh teh* and the verb/preposition *mey*. When the subject of the second clause is equivalent to that of the first, the second mention can be omitted.

2.5.4 Summary

In this section, I provided coding and distributional arguments for analysing Kelabit as having grammatical functions. I argued that the argument selected by the verbal morphology (the ASV) is subject and that other nominal arguments are core. Consequently, the function of Kelabit voice is to map different semantic arguments to subject. I subsequently reviewed a set of periphrastic constructions that can be used to map arguments to subject in lieu of morphological voice alternations. These include *inan* clauses for all arguments except actor and undergoer, and *tu'en* clauses for the undergoer.

Finally, I outlined a series of multi-clausal constructions, including relative clauses, clefts, complement clauses, adjunct clauses and co-ordinated clauses. Relative clauses and clefts are subject to the same extraction restriction found in many Western Austronesian languages, in that only the subject can be relativised or clefted (see SUBSECTION 1.4.1). Adjunct clauses may begin with a closed class of conjunctions. Similarly, complement clauses are sometimes introduced with *ngen* or *ken*. In control constructions, the shared argument must be subject in the lower clause. Finally, co-ordination is not marked with a single co-ordinating conjunction but can be achieved using *idih*, *kineh teh*, *mey* and/or particles. Much like in subordination, when arguments are shared between the two co-ordinated clauses, the second mention can be deleted, but only if it fulfils subject function in both co-ordinands. Hence,

multiclausal constructions provide additional support for our analysis of grammatical functions.

2.6 Conclusion

In this chapter, I introduced the Kelabit language of Sarawak, a Western Austronesian language spoken to varying degrees by roughly 6,000 people. It is classified as threatened in Lewis et al (2016) though in fact its vitality differs in the Kelabit Highlands as opposed to the towns. The preliminary grammar sketch in this chapter reveals that Kelabit has many similarities with other Western Austronesian languages (cf. Blust 2013). For example, it has nasal assimilation, a system of voice morphology, morphological causatives, reflexives and reciprocals, and ‘split’ subject properties (Blust 2001). Moreover, Kelabit has many typical characteristics of the languages of Sarawak, including flexible word order, multiple sets of pronouns and periphrastic voice constructions (cf. Clayre 2002, 2014).

In the following chapters, I present the voice system in Kelabit in more detail. In doing so, I explore the differences between Kelabit and other Western Austronesian languages. This allows me to address whether the two-way typology of Philippine-type and Indonesian-type is sufficient to capture syntactic differences in Western Austronesian and what implications this has for wider theoretical and historical debates.

Chapter 3

Voice Alternations

3.1 Introduction

In the previous chapter, I discussed the structure of the Kelabit language from a phonological, morphological and syntactic perspective. I established that there are parallels with other Western Austronesian languages, introduced in CHAPTER 1. In this chapter, I explore some of the differences between Kelabit and other Western Austronesian languages in terms of their voice systems. In doing so, I address the question of where Kelabit fits within Western Austronesian typology, and what it can tell us about ongoing theoretical debates.¹³⁹

In SUBSECTION 1.4.2, I introduced one of the key debates within Austronesian syntax, namely the nature of alignment and whether Western Austronesian languages can be said to have ergative or accusative alignment, or whether they represent a different system of alignment altogether. I also introduced the hypothesis that Western Austronesian languages are in the process of changing from ergative to accusative (Aldridge 2011, 2012, see SUBSECTION 3.4). Kelabit appears to be transitional between the more conservative Philippine-type languages and the more innovative

¹³⁹ An earlier version of this chapter was published as Hemmings (2015). The chapter expands upon voice systems in Western Austronesian, using the methodology presented in SUBSECTION 3.3.

Indonesian-type languages in the area (see SUBSECTION 2.2.1). Hence, it could well reveal evidence of intermediate stages in an alignment transition. Moreover, it offers an ideal opportunity to evaluate variation in Western Austronesian voice systems and the extent to which this is captured by the prevalent two-way typology.

In order to compare voice systems, and establish the place of Kelabit within the typology, we need an independent method of analysing voice. Simply applying the diagnostics of Philippine-type and Indonesian-type languages introduced in CHAPTER 1, presupposes that all Western Austronesian languages fit neatly into one of the two categories. Moreover, it further disassociates Western Austronesian voice from other voice constructions cross-linguistically (see SUBSECTION 3.2). Consequently, this chapter develops a fine-grained approach to the study of Austronesian voice and applies this to Kelabit.

The chapter is structured as follows. SUBSECTION 3.2 defines the concept of voice, drawing on cross-linguistic phenomena. SUBSECTION 3.3 presents an independent methodology for studying voice. SUBSECTION 3.4 reviews the variation in voice systems in Western Austronesian and SUBSECTION 3.5 applies the methodology to Kelabit.

3.2 Voice

In order to establish the best method of comparing voice systems, we begin by defining the term ‘voice’. The category of voice comes from the Ancient Greek tradition of *diathesis*, or formal opposition between *enérgeia* ‘action’ and *páthos* ‘experience’ (Kulikov 2011: 368).¹⁴⁰ The terms were translated into Latin as *activum* and *passivum* and survive in the modern terminology of ‘active’ and ‘passive’ (Kulikov 2011: 368).

¹⁴⁰ Probably the oldest study of voice is the Sanskrit grammar of Pāṇini in circa 500 BC (Klaiman 1991).

However, not all voice systems are of the active/passive type. Indeed, Klaiman (1991: 11) argues that voice has been used in three main ways in the literature. These can be summarised as follows:

- (1) *Conceptions of Voice*
 - a. **Verbal alternations in the syntactic functions of arguments**
Active/Passive
Ergative/Antipassive
Symmetrical Voice
 - b. **Verbal alternations in the semantic properties of arguments**
Active/Middle
 - c. **Verbal alternations in the pragmatic salience of arguments**
Direct/Inverse
Focus Systems
Subject-Object reversal

The different constructions will be illustrated in the following sections and used to build a unified functional definition of voice in SUBSECTION 3.2.4.

3.2.1 Alternations in Syntactic Functions of Arguments

The most canonical use of the term ‘voice’ signals an alternation in the mapping of arguments to grammatical functions (Kulikov 2011). In CHAPTER 1, I discussed two common examples, namely active/passive and ergative/antipassive. The symmetrical voice alternations in Western Austronesian languages also belong in this category, since the verbal morphology serves to indicate which thematic role is mapped to subject (see SUBSECTION 1.4.1 on Tagalog and Indonesian and SUBSECTION 2.5.1 on Kelabit). The only difference is that the alternations are morphologically and syntactically symmetrical. In this section, I review the morphosyntactic properties of each voice system and discuss some additional semantic and discourse correlates.

3.2.1.1 Active/Passive

As seen in CHAPTER 1, the defining morphosyntactic characteristics of passives are as follows:

- (2) **Defining Characteristics of Passives**
 - a. The passive construction is more marked than the active
 - b. The passive is syntactically intransitive/detransitivised
 - c. The undergoer is mapped to subject
 - d. The actor is mapped to an oblique

The construction can be illustrated from Latin in (3):

- (3) *Latin* (Romance)
 - a. **Active**
Miles hostem occidit.
warrior.NOM enemy.ACC kill.PRS.3SG
'The warrior kills the enemy.'
 - b. **Passive**
A milite hostis occidi-tur.
by warrior.ABL enemy.NOM kill.PRS-3SG.PASS
'The enemy is killed by the warrior.' (Kulikov 2011: 370-371)

The active clause in (3a) maps the actor to subject and the undergoer to object. It is transitive and morphologically unmarked for voice. Hence, it can be considered the basic transitive clause (Keenan & Dryer 2006). In contrast, the passive in (3b) maps the undergoer to subject, and the actor to an oblique: *a milite* 'by the warrior'. It is syntactically intransitive, though it expresses an event in which both the actor and the undergoer are inherently involved.

Cross-linguistically, languages differ as to how the oblique status of the passive actor is expressed. Most commonly, this is reflected in the use of oblique case-marking or adpositions, including instrumentals, locatives/ablatives and genitives (Keenan & Dryer 2006). However, in some languages, like Latvian, the

passive actor cannot be expressed at all (Lazdina 1966). In addition, the passive can be indicated via verbal morphology alone. In such cases, the passive actor is realised without an adposition, as shown in Haya in (4):¹⁴¹

- (4) *Haya* (Bantu)
- a. **Passive**
- | | | |
|----------|--------------------|----------|
| Ebitooke | bi-ka-cumb-w' | ómukâzi. |
| banana | they-PST-cook-PASS | woman |
- 'The bananas were cooked by the woman.'
- (Byarushengo et al 1977)

Moreover, the oblique actor can be incorporated into the passive predicate, as in Quechua in (5):

- (5) *Quechua* (Amerindian)
- a. **Active**
- | | | |
|----------|------------|-------------|
| Kuru-Ø | manzana-ta | miku-rqa-n. |
| bug-SUBJ | apple-OBJ | eat-PST-3 |
- 'The bug ate the apple.'
- b. **Passive**
- | | | | |
|------|---------------|------------|-----------|
| Kuru | miku-sqa-mi | manzana-Ø | ka-rqa-n. |
| bug | eat-PART-COMT | apple-SUBJ | be-PST-3 |
- 'The apple was bug eaten.'
- (Keenan & Dryer 2006)

Thus, there is morphosyntactic variation in passive constructions (see Keenan & Dryer 2006 for a more detailed discussion). However, they typically share the function of detransitivisation and demotion of the actor.

As well as being syntactically intransitive, the passive often has semantic and discourse properties associated with low transitivity (SUBSECTION 3.3). Active clauses are generally associated with events in which volitional actors initiate an action that impacts upon a separate participant: the undergoer (see Shibatani 2006). Passive

¹⁴¹ It remains to be seen if the Haya passive could actually be a symmetrical voice alternation.

clauses, in contrast, highlight the affectedness of the undergoer but suggest decreased agency on the part of the actor. This often corresponds to a resultative, stative or perfective interpretation and is used in contexts where the actor is either unknown, self-evident, unimportant or to be avoided for reasons of tact (Shibatani 1985). Thus, passives correspond to low degrees of semantic transitivity, as discussed in SUBSECTION 3.3.2.

In terms of discourse, the passive typically conveys that the undergoer is topical, whilst the actor is not (cf. Givón 1981). It therefore functions to foreground the undergoer and simultaneously background the actor. Hence, the passive indicates that a single argument – the undergoer – is topical, whilst the active is typically used in situations where both actor and undergoer have a degree of discourse prominence. Thus, we can add to our definition of passives the following characteristics:

- (6) **Semantic/Discourse Features of Passives**
- a. Passives are associated with low semantic transitivity
 - b. Passives are associated with low discourse transitivity

3.2.1.2 Ergative/Antipassive

The defining morphosyntactic characteristics of antipassives can be summarised in (7) (cf. Polinsky, to appear):¹⁴²

- (7) **Defining Characteristics of Antipassives**
- a. The antipassive is more marked than the ergative clause
 - b. The antipassive is syntactically intransitive
 - c. The actor is mapped to subject and receives absolutive case
 - d. The undergoer is mapped to an oblique

¹⁴² Antipassives are not restricted to languages with ergative alignment, but also found in ‘accusative’ languages (see Vincent 2013, Polinsky, to appear). Similarly, passives have been identified in ‘ergative’ languages (see Dixon 1994, van de Visser 2006). ‘Ergative’ is used as shorthand in this chapter for transitive clauses in which the actor is marked differently from other core arguments.

An alternation can be seen in Chukchi in (8):

- (8) *Chukchi* (Paleo-Siberian)
- a. **Ergative**
 Aaček-a kimi'-ən ne-nl'etet-Ø-ən.
 youth-ERG load-ABS 3PL.SUBJ-carry.away-AOR-3SG.OBJ
 '(The) young men carried away the load.'
- b. **Antipassive**
 Aaček-ət ine-nl'etet-Ø-g'et kimit'-e.
 youth-ABS ANTIP-carry.away-AOR-3PL load-INS
 '(The) young men carried away a load.' (Kulikov 2011: 381)

The ergative clause in (8a) is transitive and has two nominal arguments: an undergoer, which receives absolutive case, and an actor, which receives ergative case. This is the basic clause-type. In contrast, the antipassive in (8b) is an intransitive construction which maps the actor to subject and demotes the undergoer to an oblique, realised with instrumental case. Nonetheless, an undergoer can be presupposed in an antipassive, even when not overtly expressed (see Polinsky, to appear).

Much like the passive, there is cross-linguistic variation in how the oblique status of the undergoer is expressed. Most commonly, it is indicated through oblique case-marking, agreement and verbal morphology. However, antipassives can also be realised through 'pseudo noun incorporation' (PNI) and 'noun incorporation' (NI) constructions, where indefinite undergoers are incorporated into the predicate and have a fixed position adjacent to the verb (Polinsky, to appear). This can be seen in Tongan, in (9), and an additional Chukchi construction, in (10):

(9) *Tongan* (Polynesian)

a. **Transitive**

‘Oku puke ‘e he pepe ‘a e me’a va’inga
 PRS hold ERG DET baby ABS DET thing playing

mo e pulu lelei.
 COM DET ball good

‘The baby is holding a/the nice toy and ball.’

b. **Antipassive (PNI)**

‘Oku puke (*‘a) e me’a va’inga mo e pulu lelei
 PRS hold ABS DET thing playing COM DET ball good

‘a e pepe.
 ABS DET baby

‘The baby is holding a nice toy and ball.’ (Polinsky, to appear)

(10) *Chukchi* (Paleo-Siberian)

a. **Ergative**

ʔətt-e melotaly-ən piri-nin.
 dog-ERG hare-ABS catch-AOR.3SG:3SG

‘The dog caught a/the hare.’

b. **Antipassive (NI)**

ʔətt-ən milute-piri-ʔʔi.
 dog-ABS hare-catch-AOR.3SG

‘The dog caught a/the hare.’ (Polinsky, to appear)

PNI differs from NI in that the incorporated element can be bigger than a single noun, as long as it is not case-marked. For example, in the Tongan example in (9b) the entire phrase *e me’a va’inga mo e pulu lelei* ‘the nice toy and ball’ is incorporated but cannot be overtly case-marked with absolutive case. The NI case in Chukchi, in contrast, incorporates only the head noun *milute* ‘hare’. Both (9b) and (10b) are considered syntactically intransitive, as reflected in the word order and agreement, and therefore represent the same function as the morphologically marked antipassive (Polinsky, to appear).

Like the passive, the antipassive is associated with particular discourse functions and semantic interpretations. In terms of semantics, the antipassive is often associated with atelicity and imperfective aspects, such as the progressive, durative, inceptive, inchoative and iterative (see Polinsky, to appear, Cooreman 1994, Dixon 1994, Spreng 2010). Indeed, antipassive morphology is sometimes reanalysed as marking aspect rather than detransitivisation (cf. Comrie et al 2015). Secondly, Cooreman (1994: 51) suggests that antipassives may indicate that an event has not been successfully completed, or that the undergoer is only partially affected by the act. Finally, the antipassive is often used in cases where the undergoer is low in identifiability. This means that the undergoer is typically indefinite and non-referential. Hence, the antipassive is associated with a lower degree of semantic transitivity (see SUBSECTION 3.3.2).

In terms of discourse, antipassives are used in situations where the actor is foregrounded and the undergoer is backgrounded (see Foley & Van Valin 1984, Polinsky, to appear). Oblique case-marking typically indicates that the undergoer is obvious, generic or unimportant in discourse and will not remain under discussion in subsequent conversation. Hence, the antipassive is associated with a situation in which only the actor is topical, whilst active/ergative clauses tend to have a topical actor and undergoer (Cooreman, Fox & Givón 1984). Thus, we can add the following characteristics to the definition of antipassives:

- (11) **Semantic/Discourse Features of Antipassives**
- a. Antipassives are associated with low semantic transitivity
 - b. Antipassives are associated with low discourse transitivity

3.2.1.3 Western Austronesian Symmetrical Voice

Finally, the defining characteristics of Western Austronesian voice systems are summarised in (12), following SUBSECTION 1.3:

- (12) **Defining Characteristics of Western Austronesian Voice**
- a. The voices are equally transitive, containing two or more core arguments
 - b. In AV, the actor is mapped to subject and the undergoer is core
 - c. In UV, the undergoer is mapped to subject and the actor is core
 - d. In Philippine-type systems, peripheral arguments are mapped to subject in their respective voices.

An Indonesian-type system is illustrated for Javanese in (13) and a Philippine-type system for Cebuano in (14):

- (13) *Javanese* (Indonesian-type)
- a. **Actor Voice**
Kucing mangan iwak.
cat AV.eat fish
'The cat ate fish.'
 - b. **Undergoer Voice**
Iwak di-pangan kucing.
fish UV-eat cat
'The cat ate the fish.'
- (Hemmings 2012: 68)

- (14) *Cebuano* (Philippine-type)
- a. **Actor Voice**
Ni-hatag si Juan sa libro sa bata.
AV-give PT John PT book PT child
'John gave the book to the child.'
 - b. **Undergoer Voice**
Gi-hatag ni Juan ang libro sa bata.
UV-give PT John PT book PT child
'John gave the book to the the child.'

c. **Dative/Locative Voice**

Gi-hatag-an ang bata ni Juan sa libro.
DV-give-DV PT child PT John PT book
'John gave the child the book.'

d. **Instrumental Voice**

I-hiwa ang kutsilyo sa manga ni Maria.
IV-cut PT knife PT mango PT Maria
'Maria cut the mango with a knife.' (Shibatani 1988: 88-89)

The systems illustrated in (13) and (14) differ in particular structural properties, as discussed in SUBSECTION 1.3.1. However, they both indicate alternations in the mapping of arguments to functions. Unlike passives and antipassives, AV, UV and peripheral voices are all transitive, with two or more core arguments. Moreover, each voice is equally morphologically marked. Hence, the Austronesian alternations in (13) and (14) are analysed as 'symmetrical voice' rather than treating AV as an antipassive using PNI - like Tongan in (9) – or UV as a passive without oblique case-marking – like Haya in (4) (see Riesberg 2014).

There are several areas of morphosyntactic variation in Western Austronesian voice systems that I return to in SUBSECTION 3.4. In addition, a great wealth of studies suggest that the voice constructions have different semantic and discourse statuses (Kroeger 2004, Cooreman, Fox & Givón 1984, Gault 1999, Nolasco 2005, Norwood 2002, Donohue 2002 among others). In Philippine-type languages, UV is typically associated with high semantic and discourse transitivity, whilst AV is associated with low transitivity. In contrast, AV in some Indonesian-type languages has properties of high transitivity and UV has properties of low transitivity (SUBSECTION 3.4.1, 3.4.2). These findings are central to the alignment shift hypothesis, and are discussed in more detail in SUBSECTION 3.4. For now, I conclude that voice alternations are also

associated with particular semantic and discourse correlates in Western Austronesian, and that these seem to vary in the following ways:

- (15) **Semantic and Discourse Properties of Western Austronesian Voice**
- a. In Philippine-type languages, UV is associated with high discourse and semantic transitivity, and AV with low transitivity.
 - b. In (some) Indonesian-type languages, AV is associated with high discourse and semantic transitivity, and UV with low transitivity.

3.2.2 Alternations in Semantic Properties of Arguments

A second conception of voice is used to describe alternations in which the mapping of arguments to functions remains constant, but the semantic properties of arguments and events change. This conception of voice is most clearly represented by the active/middle alternation.

3.2.2.1 Active/Middle

The notion of middle voice goes back to the work of traditional grammarians (see Klaiman 1991). In contrast to the alternations in SUBSECTION 3.2.1, the middle is not an alternation in the mapping of arguments to functions. Instead, it is an alternation in semantic transitivity (see SUBSECTION 3.3.2). Whilst the active voice typically implies a situation in which a volitional actor acts upon a distinct undergoer, the middle voice implies that the subject has properties of both the actor and undergoer simultaneously. In other words, the subject is both the cause of the event, and the entity that is most directly affected by the action (cf. Lyons 1968). Hence, the middle voice is often seen as a midpoint between active and passive and has been variously analysed as a marker of lower transitivity, valency alternation, as having a relationship with reflexives, and

as the basis of organisation in the lexicon (Kuryłowicz 1964, Barber 1975, Klaiman 1991).

The active/middle alternation can be illustrated from Sanskrit:

- (16) *Sanskrit* (Indo-Iranian)
- a. **Active**
 Devadattaḥ kaṭaṃ karoti.
 Devadatta.NOM mat.ACC make.3SG.ACT
 ‘Devadatta makes a mat.’
- b. **Middle**
 Devadattaḥ kaṭaṃ kurute.
 Devadatta-NOM mat.ACC make.3SG.MIDDLE
 ‘Devadatta makes himself a mat.’ (Klaiman 1991: 24)

In both (16a) and (16b), the actor ‘Devadatta’ is mapped to subject and the undergoer ‘mat’ to object. This can be seen from coding and behavioural properties, such as case-marking and agreement on the verb. Therefore, there is no alternation in the mapping of arguments to functions. Nonetheless, there is formal alternation in the verbal inflection that corresponds to a different semantic interpretation. Unlike the active in (16a), the middle in (16b) indicates that the action has an effect on the subject.

Common middle situations, following Kemmer (1994) are summarised in

TABLE 3.1:

Table 3.1 Middle Situations (Kemmer 1994)

Situation	Example
Grooming or body care	wash, shave
Nontranslational motion	stretch, turn, bow
Change in body posture	sit down, kneel down
Translational motion	climb up, go away
Naturally reciprocal events	embrace, wrestle
Indirect middle	acquire, ask, take, desire
Emotion middle	be angry, grieve
Emotive speech actions	complain, lament
Cognition middle	reflect, ponder
Spontaneous events	germinate, sprout

In Kemmer's (1994) terms, these events share the fact that the actor is both 'initiator' and 'endpoint' of an event, and that the event is 'low in elaboration', without clearly distinguishable participants or sub-events. In addition, middle voice is often used to refer to facilitative constructions like 'the book sells well' or 'the book reads well', in which an actor is understood to exist but is pragmatically less important than the undergoer (Kemmer 1994: 147). These uses are similar to those in TABLE 3.1 in that the focus is on the affected entity and that the event is also low in elaboration. As a result, the middle often correlates with features of low semantic transitivity, such as irrealis mood and non-punctual aspects (Klaiman 1991).

Thus, the following characteristics identify middle voice:

- (17) **Defining Characteristics of Middle Voice**
- a. Active and middle are both syntactically transitive constructions
 - b. There is no alternation in grammatical functions
 - c. The middle is associated with lower semantic transitivity

3.2.3 Alternations in Pragmatic Salience of Arguments

Finally, the term 'voice' has sometimes also been applied to systems in which verbal alternations signal the relative pragmatic prominence of arguments, either in terms of their relative ontological status or relative informational status (Klaiman 1991). These include inverse systems, focus systems and subject-object reversal. None of these systems are uncontroversially identified as voice in the literature. However, there are several parallels with Western Austronesian. In particular, they do not involve syntactic detransitivisation.

3.2.3.1 Inverse Systems

Inverse systems are found in a range of languages, including Algonquian, Wakashan, Chukotko-Kamchatkan and Tupí Guaraní (cf. Klaiman 1991, D. Payne 1994, Whistler 1985, Comrie 1980). In inverse systems, verbal morphology indicates whether the actor outranks the undergoer or vice versa. For many Algonquian languages, the person-referencing hierarchy that represents the ontological salience of participants is as follows (Klaiman 1991: 191):

- (18) **Person-Referencing Hierachy in Algonquian**
2 > 1 > 3 proximate > 3 obviative¹⁴³

When a person higher on the hierarchy acts on a person lower on the hierarchy, the direct suffix is used. When a person lower on the hierarchy acts on a person higher on the hierarchy, the inverse suffix is used. These suffixes are known as ‘theme signs’ in the Algonquian literature (cf. Macaulay 2009). The alternation can be illustrated for Plains Cree:

- (19) *Plains Cree* (Algonquian)
- a. **Direct**
Ni-sēkih-ā-nān atim.
1-scare-DIR-1PL dog
‘We scare the dog.’
- b. **Inverse**
Ni-sēkih-iko-nān atim.
1-scare-INV-1PL dog
‘The dog scares us.’
- (Wolfart 1973: 25)

In (19a), the first person plural is acting on a third person participant, the dog. Hence, the direct theme sign *-a* is used. In (19b), conversely, the third person dog is acting on

¹⁴³ Obviation distinguishes between proximate third persons, who are salient in discourse, and obviative third persons who are not. See Macaulay (2009) for discussion of variation in prominence hierarchies.

the first plural participant. Hence, the inverse theme sign *-iko* is used to indicate that the actor is lower on the person-referencing hierarchy than the undergoer.

Since Jones (1911), some Algonquianists have analysed constructions like (19b) as passives in which the salient undergoer is mapped to subject, rather than the less salient actor (see LeSourd 1976 on Meskwaki, Rhodes 1994 on Ojibwa). This analysis is based on two arguments: verbal agreement and functional similarities between the inverse and the passive. To illustrate the morphological argument, consider the following data from Meskwaki:

- (20) *Meskwaki Agreement*
- a. **Animate Intransitive**
Ke-we·wenesi.
2-be.pretty
'You are pretty.'

 - b. **Direct**
Ke-pemen-a·-w-a.
2-take.care.of-DIR-3-SG
'You take care of him.'

 - c. **Inverse**
Ke-pemen-ekw-w-a.
2-take.care.of-INV-3-SG
'He takes care of you.'
- (Dahlstrom, nd)

In (20), the prefix *ke-* indicates agreement with a second person argument. In (20b) the second person is the actor and the theme sign is direct, whilst in (20c) the second person is the undergoer, and the theme sign is inverse. If theme signs are taken to indicate an alternation in the mapping of arguments to functions, then we could simply

state that the prefix agrees with the subject. This would arguably give a simpler account of the morphology.¹⁴⁴

Moreover, the inverse is in many ways functionally similar to a passive. For example, in the context of two third person arguments, Algonquian languages like Meskwaki obligatorily use a system of obviation to mark one of the third person arguments as proximate, or central to the discourse, and one as obviative, or less central to the discourse. If the proximate acts upon the obviative then the direct construction is used. If the obviative acts upon the proximate, the inverse construction is used. Hence, the inverse indicates that the undergoer is more discourse topical than the actor, much like the passive (see SUBSECTION 3.2.1.1).

However, the inverse is distinct from the passive in that it represents a transitive construction with two nominal arguments. This can be seen in the contrast between intransitive predicates, such as (20a), which take a single agreement prefix, and direct/inverse constructions in (20b) and (20c), which agree with two arguments. For this reason, Perlmutter & Rhodes (1988) suggest the term ‘reversal’ and analyse Algonquian inverse systems in a similar manner to Western Austronesian. Under such

¹⁴⁴ Nb. agreement in Meskwaki is somewhat more complicated than presented above. For example, there are cases in which direct and inverse constructions do not have the same marking, as in (i) and (ii):

- (i) **Direct**
Ne-pemen-a.-pena.
1-take.care.of-DIR-1PL
‘We (EXCL) take care of him/them.’
- (ii) **Inverse**
Ne-pemen-ekw-na.n-a.
1-take.care.of-INV-1PL-3SG
‘He takes care of us (EXCL).’ (Dahlstrom, nd)

Dahlstrom (nd) takes this as evidence against the morphological argument for a symmetrical voice analysis. Note that non-subject actors are sometimes marked differently from non-subject undergoers in Western Austronesian. These alternations are nonetheless considered symmetrical. More interesting are the syntactic arguments against a symmetrical voice analysis in Meskwaki (Dahlstrom, nd). Evidence from possessor raising suggests that the undergoer is the object of an inverse clause, and the actor the subject. In this way, Meskwaki crucially differs from Western Austronesian.

an analysis, the direct construction is analysed as AV-like and the inverse construction as UV-like.

Rhodes (1994) supports this analysis by demonstrating that the undergoer has subject properties in the inverse construction. This can be seen in Ojibwa from the ‘copying to object’ construction, in which the subject of a complement clause is copied as the object of the matrix clause (cf. Dahlstrom, nd). Importantly, non-subject arguments cannot be copied. For example, in Ojibwa the actor of a direct construction can be copied to object, whilst an undergoer cannot:

(21) *Copying to Object in Ojibwa (Direct)*

a. **Actor copied to object**

Ngikenmaag	ninwag	gii-baashkzwaawaad
ni-gikenim-aa-ag	aniniw-ag	gii-baashkizw-aa-waa-d
1-know-3.ANIM.OBJ-3P	man-PL	PST-shoot-3.ANIM.OBJ-3P-3SUBJ

Maagiiyan.
 Maagii-an
 Marge-OBV
 ‘I know that the men (prox) shot Marge (obv).’
 (know agrees with ‘men’)

b. **Undergoer copied to object**

*Ngikenmaa	Maagiiyan	gii-baashkzwaawaad
ni-gikenim-aa	Maagii-an	gii-baashkizw-aa-waa-d
1-know-3.ANIM.OBJ	Marge-OBV	PST-shoot-3.ANIM.OBJ-3P-3.SUBJ

ninwag.
 aniniw-ag
 man-PL
 For: ‘I know that the men (prox) shot Marge (obv).’
 (know agrees with ‘Marge’) (Rhodes 1994: 439)

In contrast, in the inverse construction, the undergoer can be copied to object and not the actor:

(22) *Copying to Object in Ojibwa (Inverse)*

a. **Undergoer copied to object**

Ngikenmaa Maagii gii-baashkzogod ninwan.
ni-gikenim-aa Maagii gii-baashkizw-igo-d aniniw-an
1-know-3.ANIM.OBJ Marge PST-shoot-INV-3SUBJ man-OBV
'I know that the men (obv) shot Marge (prox).'

(know agrees with 'Marge')

b. **Actor copied to object**

*Ngikenmaag ninwan gii-baashkzogod Maagii.
ni-gikenim-aa-ag aniniw-an gii-baashkizw-igo-d Maagii
1-know-3.ANIM.OBJ-3P man-OBV PST-shoot-INV-3.SUBJ Marge
For: 'I know that the men (obv) shot Marge (prox).'

(know agrees with 'men')

(Rhodes 1994: 439-440)

Hence, the Ojibwa inverse is similar to Austronesian UV constructions, in that the inverse undergoer has subject properties.¹⁴⁵

However, Dahlstrom (1991) demonstrates that this is not true of all Algonquian languages. In Plains Cree, for example, a different pattern is found in the 'copying to object' construction. Only the actor can be copied, regardless of whether the construction is direct or inverse:

(23) *Copying to Object in Plains Cree (Direct)*

a. **Actor is copied to object**

Nikiske·yima·w George e·=sa·kiha·t okosisa.
know.1-3.INDP.IND George love.3-3'.CONJ his.son.OBV
'I know George (prox) loves his (prox) sons (obv).'

b. **Undergoer is copied to subject**

*Nikiskeyimima·wa George e·=sa·kiha·t okosisa.
know.1-3'.INDP.IND George love-3-3'.CONJ his.son.OBV
For: 'I know George loves his sons.'

(Dahlstrom 1991: 72-73)

¹⁴⁵ See Fry & Hamilton (2014) for similar results relating to Mi'gmaq.

- (24) *Copying to Object in Plains Cree (Inverse)*
- a. **Actor is copied to object**
 Nikiske·yimima·wa George e·=sa·kihikot okosisa.
 know.1-3'/INDP.IND George love.3'-3'/CONJ his.son.OBV
 'I know George (prox) loves his (prox) sons (obv).'
- b. **Undergoer is copied to subject**
 *Nikiske·yima·w George e·=sa·kiha·t okosisa.
 know.1-3'/INDP.IND George love-3'-3'/CONJ his.son.OBV
 For: 'I know George loves his sons.' (Dahlstrom 1991: 73)

Hence, the mapping of arguments to functions does not appear to have changed in Plains Cree. Unlike in Ojibwa, the actor has subject properties in both the direct and the inverse.¹⁴⁶ What this shows is that some inverse systems may be analysable as 'symmetrical voice', but others do not involve the remapping of arguments to functions.

Consequently, the main characteristics of inverse systems are as follows:

- (25) **Defining characteristics of the inverse**
- Both direct and inverse are syntactically transitive & morphologically marked
 - The direct is used when the actor is more prominent than the undergoer
 - The inverse is used when the undergoer is more prominent than the actor
 - An alternation in the mapping of grammatical functions is not necessary

3.2.3.2 Focus Systems

Focus systems signal changes in the information structural status of arguments. They are common in ergative Mayan languages and usually occur in addition to antipassive and passive constructions (Grinevald & Peake 2012). Like inverse systems, focus

¹⁴⁶ Dahlstrom (1991) provides similar evidence in terms of quantifier floating. These are properties that Schachter (1976) identifies as reference-related, and Manning (1996) uses to identify subject as opposed to actor. Hence, the patterns are in contrast to those of Western Austronesian.

systems do not affect the syntactic transitivity of clauses or the linking of arguments to functions. Instead, they identify a particular argument as having information structure salience. This can be illustrated from Chajul Ixil:

- (26) *Chajul Ixil* (Mayan, Mamean)
- a. **Ergative**¹⁴⁷
 Kat in-q'os axh.
 ASP 1SG.ERG-hit 2SG.ABS
 'I hit you.'
- b. **Actor Focus**
 In kat q'os-on axh.
 1SG.ABS ASP hit-AF 2SG.ABS
 'It was I who hit you.'
- c. **Ergative**
 A-k'oni in ta'n uula.
 2SG.ERG-shoot 1SG.ABS with sling
 'You shot me with a sling.'
- d. **Instrument Focus**
 Uula a-k'oni-b'e in.
 sling 2SG.ERG-shoot-IF 1SG.ABS
 'With a sling you shot me.' (Klaiman 1991: 34, 244)

In Chajul Ixil, the basic ergative clauses in (26a) and (26c) are transitive and take two nominal arguments: an ergative actor, signalled via agreement prefixes on the verb, and an absolutive undergoer. The neutral word order is verb-initial and the predicate is unmarked for voice. In (26b) and (26d) we see two 'focus' constructions. In (26b), the predicate is marked with the suffix *-on* and the actor appears pre-verbally. In (26d), the predicate is marked with the suffix *-b'e* and the instrument appears pre-verbally. It is realised as a core nominal argument, unlike in (26c), where the instrument is non-prominent and realised as a PP. These serve to place particular information structure

¹⁴⁷ Ayres (1983) suggests that the undergoer cannot be given marked informational salience in the same manner. Perhaps this is because it has salience in the basic ergative construction.

salience on the actor and instrument, indicating contrastive, new or emphasised interpretations (Klaiman 1991, Ayres 1983).

In many Mayan languages, focus constructions are necessary in the context of pragmatic focus, wh-question and relative clause constructions (see Aissen 1999, Stiebels 2006).¹⁴⁸ This can be illustrated with Q'eqchi in (27) and Quiché in (28):

- (27) *Q'eqchi* (Mayan, Quichean)
- a. **Ergative (relativising on undergoer)**
X-x-cam li ixk [li x-r-il li cuink].
REC.PST-3.ERG-die the woman that REC.PST-3.ERG-see the man
'The woman [that the man saw] died.'
- b. **Actor Focus (relativising on actor)**
X-x-cam li ixk [li x-il-o-c
REC.PST-3.ERG-die the woman that REC.PST-see-AF-NON.FUT.INTR

r-e li cuink].
3.ERG.DAT the man
'The woman [that saw the man] died.' (Berinstein 1985: 167)
- (28) *Quiché* (Mayan, Quichean)
- a. **Instrument Focus (relativising on the instrument)**
X-Ø-inw-elaq'a-j lee ch'iich'
ASP-3SG.ABS-1SG.ERG-steal-SUFFIX the machete

[x-Ø-u-rami-b'e-j lee achih r-ee lee chee7].
ASP-3SG.ABS-3SG.ERG-cut-IF-SUFFIX the man 3SG-GEN the tree
'I stole the machete that the man used to cut the tree.'
(Norman 1978: 463)

Only the undergoer can be relativised from an ergative clause, as in (27a). In order to relativise the actor, an actor focus construction is used, as in (27b). To relativise on the instrument, an instrument focus construction is used, as in (28). Hence, there are

¹⁴⁸ See Stiebels (2006) for discussion of how Mayan languages differ in this respect.

functional and syntactic similarities between Mayan focus constructions and the Western Austronesian voice alternations in SUBSECTION 1.4.1.

Actor focus (AF) constructions, such as (26b) and (27b), have sometimes been analysed as antipassives (Berinstein 1985, Pinkerton 1978).¹⁴⁹ In Q'eqchi, this is supported by the fact that AF is morphologically and syntactically intransitive. The predicate does not agree with the ergative argument, unlike (27a), and takes the TAM marker *-c*, which only co-occurs with intransitive predicates (Stiebels 2006). Moreover, the undergoer is an oblique and marked with the dative case. However, in Ixil and other Mayan languages, although AF verbs only agree with the absolutive argument, there is evidence to suggest that they are syntactically transitive. Firstly, actor and undergoer are both realised as core nominals without prepositions (cf. Aissen 1999, Ayres 1983, Smith-Stark 1978). Secondly, AF clauses in Ixil do not take intransitive verbal morphology, such as the *-i* suffix that is used in the context of tense/aspect markers like *kat* (Ayres 1983). Furthermore, focus constructions have a different discourse function to passives and antipassives. Focus constructions serve to indicate the discourse prominence of the preverbal constituent, whilst passives and antipassives serve to defocus the prominence of actors and undergoers respectively (Ayres 1983). Consequently, in languages like Ixil, AF constructions are more commonly analysed as 'focus' (Klaiman 1991), 'verbal indexing' (Ayres 1983) or inverse constructions (Aissen 1999, Zavala 1997).¹⁵⁰

¹⁴⁹ Similarly, Norman (1978) analyses *-b'e* as an applicative, indicating the promotion of the instrument to direct object, and the demotion of the undergoer to an oblique. There are many syntactic tests that support this observation, including the use of oblique genitive marking in (28a) and patterns of passivisation (see Norman 1978: 462). Norman (1978) argues that the applicative function is the historically original function and that it develops into a focus marker in languages like Ixil, as discussed below.

¹⁵⁰ Arguments for an inverse analysis are largely functional. AF constructions tend to occur in Mayan languages where the undergoer is more prominent than the actor (see Aissen 1999, Duncan 2003).

The reason that the Ixil constructions are treated as ‘focus’ rather than symmetrical voice is that, like Plains Cree, the verbal alternations do not signal remapping of arguments to functions. Two main arguments support this analysis: coding patterns and control of agreement (Ayres 1977, 1983, Norman 1978). In Ixil, core arguments are expressed as nominals and obliques as prepositional phrases. In both (26c) and the IF construction in (26d), the actor is expressed through an ergative agreement prefix and the undergoer using the absolutive pronoun. Hence, coding suggests that no alternation in grammatical functions has taken place (see Ayres 1983). Moreover, a property of objects in Mayan languages is that they control agreement on the predicate. In the IF construction, it is the undergoer that controls absolutive agreement, as opposed to the instrument. This can be seen in (29):

- (29) *Ixil Agreement Patterns*
 a. **Instrument Focus**
 Axh la7 in-paxi-b’e-Ø u ispeeja.
 2SG.ABS ASP 1SG.ERG-break-IF-3SG.ABS the window
 ‘I’m going to use you to break the window.’ (Norman 1978: 465)

The predicate in (29) takes a zero-morpheme, indicating agreement with the 3SG undergoer, ‘the window’, as opposed to the 2SG instrument. This suggests that the undergoer retains its core status in IF and that no alternation in grammatical functions has taken place.¹⁵¹

Consequently, focus systems in languages like Ixil reflect changes in the relative discourse prominence of arguments without affecting either their mapping to

¹⁵¹ Note that the fact that AF and IF are associated with relativisation, wh-questions and focus constructions in Ixil suggests that there could be an alternation in subject along the same lines as in Western Austronesian. Exploring this is beyond the scope of the current dissertation but could be a fruitful area for future research.

grammatical functions or the syntactic transitivity. Thus, the defining characteristics are as follows:

- (30) **Defining Characteristics of Focus Systems**
- a. Focus systems do not trigger alternations in grammatical functions
 - b. Focus constructions are equally transitive, though may differ morphologically
 - c. The focus marker indicates particular information structure prominence for that argument
 - d. Actor focus is typically associated with contexts in which the undergoer is more prominent than the actor.

3.2.3.3 Subject-Object Reversal

Finally, the subject-object reversal construction in Bantu also seems to reflect differences in information structure (cf. Marten & van der Wal 2014, Zerbian 2006).¹⁵²

Bantu languages are generally SVO with nominative/accusative alignment. In basic transitive clauses, the actor is mapped to subject and triggers agreement on the verb. The undergoer is mapped to object and appears in the immediately post-verbal position (Zerbian 2006: 361). However, in subject-object reversal constructions the undergoer is realised in the pre-verbal position and triggers agreement on the verb. The actor occurs post-verbally and is necessarily interpreted as focussed (Marten & van der Wal 2014, Ndayiragije 1999). This can be seen in Kirundi in (31):

- (31) *Kirundi* (Bantu)
- a. **Active**

Petero	a-á-ra-guze	ibitabo.
Peter	3SG-PST-ANTIF-buy.PFV ¹⁵³	books

‘Peter bought books.’

¹⁵² This could be considered part of a larger system of inversion in which a number of semantic arguments can appear in pre-verbal position and trigger agreement on the verb (see Marten & van der Wal 2014). Hence, it is comparable to Philippine-type voice systems.

¹⁵³ See below for discussion of the antifocus marker *ra-*. This is independent of subject-object reversal.

b. **Subject-Object-Reversal**

Ibitabo	bi-á-guze	Petero.
Books	3PL-PST-buy.PFV	Peter
‘Peter (not John) bought books.’		(Ndayiragije 1999: 412)

Both clauses in (31) are transitive, with two core nominal arguments *Petero* ‘Peter’ and *ibitabo* ‘books’. However, in (31a) the actor is pre-verbal and triggers an *a*-agreement prefix, whilst in (31b) the undergoer is pre-verbal and triggers a *bi*-agreement prefix. As discussed in Marten & van der Wal (2014), clauses like (31a) are most likely to occur in contexts where the actor is the information structure topic. In contrast, clauses like (31b) occur in contexts where the undergoer is the topic and the actor represents focus information. In (31b), this results in a contrastive interpretation.¹⁵⁴ There are some animacy restrictions, as subject-object reversal is only possible when the actor is more animate than the undergoer (Morimoto 2003). Similar constructions are found in Swahili (Whiteley 1972, Whiteley & Mganga 1969), Dzamba (Givón 1979), Luguru (Mkude 1974, Marten & van der Wal 2014) and Kilega (Kinyalolo 1991), among others.

Since the undergoer controls agreement in subject-object reversal, such constructions have been treated as non-canonical passives (Kimenyi 1980, Hamlaoui & Makasso 2013). However, they differ from passives in a number of ways, as seen if we compare (31b) with a canonical passive in Kirundi:

(32) *Kirundi*

a. **Passive**

Ivyo	bitabo	bi-á-ra-guz-u-e	(na	Petero).
Those	books	3PL-PST-ANTIF-buy-PASS-PFV	(by	Peter)
‘Those books were bought (by Peter).’			(Ndayiragije 1999: 412)	

¹⁵⁴ See Marten & van der Wal (2014) and Ndayiragije (1999) for further discussion of information structure correlates. In Swahili, subject-object reversal has been argued to represent an event as particularly unexpected or noteworthy (cf. Whiteley & Mganga 1969).

In (32), the undergoer also appears pre-verbally and triggers agreement on the verb. However, it takes additional morphological marking in the form of the passive suffix *-u*. Similarly, the actor is demoted and optionally expressed through an oblique PP. Hence, the clause is syntactically intransitive. The subject-object reversal in (31b) is morphologically unmarked and the post-verbal actor remains a core argument. It cannot be deleted and is realised as a core nominal argument, rather than an oblique PP (Marten & van der Val 2014, Ndayiragije 1999). Finally, the passive is compatible with the antifocus marker *ra-*, which indicates a discourse-neutral context without focus, whilst subject-object reversal is not (Ndayiragije 1999). Given these differences, subject-object reversal has sometimes been analysed as a sort of symmetrical voice alternation in which the undergoer is mapped to subject and the actor to object, though one that is not indicated via overt verbal morphology (Morimoto 2006).

This is supported by the fact that the pre-verbal undergoer in constructions like (31b) has a number of apparent subject properties. In addition to controlling agreement, it can undergo right-dislocation, like the subject in a basic transitive clause:

(33) *Kirundi Right Dislocation*

a. **Active**

Abâna	ba-á-ra-somye	igitabo.
children	3PL-PST-ANTIF-read.PFV	book

‘Children read a book.’

b. Ba-á-ra-somye igitabo, abo bâna.
 3PL-PST-ANTIF-read.PFV book DEM children
 ‘They read a book, those children.’

c. **Subject-Object-Reversal**

Igitabo	ki-á-somye	abâna.
Book	3SG-PST-read.PFV	children

‘Children (not parents) read a book.’

- d. *Ki-á-somye* *abâna,* *ico* *igitabo.*
 3SG-PST-read.PFV children DEM book
 ‘That book, children (not parents) read it.’ (Ndayiragije 1999: 422)

However, Morimoto (2006) and Zerbian (2006) provide a number of arguments against treating subject-object reversal as symmetrical voice. Firstly, as Kimenyi (1980: 145) notes, apart from control of agreement, which arguably facilitates right dislocation, there are very few syntactic tests that readily identify the pre-verbal undergoer as subject. For example, control of raising and equi NP deletion cannot be used, since Kirundi has no equivalent raising verbs and there can never be gaps in coordinate clauses since obligatory verbal agreement prefixes function as anaphoric topic pronouns (Morimoto 2006: 168).

Furthermore, the post-verbal actor in subject-object reversal constructions does not have typical object properties. Firstly, it cannot be passivised or expressed using an object marker (Morimoto 2006: 169). Secondly, it cannot be relativised, unlike objects in basic transitive clauses:

(34) *Kirundi Object Relativisation*

a. **Transitive**

Igitabo [*abâna* *ba-á-riko* *ba-soma*]...
 Book children 3PL-PST-be 3PL-read.IPFV
 ‘The book that the children were reading...’

(Ndayiragije 1999: 420)

b. **Subject-Object-Reversal**

**Umuntu* [*ivyo* *bitabo* *bi-á-somye*]...
 Person those books 3PL-PST-read.PFV
 For: ‘The person who read those books...’

(Ndayiragije 1999: 428)

Thirdly, the post-verbal actor doesn't appear immediately after the verb, like typical objects, but rather clause-finally, as demonstrated by the position of adverbs:

- (35) *Kirundi Adverb Order*
- a. **Active**
 Yohani a-á-ra-oógeje imiduga néezá.
 John 3SG-PST-ANTIF-wash.PFV cars well
 ‘John washed cars well.’
- b. *Yohani a-á-ra-oógeje néezá imiduga.¹⁵⁵
 John 3SG-PST-ANTIF-wash.PFV well cars
 ‘John washed cars well (not trucks).’ (Ndayiragije 1999: 416)
- c. **Subject-Object-Reversal**
 Imiduga yi-á-oógeje néezá Yohani.
 cars 3PL-PST-wash.PFV well John
 ‘John (not Peter) washed cars well.’
- d. *Imiduga yi-á-oógeje Yohani néezá.
 cars 3PL-PST-wash.PFV John well
 For: ‘John washed the cars well.’ (Ndayiragije 1999: 417)

In pragmatically neutral contexts, indicated through the antifocus *-ra-* affix in (35a-b), the object occurs directly after the verb. However, in the subject-object reversal construction, the actor must occur clause-finally and cannot occur before the adverb. This reinforces the idea that it is focus but not object.

Finally, the post-verbal actor retains some subject properties. For example, the actor controls the interpretation of subject in a control construction, regardless of whether the clause is active or subject-object reversal:

- (36) *Kirundi Control Constructions*
- a. **Active**
 Yohani_i a-á-ra-emeye [pro_i kugura iyo modoka].
 John 3SG-PST-ANTIF-accept.PFV INF.buy that car
 ‘John agreed to buy that car.’ (Ndayiragije 1999: 417)

¹⁵⁵ Note that when the antifocus affix *ra-* is not used, both orders are possible with different interpretations. When the object is final it is focused, i.e. washed cars and not trucks. When the adverb is final it is focused, i.e. washed well and not badly (Ndayiragije 1999).

b. **Subject-Object-Reversal**

Iyo modoka_i i-á-emeye [pro_i kugura] Yohani.
that car 3SG-PST-accept.PFV INF.buy John
'John (not Peter) agreed to buy that car.' (Ndayiragije 1999: 417)

In both (36a) and (36b), it is 'John' that controls the subject in the subordinate clause. Hence, Morimoto (2006) concludes that the actor retains the grammatical function of subject in the subject-object reversal construction. Given these effects, Morimoto (2000, 2006) argues that the agreement is actually topic agreement, rather than subject agreement, which could explain the patterns of right dislocation. Hence, no alternation in the mapping of arguments to functions occurs and the defining characteristics of Bantu Subject-Object reversal are:

(37) **Defining Characteristics of Subject-Object Reversal**

- a. The subject-object reversal is syntactically transitive
- b. There is no additional morphological marking of voice
- c. The undergoer appears in pre-verbal position and triggers agreement and the actor occurs in post-verbal position
- d. There is little evidence for an alternation in grammatical functions, but rather the construction indicates that the actor is focus and the undergoer is topic.

3.2.4 Summary

In the previous sections, I surveyed a selection of putative voice systems in the world's languages. This revealed a high degree of cross-linguistic variation in what has been treated as 'voice'. Firstly, not all voice alternations involve syntactic detransitivisation. Secondly, not all voice systems involve the remapping of arguments to syntactic functions, though this is often the canonical understanding of voice (cf.

Kulikov 2011: 371).¹⁵⁶ In this section, I summarise the features that these constructions share in common in order to arrive at a functional definition of voice.

Firstly, the alternations are reflected in the verbal morphology or through other morphosyntactic means (Kulikov 2011: 371). Secondly, the voice systems involve a proto-typical transitive construction, with particular syntactic, semantic and discourse properties, as well as alternations, in which participants acquire non-default syntactic functions, semantic entailments or discourse statuses. Hence, voice can be understood in its widest sense as a grammatical category that combines morphology, syntax, semantics and discourse to indicate changes in the participant's relation to the presentation of an event. In other words, voice can be defined as follows, in the words of Weber (2011):

Grammatical Voice is manifested in systems in which alternations in the shapes of predicates or whole constructions signal alternations in the configurations of the syntactic and/or pragmatic status of (semantic) arguments of a predicate.

Consequently, a comparison of voice systems must take into account the levels of morphology, syntax, semantics and discourse.

3.3 Methodology for Studying Voice

In SUBSECTION 3.2, I argued that voice represents alternations between 'basic' transitive clauses and more marked constructions, whether they are marked at the level of syntax, semantics or discourse/information structure. This allows us to formulate a

¹⁵⁶ As discussed above, many do not consider the alternations in SUBSECTION 3.2.3 to be 'voice'. Nonetheless, I attempt to make my definition as broad as possible. The methodology presented in SUBSECTION 3.3 applies regardless of whether voice is understood in the canonical sense, or in the broader sense.

methodology for entering into the Western Austronesian alignment debate. In order to analyse alignment in a language with multiple transitive clauses, we need to identify which of the clauses is basic (cf. Kroeger 2004). If it is the one in which the undergoer is mapped to subject, then we have ergative alignment. If it is the one in which the actor is subject, then we have accusative alignment. In the previous section, I defined voice as pertaining to various structural levels, including morphology, syntax, semantics and discourse. Each of these levels can provide tests for identifying the basic transitive clause.

3.3.1 Morphosyntax

In many languages the opposition between basic and non-basic clauses is grammaticalised in morphology and syntax through morphological markedness and syntactic valency change. This was seen in both the Latin example in (3) and the Chukchi example in (8). The basic clauses in (3a) and (8a) have predicates that are morphologically unmarked for voice and syntactically transitive. The passive and antipassive alternations in (3b) and (8b), in contrast, involve additional morphology and detransitivisation. Thus, morphosyntax can help to establish which clause is the basic clause. A basic clause will be less morphologically marked than non-basic clauses, and a syntactically transitive clause will have two core nominal arguments, whilst a syntactically intransitive clause has only one. There are both cross-linguistic and language-specific tests for core syntactic status, as illustrated in SUBSECTION 1.4.2 (cf. Arka 2005). These are discussed in relation to Western Austronesian in SUBSECTION 3.4 and Kelabit in SUBSECTION 3.5.

However, as shown in SUBSECTION 3.2.1.3, many Western Austronesian languages have morphologically and syntactically symmetrical voice alternations in

which no basic transitive clause can be identified on the basis of morphology and syntax alone. In these situations, it falls to the levels of semantics and discourse to establish which clause is basic.

3.3.2 Semantics

Transitivity, as many studies have suggested, is not purely a syntactic notion (Croft 1994, Kemmer 1994). It can also be defined in semantic terms, where it is viewed as a cline rather than a binary distinction (Hopper & Thompson 1980, Kittilä 2011, Nagaya 2009a). Typically, semantic transitivity is defined as the transfer of an action from one participant to another, from the ‘source of action’ to the ‘most affected entity’ (cf. Nolasco 2005). Hopper & Thompson (1980: 252) suggest that this can be broken down into ten semantic parameters that have high and low transitivity values respectively:

Table 3.2 Transitivity Parameters following Hopper & Thompson (1980)

	High	Low
a. No. of Arguments	two or more participants	one participant
b. Kinesis	action	state
c. Aspect	telic	atelic
d. Punctuality	punctual	non-punctual
e. Volitionality	volitional	non-volitional
f. Affirmation	affirmative	negative
g. Mode	realis	irrealis
h. Agency	A high in agency	A low in agency
i. Affectedness of U	U totally affected	U not affected
j. Individuation of U	U highly individuated	U non-individuated

In other words, semantic transitivity is defined in terms of features related to the event (aspect/mood), features related to the actor (volitionality/agency) and features related to the undergoer (affectedness/individuation). The most proto-typically transitive event is one in which a given entity actively and volitionally initiates a complete

punctual effect on another distinct participant, who is totally affected by it. Importantly, semantic transitivity is distinct from syntactic transitivity, in that many syntactically ‘transitive’ constructions, such as reflexives, reciprocals and middle voice constructions, are semantically low in transitivity (cf. Givón 1994).

Assuming that these factors can give us a rough metric for the semantic transitivity of a given clause, we can use TABLE 3.2 as a means of identifying whether clauses differ in their semantic transitivity. Taking each clause in turn, a value of 1 can be assigned for each of the parameters with high transitivity properties and a value of 0 for each with low properties. These are added together to give a quantified semantic transitivity value from 0 - 10, which can be averaged across a text or corpus. This approach makes several assumptions. Firstly, it assumes that each of the properties is equally important to semantic transitivity, which may or may not be the case in a specific language.¹⁵⁷ Secondly, it treats the properties as being logically independent of one another, which is most probably not the case. For example, a low value for the number of participants automatically entails a low value for all the factors relating to U, as there is no distinct undergoer identifiable in that example. Finally, it assumes a binary distinction between low and high values, when many semantic categories are thought of as scalar.¹⁵⁸ Despite these difficulties, it remains a good starting point for comparing semantic transitivity within a particular language until a better model of cross-linguistic semantic transitivity is available.

¹⁵⁷ See Nolasco (2005) with regard to Philippine-type languages.

¹⁵⁸ See Timberlake (1977) for an early discussion of how the individuation of U involves a number of scales, including animacy, referentiality, number, person and more.

3.3.3 Discourse

The basic transitive clause can also be defined in discourse terms using token frequency and topicality measures (Givón 1983). The most frequent clause-type in discourse is usually taken to be the most basic (Kroeger 2004). To establish relative frequency, all clauses are coded for transitivity and those with AV marking and UV marking expressed as a percentage of the total. One would expect basic clauses to be more frequent than marked clauses.

The second test used to establish which clause is basic is the relative discourse topicality of arguments. According to Givón (1983), in active clauses the actor tends to have high topicality and the undergoer tends to have lower topicality but remain topical. Similarly, Cooreman, Fox & Givón (1984) show that ergative languages like Chamorro have the same patterns for ergative clauses: the actor tends to have high topicality and the undergoer tends to have lower topicality. Antipassive clauses, on the other hand, tend to have mid topicality for actors and even lower topicality for undergoers, whilst passive clauses tend to have high topicality for undergoers and low topicality for actors. Inverse clauses are functionally opposite to active clauses since two arguments have relative topicality but the undergoer is higher than the actor. This is summarised in TABLE 3.3:

Table 3.3 Topicality of Arguments (Cooreman 1987)

	Topicality of Arguments		
Active/Ergative	Actor	>	Undergoer
Inverse	Undergoer	>	Actor
Passive	Undergoer	>>	Actor
Antipassive	Actor	>>	Undergoer

The topicality of arguments can be assessed following Givón's (1983) quantitative metrics: referential distance (RD) and topical persistence (TP). RD is

calculated by counting back the number of clauses to the last mention of the argument. Thus, if an argument has been mentioned in the previous clause it will have the minimum value of 1 for RD. Givón (1983) arbitrarily sets the maximum value for RD at 20, which is also used for new referents. Though there are certain conceptual problems with this, for the sake of comparability the same convention is adopted in this thesis. TP is the mirror image to RD and counts forward the number of clauses in which the participant remains a semantic argument of the predicate, regardless of how it is encoded. Again, in keeping with Givón (1983) and Walters (1994), the minimum value of 0 is assigned if the participant is not a semantic argument in the following clause, whilst no maximum value is imposed. As Walters (1994: 132) discusses, RD bears an inverse relationship, whilst TP bears a direct relationship to the topicality of the arguments. He therefore devises a method of scaling the results and bringing them into a single figure. This is achieved by dividing $1/RD$ to bring the scale between 0.05 (approximately 0) and 1. Similarly, he divides TP by 3, as roughly the highest measured average. This brings the measure of TP to between 0-1 and removes the inverse correlation between the two measurements, allowing for straightforward averaging of the two metrics.

3.3.4 Summary

In summary, if voice is defined as a grammatical category that brings together morphological, syntactic, semantic and discourse factors, then any comparison of voice systems should consider all of these levels in order to establish which clause-type is basic. Consequently, a working methodology for analysing voice constructions in Kelabit is as follows:

Table 3.4 Methodology for Comparing Voices

Morphology	Is one of the voices more marked than the others?
Syntax	Is one of the voices syntactically transitive but not the others?
Semantics	Is one of the voices more semantically transitive than the others?
Discourse	Is one of the voices more ‘active/transitive’ in discourse terms?

This approach allows us to identify which, if any, of the clauses is more basic than the others. In turn, we can position voice systems on a scale from ergative to accusative, where symmetry at any one level is no longer problematic. We need not presuppose that all the levels necessarily select the same clause-type as basic and can identify if any of the levels provide conflicting evidence that could represent different stages in an alignment transition. I now turn to compare the morphosyntax, semantics and discourse properties of Western Austronesian languages, before applying these tests to Kelabit in SUBSECTION 3.5.

3.4 Western Austronesian Voice

In SUBSECTION 1.4.3, I argued against ergative and accusative analyses of Western Austronesian, on the basis that both Philippine-type and Indonesian-type languages have morphosyntactically symmetrical voice alternations. However, I have now defined voice and alignment as relating not only to morphology and syntax, but also to semantics and discourse. Consequently, in this section, I return to the alignment debate by considering the semantic and discourse differences between Western Austronesian voice systems. Using the methodology outlined in SUBSECTION 3.3.4, I explore which voice construction can be considered basic in proto-typical Philippine-type languages, proto-typical Indonesian-type languages and a selection of languages in Borneo and Sulawesi that are sometimes considered transitional (Ross 2002). This functions as a background against which to compare Kelabit, and also as

a preliminary exploration as to whether the two-way typology can capture important aspects of variation within Austronesian syntax.

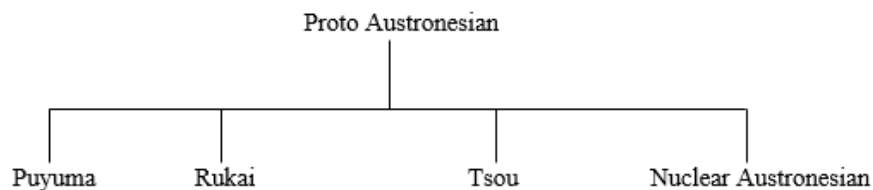
3.4.1 Philippine-type

Philippine-type languages typically have a four-way system of voice alternations (Arka & Ross 2005: 7, SUBSECTION 1.3.1). Peripheral roles, such as the locative and benefactive, have their own voice constructions and there is case-marking of nominal arguments to reflect their role within the voice system (Arka 2002). Philippine-type voice is generally considered to be the most conservative Western Austronesian voice system and has been reconstructed for Proto-Austronesian (Arka & Ross 2005, Adelaar 2005: 6).¹⁵⁹

Table 3.5 Proto-Austronesian Voice (Adelaar 2005: 6, following Ross 2002)

	Actor	Undergoer	Location	Circumstantial
Indicative				
Neutral	<um>V	V-en	V-an	Si-V
Perfective	<umin>V	<in>V	<in>V-an	Si-<in>V
Durative	<um>-R-V	R-V-en	R-V-an	Si-R-V
Non-Indicative				
Atemporal	V	V-u	V-i	V-áni
Projective	<um>V-a	V-aw	V-ay	V-ánay

¹⁵⁹ See Ross (2009) for an alternative view. He argues that three Formosan languages, Puyuma, Rukai and Tsou, do not share the innovation of a Philippine-type voice system. He therefore posits four primary subgroups of Austronesian, in contrast to the more commonly accepted ten primary subgroups outlined in SUBSECTION 1.2. This is illustrated below:



Nuclear Austronesian includes the remaining Formosan languages and Malayo-Polynesian and Ross (2009) reconstructs the ‘Philippine-type’ voice system to this stage. Zeitoun & Teng (2016) demonstrate that Saaroa and Kankana’nu also do not share the full innovations of Nuclear Austronesian and propose a modification to Ross (2009) accordingly.

Reflexes of the reconstructed PAn voice markers, particularly **-um-* and **-in-*, are found in many of the languages of Taiwan, Borneo, Sulawesi, Madagascar and the Philippines.

Whilst Cebuano and Tagalog reflect the proto-typical four-voice system, there are a number of Philippine-type languages, particularly in Taiwan and Northern Borneo, that have three-way voice systems. These differ according to the semantic roles that are represented. For example, the Formosan language Kavalan has an instrumental voice, but no benefactive or locative voice, as in (38).

- (38) *Kavalan* (Formosan)
- a. **Actor Voice**
q-<m>aRat saku 'nay 'tu mutun.
<AV>bite cat that OBL rat
‘That cat bit a rat.’
- b. **Undergoer Voice**
qaRat-an na saku mutun 'nay.
bite-UV GEN cat rat that
‘A cat bit that rat.’
- c. **Instrumental Voice**
ti-tabu na tina-ku tu baut ya biRi.
IV-wrap GEN mother-1SG.POSS OBL fish NOM leaf
‘My mother wrapped fish with the leaf.’
(Li & Tsuchida 2006: 26-27)

In contrast, Kadazan Dusun of Sabah has a benefactive voice, but no instrumental voice:¹⁶⁰

- (39) *Kadazan Dusun* (Sabahan)
- a. **Actor Voice**
Mog-ovit i ama' di tanak do buuk.
AV-bring PT father PT child PT book
‘Father is bringing the child a book.’

¹⁶⁰ Thao (Formosan) has actor voice, undergoer voice and locative voice (Blust 2013: 450).

b. **Undergoer Voice**

Ovit-on di ama' di tanak i buuk.
bring-UV PT father PT child PT book
'Father is bringing the child the book.'

c. **Benefactive Voice**

Ovit-an di ama' i tanak do buuk.
bring-BV PT father PT child PT book
'Father is bringing the child a book.' (Clayre 1991: 415)

Nonetheless, both Kadazan and Kavalan can be considered 'Philippine-type' as they encode voices for peripheral arguments, even if this is reduced from the PAn system in TABLE 3.5. Similarly, in both Kavalan and Kadazan Dusun nominal arguments are preceded by markers that indicate which semantic argument is mapped to subject. Let us now consider the morphosyntax, semantics and discourse properties of the voice constructions.

3.4.1.1 Morphosyntax

As discussed in SUBSECTION 1.4.2.1, many Philippine-type languages are morphologically and syntactically symmetrical, since the various alternations are equally morphologically marked and syntactically transitive (Kroeger 1993, Riesberg 2014). In this case, morphology and syntax do not provide evidence for treating any of the voices as more basic than the other.

However, not all languages with Philippine-type properties are equally symmetrical. Indeed, some display morphological and syntactic asymmetries. For example, in Pangutaran Sama UV is unmarked, whilst AV takes a prefix (see SUBSECTION 3.4.2.1 for similar discussion in relation to Balinese):

- (40) *Pangutaran Sama*
- a. **Undergoer Voice**
 Tigad onde' so.
 UV.cut child snake
 'The child cut the snake.'
- b. **Actor Voice**
 Mag-tigad onde' so.
 AV-cut child snake
 'The child cut a snake.'
- (Kroeger 2004: 302)

This would support an analysis of UV as basic at the level of verbal morphology and consequently an ergative analysis.¹⁶¹

In Tagalog there is a morphological argument for treating UV as unmarked in realis contexts, and AV as unmarked in irrealis contexts (B. Blake 1990, Himmelmann 1991, Kroeger 1993). Consider the paradigm of voice markers in TABLE 3.6:

Table 3.6 Tagalog Verbal Paradigm (Katagiri 2005: 159)

	Realis			Irrealis
	Infinitive	Perfective	Imperfective	Contemplated
AV	b<um>ili	b<um>ili	b<um>ibili	bibili
UV	bilh-in	<u>bin</u> ili	<u>bin</u> ibili	bibilh-in
LV	bilh-an	<u>bin</u> ilih-an	<u>bin</u> ibilh-an	bibilh-an
BV	i-bili	i- <u>bin</u> ili	i- <u>bin</u> ibili	i-bibili
IV	ipam-bili	ipin <u>am</u> -bili	ipin <u>am</u> -bibili	ipam-bibili

Katagiri (2005) argues that *-in-* (underlined in TABLE 3.6) should be analysed as a mood/aspect marker rather than a voice marker on the basis that it occurs in all perfective/imperfective forms of non-actor voices. Indeed, in languages like Ilokano, *-in-* even occurs in AV (cf. Katagiri 2005: 160):

¹⁶¹ Nb. there is at least one Formosan language for which morphological arguments can be made for treating AV as basic: Puyuma (Teng 2005: 139). In Puyuma, non-actor voices are always encoded through dedicated morphology, whilst the actor voice is sometimes expressed through zero-marking. This might support Ross's (2009) theory, in which Puyuma is distinct from Nuclear Austronesian languages.

- (41) *Ilokano*
Actor Voice
- | | | |
|-----------------|-------------------|---------|
| a. Root | <i>panaw</i> | ‘leave’ |
| b. Infinitive | <i>pumanaw</i> | |
| c. Perfective | <i>pimmanaw</i> | |
| d. Imperfective | <i>pimampanaw</i> | |

If this analysis is adopted, then TABLE 3.6 suggests that UV predicates are unmarked for voice in realis perfective and imperfective contexts. In contrast, AV is unmarked in irrealis contexts. Hence, morphological markedness differs in realis and irrealis mood. This has been taken to support an analysis of Tagalog as split-ergative, conditioned by mood/aspect (B. Blake 1990, Katagiri 2005). However, if realis clauses are taken to be more proto-typically transitive, following Hopper & Thompson (1980), then TABLE 3.6 could equally be taken to support an analysis of UV as basic.

Finally, there are syntactic phenomena in Philippine-type languages, which support an analysis of UV as basic. For example, in Kapampangan the cross-referencing or person-marking system suggests that UV constructions involve two grammatically distinct participants, whilst AV constructions only involve one:

- (42) *Kapampangan*
- a. **Actor Voice**
- | | | | | |
|-------------------------|-------------|---------|-----|-------|
| Mamangan | ya=ng | bayabas | ing | anak. |
| AV.eat | 3SG.NOM=LNK | guava | NOM | child |
| ‘the child ate guavas.’ | | | | |
- b. **Undergoer Voice**
- | | | | | |
|-------------------------|----------------------------|---------|-----|-------|
| Pengan ne | ning | pusa | ing | asan. |
| UV.eat | 3SG.GEN (na)+ 3SG.NOM (ya) | GEN cat | NOM | fish |
| ‘the cat ate the fish.’ | | | | |
- (Nolasco 2005)

In AV in (42a), only the absolutive argument is cross-referenced in *yang*. However, in UV in (42b), both the ergative and the absolutive argument are cross-referenced in the particle *ne*, which combines the ergative *na* and the absolutive *ya*. Nolasco (2005: 22)

therefore concludes that transitivity has been fully grammaticalised in Kapampangan and that the alignment is ergative.

Hence, Philippine-type languages vary in their morphosyntax. Whilst some languages are morphosyntactically symmetrical, others display morphological and syntactic asymmetries. In general, these tend to suggest that UV is basic, and therefore support an ergative analysis. In the next sections, I consider semantics and discourse, which present a much stronger motivation for this account.

3.4.1.2 Semantics

As discussed in SUBSECTION 1.4.2.1.2, the main evidence for the ergative hypothesis comes from the semantic parallels between UV and ergative/transitive clauses, and AV and antipassives (T. Payne 1982, Aldridge 2004). Many studies identify UV as high in semantic transitivity (cf. Nolasco 2005, Nagaya 2009a). For example, Nolasco (2005) analyses semantic transitivity in a corpus of Ilokano magazine stories, Cebuano folktales and Tagalog romance novels. He concludes that the UV construction represents high transitivity, since UV constructions are correlated with high intensity action, telicity, punctuality and volitionality. For example, consider (43):

- (43) *Cebuano*
- a. **Undergoer Voice**
- | | | | | |
|----------------|-----|------|-----|-----------|
| N-ahibalo-an | ni | Juan | ang | tinaguan. |
| PST-ka.know-UV | GEN | Juan | NOM | secret |
- ‘Juan discovered the secret.’ (punctual)
- b. **Actor Voice**
- | | | | | |
|---------------|-----|------|-----|-----------|
| N-akahibalo | si | Juan | sa | tinaguan. |
| PST-paka.know | NOM | Juan | GEN | secret |
- ‘Juan knows the secret.’ (non-punctual) (Shibatani 1988: 104)

In (43a), the UV construction has a punctual, dynamic interpretation. In (43b), however, the AV construction has a non-punctual, stative interpretation. Therefore, UV can be considered the basic transitive clause on a semantic level, which supports an ergative analysis (Nolasco 2005: 22).

Corpus studies also reveal semantic similarities between AV clauses and antipassives (Aldridge 2004, Nagaya 2009a). In Philippine-type languages, there is a constraint against definite undergoers in AV (SUBSECTION 1.4.2.1.2). In fact, the non-subject undergoer is typically interpreted as indefinite, nonspecific and non-presuppositional, which are cross-linguistic properties of the antipassive (Cooreman 1994, T. Payne 1982).¹⁶² This can be illustrated in Tagalog:

- (44) *Tagalog*
- a. **Actor Voice**
 Nagluto ang babae ng/*sa manok.
 AV.PFV.cook NOM woman a/*the chicken
 ‘The woman cooked a/*the chicken.’
- b. **Undergoer Voice**
 Niluto ng babae ang manok.
 UV.PFV.cook GEN woman NOM chicken
 ‘The woman cooked the chicken.’ (Katagiri 2005: 167)

- (45) *Tagalog*
- a. **Actor Voice**
 *Pumatay si Juan ng aso.
 AV.PFV.kill NOM Juan GEN dog
 For: ‘Juan killed a dog.’
- b. **Undergoer Voice**
 Pinatay ni Juan ang aso.
 UV.PFV.kill GEN Juan NOM dog
 ‘Juan killed the/a dog.’ (Katagiri 2005: 169)

¹⁶² See Aldridge (2004) for corpus examples.

In Tagalog, the AV undergoer is typically indefinite, as in (44a). Moreover, AV is ungrammatical in contexts where the undergoer is highly affected, as in (45a).¹⁶³ Thus, AV correlates with low degrees of affectedness and identifiability in the undergoer, which is common in antipassives (SUBSECTION 3.2.1.2) and suggests that AV is lower in semantic transitivity than UV.

Nonetheless, definiteness constraints vary in Philippine-type languages. For example, in Cebuano both definite and indefinite undergoers occur in AV, but with different case-markers:

- (46) *Cebuano*
- a. **Actor Voice**
 Miluto ang babaye ug/sa manok.
 AV.PFV.cook NOM woman a/the chicken
 ‘The woman cooked a/the chicken.’
- b. **Undergoer Voice**
 Giluto sa babaye ang manok.
 UV.PFV.cook GEN woman NOM chicken
 ‘The woman cooked the chicken.’ (Katagiri 2005: 167)

Similarly, both AV and UV are possible alternatives in contexts where the undergoer is highly affected:

- (47) *Cebuano*
- a. **Actor Voice**
 Mipatay si Juan ug/sa ero.
 AV.PFV.kill NOM Juan a/the dog
 ‘Juan killed the dog.’
- b. **Undergoer Voice**
 Gipatay ni Juan ang ero.
 UV.PFV.kill GEN Juan NOM dog
 ‘Juan killed the dog.’ (Katagiri 2005: 169)

¹⁶³ See Ceña (1977: 6-7) for examples in Tagalog where both AV and UV are possible. These tend to indicate different degrees of affectedness.

This implies that Cebuano AV may be higher in semantic transitivity than the Tagalog equivalents in (44) and (45). Hence, Philippine-type languages differ in the semantic properties associated with AV.

Overall, semantic evidence seems to support an analysis of UV as the basic transitive clause in Philippine-type languages. UV tends to have high semantic transitivity and AV tends to be lower in semantic transitivity. However, Philippine-type languages are not uniform in their treatment of AV, and some allow definite undergoers more readily than others. This is in keeping with the Aldridge (2011) view of alignment shift, which I return to in SUBSECTION 3.5.

3.4.1.3 Discourse

Finally, frequency and topicality measures also support an analysis of UV as basic in Philippine-type languages. A number of studies show that UV constructions are more frequent than AV (Kroeger 2004) though this may be affected by genre (see SUBSECTION 5.5.2). For example, Gault (1999) found that 75% of transitive clauses in Sama Bangingi texts were UV and other minor clause types, whilst only 25% were AV constructions. Similarly, Cooreman, Fox & Givón (1984) found that 59% of the transitive clauses in their sample of Tagalog (166 of 281) used UV, as opposed to 24% AV. Finally, Walters (1994) found that 75% of transitive clauses in the Cebuano narrative in Wolff (1967) were UV and only 25% AV.¹⁶⁴ Hence, frequency counts overwhelmingly support an analysis of UV as basic.

Furthermore, topicality measures also support an analysis of UV as basic. Firstly, UV has been shown to correlate with foregrounded clauses (see Hopper & Thompson 1980 on Tagalog). Secondly, UV clauses have been shown to have the

¹⁶⁴ Walters (1994) does not include *ma*- clauses.

topicality metrics of active/ergative clauses. For example, Walters (1994) analysed the topicality patterns of UV and AV in a Cebuano narrative using the methodology in SUBSECTION 3.3.3. The results are summarised in TABLE 3.7:

Table 3.7 Scaled Average Topicality (SAT) of Arguments in Cebuano (Walters 1994: 134)

	SAT of Actor	SAT of Undergoer
AV	0.41	0.18
UV	0.89	0.18

On average, UV clauses have a highly topical actor, and an undergoer with lower topicality. Walters (1994) interprets this as the topicality patterns expected of a basic active clause. In contrast, AV seems to have the topicality patterns of an antipassive with a mid-topicality actor and a low-topicality undergoer. Hence, discourse measures would also support an ergative analysis, in which UV is the basic transitive clause and AV is antipassive-like.

In summary, there are a number of arguments for treating UV as the basic clause-type in Philippine-type systems and these can be summarised as follows:

Table 3.8 Philippine-type Voice Systems

Level of analysis	AV transitivity	UV transitivity
Morphology	✓ (?)	✓
Syntax	✓ (?)	✓
Semantics	x	✓
Discourse	x	✓

In proto-typical examples, the voice alternations are morphosyntactically symmetrical. However, UV is generally associated with higher semantic and discourse transitivity than AV. This would support an analysis of Philippine-type languages as ergative at discourse and semantic levels.

3.4.2 Indonesian-type

The Indonesian-type voice system is generally agreed to constitute a historical innovation (Adelaar 2005).¹⁶⁵ Like Philippine-type systems, they are morphosyntactically symmetrical. However, they have innovated a two-way system of morphologically-marked alternations, true passive constructions, and applicative suffixes that attach in both AV and UV (SUBSECTION 1.3.1). These systems are found in a number of languages from Standard Indonesian (Musgrave 2002) to Balinese (Arka 2003) and several languages spoken in Borneo, such as Mualang, a Malayic language in Western Borneo (Tjia 2007) and Manyaan, a Barito language in Central Kalimantan (Gudai 1985).

AV is typically expressed through a nasal prefix, such as *meN-* in Indonesian or *N-* in Javanese (Crouch 2009). The nasal undergoes substitution or assimilation, depending on whether the root begins with a voiced consonant, voiceless consonant or a vowel (cf. Blust 2004). Ross (2002) treats the prefix as an innovation that replaces PAN **-um-*. It most likely derives from the proto Malayo-Polynesian prefix **maŋ-*, which has cognates in Philippine-type languages where it is generally used as an intransitive prefix (cf. Blust 2013). UV is also marked with a prefix, such as *di-* in Indonesian, which is normally oral rather than nasal (see Crouch 2009). The *di-* prefix is sometimes analysed as a development from *-in-*, via *ni-* (Blust 2013: 452). However, its provenance remains controversial and has been subject to many hypotheses (see Adelaar 2009). For example, Adelaar (2009) suggests that it may be cognate with the locative preposition *di*. Hence, the PAN voice markers **-um-* and **-in-* do not play as

¹⁶⁵ See Starosta et al (1982) and Wolff (1996) for discussion of potential historical developments towards the Indonesian-type system. An alternative theory exists, which argues that proto Malayo-Polynesian had a three-way voice system like Chamorro or Nias, rather than the Philippine-type system (see Donohue 2007b). Nonetheless, the majority view is that Philippine-type voice systems are reconstructable back to earlier stages of the language, as discussed in SUBSECTION 3.4.1.

central a role as in Philippine-type languages, and tend to survive only as reflexes (Blust 2013: 454).

Many varieties of Malay/Indonesian differ from the standard in their morphosyntax. In some cases, varieties do not have morphological voice oppositions at all. Consider Kelantan Malay:

(48) *Kelantan Malay*

a. **Actor Voice?**

Kucī makē ikē.
cat eat fish
'The cat ate the fish.'

b. **Undergoer Voice?**

Ikē ɲə makē k~kucī.
fish UV? eat NON.SUBJ~cat
'The fish was eaten by the cat.'

(Donohue 2007b: 80)

In (48b), the UV construction involves a separate particle *ɲə* and reduplication to signal that the actor is a non-subject argument. It is not clear if the actor has been demoted and hence if this functions as UV or passive. In Papuan Malay there does not seem to be a passive or an undergoer voice and consequently all sentences are described as 'active' (Donohue 2007b). Hence, Donohue (2007b) argues that Malay/Indonesian varieties are moving towards the typologically more common systems in which actors are not mapped to the function of object. For discussion of variation in Malay/Indonesian see Donohue (2007b) and Cole, Hermon & Yanti (2008).

3.4.2.1 Morphosyntax

Like Philippine-type languages, many Indonesian-type voice systems are morphologically and syntactically symmetrical (see Riesberg 2014, SUBSECTION 1.4.2.2.3). Nonetheless, there are some morphosyntactic asymmetries that are relevant

to the question of alignment. These paint a more varied picture of Indonesian-type languages, in keeping with the view that they may be at various stages in the process of alignment shift.

Firstly, whilst languages such as Madurese have morphologically marked AV and UV alternations, regardless of the animacy/definiteness of arguments (Davies 2005, 2010),¹⁶⁶ other languages display morphological asymmetries, as shown in CHAPTER 1. In some cases, UV appears less marked, perhaps reflecting development from an ‘ergative’ Philippine-type system. In Javanese, for example, UV predicates are morphologically unmarked when the actor is a speech act participant (see SUBSECTION 1.3.1 for similar patterns in Indonesian):

- (49) *Javanese*
 a. **Undergoer Voice (1SG actor)**
 Surat wis tak=kirim.
 letter PFV 1SG=send
 ‘I sent the letter.’ (elicitation, fieldnotes)

Moreover, in Balinese, UV is morphologically unmarked, even when the actor is a third person:

- (50) *Balinese*
 a. **Actor Voice**
 Tiang nyepak cicing-e.
 N-sepak
 1SG AV-kick dog-DEF
 ‘I kicked the dog.’
 b. **Undergoer Voice**
 Cicing-e sepak tiang.
 dog-DEF UV.kick 1SG
 ‘The dog was kicked by me.’ (Artawa 1998: 8)

¹⁶⁶ Though there is also a bare construction in Madurese which is accepted by at least some speakers and attested in natural discourse. See Davies (2005: 201-2) for discussion.

These facts might support an analysis of UV as basic at the level of morphology. Accordingly, morphological arguments have been given to support an ergative analysis of Balinese (see SUBSECTION 1.4.2.2.2). However, there are also languages, such as Nias, where AV is unmarked and UV is marked (Weber 2011). Hence, Indonesian-type languages vary in their morphology.

Indeed, like *-in-* in Tagalog (SUBSECTION 3.4.1.1), verbal morphology has been argued to mark aspect or other categories, rather than voice, in some Indonesian-type languages (Soh & Nomoto 2009, Donohue 2007b). For example, in Indonesian *di-* is sometimes claimed to represent formality rather than undergoer voice, since it can be used in contexts that appear active. For example, consider the following utterance in the context of buying cassettes at a shop:

- (51) *Indonesian*
 a. Dua ini ya pak mau di-beli?
 two this AG sir want di-buy
 ‘So it’s these two that you want to buy, is it?’
 (Donohue 2007b: 124)

In (51), *di-* does not seem to mark UV but rather signals a formal rather than informal exchange.

Similarly, in some varieties of Sasak, the nasal prefix gives a contrastive predicate focus reading rather than indicating actor voice:

- (52) *Ngenó-ngené Sasak*
 Ali m-pantòk tèmòk.
 Ali N-hit wall
 ‘Ali hit the hill (he didn’t kick it).’
 (Austin 2013: 36)

This leads to the argument that voice alternations in some Indonesian-type languages may not be morphologically marked at all (see also SUBSECTION 3.4.2.2).

In addition to morphological asymmetries, there are also syntactic asymmetries in Indonesian-type languages (Riesberg 2014). For example, consider the behaviour of secondary predicates in Balinese. Secondary predicates can modify both the subject and the non-subject argument in AV, but only the subject in UV. Similarly, secondary predicates cannot modify obliques:

- (53) *Balinese*
- a. **Actor Voice**
 Tiang nguber i Nyoman ibi malalung.
 1 AV.chase PERS Nyoman yesterday naked
 ‘I chased Nyoman yesterday and I was naked.’
 Or: ‘I chased Nyoman yesterday and he was naked.’
 (Arka 2003: 56)
- b. **Undergoer Voice**
 I Nyoman uber tiang malalung.
 Pers Nyoman UV.chase 1 naked
 ‘Nyoman was chased by me and he was naked.’
 *‘Nyoman was chased by me and I was naked.’ (Arka 2003: 57)
- c. **Obliques**
 Ia matakon teken anak-e ento ibi punyah.
 3 question to person-DEF DET yesterday drunk
 ‘He_i asked the person_j questions yesterday and he_{i/*j} was drunk.’
 (Arka 2003: 57)

This suggests that the non-subject argument may be less core in UV than in AV, which supports an analysis of AV as basic at the level of syntax. Overall, Riesberg (2014: 84) argues that there is more evidence for treating AV and UV as symmetrical in Balinese, than patterns like (50) and (53). For example, complex quantifiers, such as *ajak makejang* ‘all’, present the same patterns of symmetry as Indonesian in SUBSECTION 1.4.2.2.3.¹⁶⁷ Nonetheless, this goes to show that syntactic symmetry is perhaps a matter of degree, rather than a straightforward dichotomy.

¹⁶⁷ Though simple quantifiers pattern with secondary predicates (see Riesberg 2014, Arka 2003).

Hence, morphology and syntax generally do not determine which clause-type is basic in Indonesian-type languages any more than in Philippine-type languages. Nonetheless, languages differ in their degree of morphological and syntactic symmetry (see Arka 2005) and these differences could perhaps be argued to reflect different points in the transition from ergative to accusative (SUBSECTION 3.5).

3.4.2.2 Semantics

Morphosyntactic variation is mirrored at the levels of semantics and discourse. Some semantic studies support an analysis of UV as high in semantic transitivity. For example, Wouk (1989, 1996, 2004) found that Standard Jakartan Indonesian *di*-clauses are associated with high transitivity properties, such as indicative mood and dynamic semantics. Moreover, they typically had animate, individuated and referential undergoers. In contrast, *N*-clauses are associated with irrealis mood, stative semantics and often have non-referential undergoers.

Similarly, in other varieties of Indonesian, the UV prefix is said to correlate with perfective aspect, whilst the AV prefix correlates with imperfective aspect (Rafferty 1982, Soh & Nomoto 2008). Soh & Nomoto (2009, 2015) suggest that *meN*-clauses in Malay are associated with atelicity, as shown through the comparison of clauses that can either take the *meN*- prefix or not:

- (54) *Malay*
- a. **No voice prefix**
- | | | | | |
|---------------------|---------|------------|------------|---------|
| Malaysia akan bina | se-buah | makmal | pengawasan | nuklear |
| Malaysia will build | one-CLF | laboratory | control | nuclear |
- di Bukit Ibam.
in Bukit Ibam
'Malaysia will build a nuclear control laboratory in Bukit Ibam.'

b. **Actor voice prefix**

Malaysia akan mem-bina se-buah makmal pengawasan
 Malaysia will AV-build one-CLF laboratory control

nuklear di Bukit Ibam.
 nuclear in Bukit Ibam

‘Malaysia will be building a nuclear control laboratory in Bukit Ibam.’
 (Soh & Nomoto 2009: 152)

This would suggest that overtly marked AV clauses correlate with low semantic transitivity in the sense of Hopper & Thompson (1980). Similar contrasts are found in northern and eastern Sasak, where the nasal prefix can indicate ongoing action and/or non-referential undergoers, which are both characteristics of low semantic transitivity (Austin 2013: 44).

Finally, many studies suggest a correlation between UV and foregrounded clauses, and AV and backgrounded clauses that was also found in Philippine-type languages (SUBSECTION 3.4.1.3). For example, consider Madurese:

(55) *Madurese*

a. **Actor Voice**

Reng lake’ gelle’ ngrabadi Bang.Pote.
 person male previous AV.care.APPL Garlic
 ‘The man took care of Garlic (name).’

b. **Undergoer Voice**

Samper gelle’ etabang bi’ Bang.Pote.
 cloth previous UV.search with Garlic
 ‘Garlic (name) searched for the cloth.’ (Davies 2005: 213)

Davies (2005) argues that (55a) provides background information, whilst (55b) expresses an important event in the main storyline. Similar patterns are said to hold for Classical Malay (see Hopper 1979), which was the literary language of the Malacca Empire from the 17th-19th century (Cumming 1991). Hence, semantics might support an analysis of UV as basic in these cases.

However, Indonesian-type AV is unlike its Philippine-type equivalent in that there is no definiteness constraint against definite undergoers in AV. This is seen in the Balinese alternation in (50), repeated below:

- (56) *Balinese*
- a. **Actor Voice**
- | | | |
|-------|---------|-----------|
| Tiang | nyepak | cicing-e. |
| 1SG | AV.kick | dog-DEF |
- ‘I kicked the dog.’
- b. **Undergoer Voice**
- | | | |
|----------|---------|--------|
| Cicing-e | sepak | tiang. |
| dog-DEF | UV.kick | 1SG |
- ‘The dog was kicked by me.’ (Artawa 1998: 8)

In both AV and UV, the undergoer *cicing* ‘dog’ can be marked with the definiteness suffix *-e* (see Wouk 2010 for similar facts in Sasak). Hence, even if Indonesian-type AV has correlations with imperfective aspect, it is unlike typical antipassives. Consequently, *meN-* is more often interpreted as a marker of transitivity than intransitivity (Cole & Hermon 1998).

Furthermore, in some Indonesian-type languages *di-* clauses have semantic properties associated with passives. For example, in a corpus study in Nomoto & Kartini (2014), *di-* clauses most commonly appeared without an overt actor, which is a cross-linguistic tendency of passives (Keenan & Dryer 2006). Moreover, the actor in such cases is often non-specific and low in identifiability. Hence, semantics also suggests a transition, with some Indonesian-type languages maintaining the more conservative patterns of UV as high in transitivity and AV as low in transitivity, whilst other have developed AV clauses that are higher in transitivity and UV clauses with the semantic properties of passives.

3.4.2.3 Discourse

Discourse measures also show variation. AV tends to be more frequent than in Philippine-type languages, which supports an analysis of AV as transitive rather than antipassive. In a number of Indonesian-type languages, AV and UV are fairly equal in terms of discourse frequency. For example, Pastika (1999) suggests that AV and UV appear to have roughly equal distribution in Balinese spoken texts and AV outnumbers UV at 70% to 30% in written texts. Similarly, Davies (2005: 212) reports that AV and UV transitive clauses are roughly equal in a corpus of Madurese folktales and historical narratives, with 45-53% AV and 47-55% UV.¹⁶⁸ As for Indonesian, Cumming (1995: 255) found that AV clauses occurred 72.7% of the time in the Modern Indonesian corpus. Interestingly, this was not the case in Classical Malay, where the percentage of *di-* clauses was 73% (Cumming 1991: 162). Similarly, the asymmetrical languages spoken in Eastern Flores have AV clauses that are basic in terms of distribution and frequency as well as morphosyntax (see SUBSECTION 1.3.2). Hence, a strong argument can be made for treating this as evidence of a transition from discourse ergative to discourse accusative.

Topicality measures also provide mixed results. There is some argument for treating UV as the basic transitive clause at a discourse level, since it can be used to describe a sequence of events performed by the same actor (Kaswanti Purwo 1988: 205). This is illustrated in Malay:

¹⁶⁸ This is higher still if cases of UV morphology + applicative suffixes are treated as distinct voices (see Davies 2005). Then the proportion of AV to UV clauses is 48% AV to 38.3% UV.

- (57) *Malay*
- a. **Undergoer Voice**
- Tudung botol di-rentap keluar, kemudian... di-jurus=nya minyak
 cover bottle UV-pull go.out afterward UV-pour=3SG oil
- tanah itu ke atas lantai... botol itu lalu di-himbau=nya ke
 earth that to top floor bottle that then UV-throw=3SG to
- tengah sungai.
 middle river
- ‘She pulled off the bottle cap, then... she poured the kerosene onto the
 floor of the boat... then she threw the bottle to the middle of the river.’
 (Kroeger 2014: 22)

In (57), the *di-* clauses describe events where the actor, and not the undergoer, is high in topic continuity. As such, the actor is either expressed using the clitic pronoun =*nya* or through zero anaphora. Moreover, Himmelmann (2005a: 136) and Gil (2008) claim that UV clauses are often acquired earlier in Indonesian-type languages, suggesting that they could be basic.

However, AV clauses typically have the discourse characteristics of active clauses rather than antipassives (see Wouk 1999: 105 on Selong Sasak). Moreover, at least in Indonesian, there is evidence for treating AV as discourse active and UV *di-* clauses as discourse passive. This can be seen if we consider a traditional story, *Kuda Oncesrawa*, written in Standard Indonesian as part of the blog *I made winangun arta*.¹⁶⁹ The story compares in terms of length and genre with Walters (1994) and the Kelabit story discussed in SUBSECTION 3.5.3. In total, there were 51 unambiguously transitive clauses. 38 clauses, or 74.5%, were in AV, and 13, or 25.5%, were in UV. Applying the methodology in SUBSECTION 3.3, the following topicality measures emerge:

¹⁶⁹ <http://winangun.blogspot.com/2008/05/bilingual-story-indonesian-english.html>. This is a traditional story said to come from Bali. Details about the origin of the author are not given though the text is written in Standard Indonesian, and translated into English in tandem with the Indonesian.

Table 3.9 Scaled Average Topicality (SAT) in Indonesian

	SAT of Actor	SAT of Undergoer
AV	0.53	0.36
UV	0.27	0.53

TABLE 3.9 presents the opposite patterns to Cebuano (SUBSECTION 3.4.1.3). AV has the discourse properties of an active/transitive clause, in that both actor and undergoer are topical, but the actor is higher in topicality than the undergoer. In contrast, UV has topicality patterns that are consistent either with an analysis of UV as a passive or a type of inverse construction, since the undergoer is higher in topicality than the actor and the actor is relatively low.¹⁷⁰ Hence, discourse evidence in Indonesian supports an analysis whereby AV is basic, and UV passive, which suggests accusative alignment at the level of discourse.

Consequently, there is some evidence for treating languages like Indonesian as accusative at semantic and discourse levels, since AV has the semantic and discourse characteristics of an active clause and UV has the semantic and discourse characteristics of a passive. This is summarised in TABLE 3.10:

¹⁷⁰ As discussed in SUBSECTION 1.3.1, some consider that *di-* clauses are passives and reserve the use of UV solely for bare-predicate constructions, and/or constructions with third person pronominal agents marked *di- -nya*. In TABLE 3.9, the 13 UV clauses include the following types:

- (i) *di-* + NP actor (3)
- (ii) *di-* + zero actor (4)
- (iii) *di-* + *-nya* (2)
- (iv) bare predicates with first/second person actors (4)

If constructions (iii) and (iv) are truly different from (i) and (ii), then we might expect to find differences in the topicality measures in the story *Kuda Oncesrawa*. However, this is not the case. The scaled average topicality of the actor is 0.26 (iii and iv) vs 0.27 (i and ii) and the undergoer is 0.53 (iii and iv) vs 0.51 (i and ii). This suggests that there may not be a huge discourse difference between the constructions, though the number of examples is too small to draw any firm conclusions.

Table 3.10 Voice in Indonesian

Level of analysis	AV transitivity	UV transitivity
Morphology	✓	✓
Syntax	✓	✓
Semantics	✓ (?)	?
Discourse	✓	? (x)

However, the treatment of UV varies in other Indonesian-type languages. For example, in Balinese UV maintains properties of basic clauses. Thus, Indonesian-type languages appear to differ in their alignment systems which could well reflect a transition from ergative to accusative. At the very least, this calls into question the validity of ‘Indonesian-type’ as a meaningful typological grouping. This becomes even more apparent when we consider the variation in voice systems of Borneo and Sulawesi.

3.4.3 Borneo and Sulawesi

Many languages in Borneo and Sulawesi can be considered ‘transitional’, since they have a mixture of Philippine-type and Indonesian-type characteristics, as well as unique areal features of their own. This is demonstrated in Clayre (2014)¹⁷¹ and van den Berg (1995), who present a typology of voice systems in Middle Borneo and Sulawesi, summarised as follows:

- (58) *Voice Systems in Middle Borneo and Sulawesi*
- a. **Full Systems**
 - Three or more voices
 - Verbal affixation
 - Nominal and pronominal marking
 - Three sets of pronouns

¹⁷¹ Summarising a survey of Apad Uat, Kenyah, Kayan and Rejang-Baram languages, conducted in 1996.

b. **Reduced Systems**

Two voices
Reduced/no nominal marking
No more than two pronoun sets
Word-order variation

c. **No Voice System**

The only remnants are found in subordinate/relative clauses

Languages with full systems are essentially Philippine-type, as discussed in SUBSECTION 3.4.1, whilst languages with no voice system resemble the asymmetrical languages discussed in SUBSECTION 1.3.2. The most interesting group of languages are those with ‘reduced voice systems’ as these differ from both Philippine-type and Indonesian-type languages, at morphosyntactic, semantic and discourse levels.

3.4.3.1 Morphosyntax

Full system languages in Clayre (2014) and van den Berg (1995) include Lundayeh in Northern Sarawak and Tondano in Northeast Sulawesi. The voice systems are illustrated in (59) and (60):

(59) *Lundayeh*

a. **Actor Voice**

Ieh ni'er negku.
3SG.1 AV.see 1SG.3
'He saw me.'

b. **Undergoer Voice**

Beli-en ku lal neh ku usin nih.
buy-UV.IRR 1SG.2 hen DEM with money DEM
'I'll buy the hen with this money.'

c. **Instrumental Voice**

Pimeli ku lal usin nih.
IV.buy 1SG.2 hen money DEM
'I'll use this money to buy the hen.'

(Clayre 2014: 132-133)

- (60) *Tondano*
- a. **Actor Voice**
 Si tuama k<um>eong roda wo ntali witu lalan.
 PT man <AV>pull cart with rope on road
 ‘The man will pull the cart on the road with the rope.’
- b. **Undergoer Voice**
 Roda keong-ən ni tuama wo ntali witu lalan.
 cart pull-UV PT man with rope on road
 ‘The man will pull the cart on the road with the rope.’
- c. **Instrumental Voice**
 Tali i-keong ni tuama roda witu lalan.
 rope IV-pull PT man cart on road
 ‘The man will pull the cart on the road with the rope.’
- d. **Locative Voice**
 Lalan keong-an ni tuama roda wo ntali.
 road pull-LV PT man cart with rope
 ‘The man will pull the cart on the road with the rope.’

(Sneddon 1975)

Lundayeh has a three-voice system of alternations, like Kadazan Dusun or Kavalan in SUBSECTION 3.4.1. However, unlike other Philippine-type languages, there is no case-marking of nominal arguments. Tondano has the typical four-voice system and uses reflexes of PAn morphology. However, pre-nominal particles only occur with actor arguments and the clauses are subject-initial rather than verb-initial (see CHAPTER 5). Hence, even the ‘Philippine-type’ languages in Borneo and Sulawesi are somewhat different from the languages of the Philippines.

The majority of the languages surveyed in Clayre (2014) and van den Berg (1995) are considered ‘reduced systems’. However, this is a fairly heterogeneous group. Some languages, such as Sa’ban in Northern Sarawak, have Indonesian-type two-voice systems:

- (61) *Sa'ban*
- a. **Actor Voice**
- | | | | |
|--------|--------|---------|---------------------|
| Yuet | noknai | n-toe | éek. ¹⁷² |
| monkey | this | AV-drop | 1SG |
- ‘This monkey drops me.’
- b. **Undergoer Voice**
- | | | | |
|--------|--------|---------|------|
| Yuet | noknai | i-toe | éek. |
| monkey | this | UV-drop | 1SG |
- ‘I dropped the monkey.’ (Clayre 2014: 138)

Like Indonesian-type languages, Sa'ban has two symmetrical voices, marked with a nasal and oral prefix. However, unlike Indonesian-type languages, Sa'ban does not have applicatives or a true passive construction (Clayre 2014).¹⁷³

Moreover, many of the languages of South Sulawesi have developed voice systems not found elsewhere, in which the voice system is reduced and the transitivity of the predicate overtly coded. There is widespread use of pronominal affixes but little nominal marking. This can be illustrated from Uma:

Table 3.11 *Transitivity in Uma* (van den Berg 1995)

Transitivity	Actor	Undergoer	Verbal Morphology
Stative	NP/person suffix	-	<i>mo-/∅</i>
Intransitive	NP/person suffix	-	<i>mo-/ma-/∅</i>
Detransitive	NP/person suffix	-	<i>me-</i>
Incorporated	person suffix on NP	NP	<i>N-/mpo-</i>
Antipassive	person suffix	NP/-	<i>N-/mpo-</i>
Actor Voice	NP/-	NP/person suffix	<i>N-/mpo-</i>
Undergoer Voice	person prefix	NP/person suffix	-

¹⁷² *noknai* is glossed as ‘this’ – however, it is very similar to Kelabit *nuk na'ah ih* ‘the aforementioned’ and may be subject to the same process of assimilation (see SUBSECTION 2.3.5.2). In Kelabit, this has a discourse-structuring function and refers back to a previously mentioned referent.

¹⁷³ Two-voice systems are also found in Berawan, Melanau, Sebop and other languages. However, unlike Sa'ban, variant pronoun forms are used for actor subjects and actor non-subjects (see CHAPTER 4 for discussion). Moreover, varieties like Mukah Melanau use the voice markers AV *-əm-* and UV *-ən-* among others, which are cognate with PAn morphology (cf. Blust 2013: 402).

The voice alternations are illustrated in (62):

- (62) *Uma*
- a. **Actor Voice**
- | | | |
|-----------------|-----------|-------|
| Tuama-ku | m-po-'oli | once. |
| Father-1SG.POSS | AV-TR-buy | rice |
- ‘My father bought rice.’
- b. **Undergoer Voice**
- | | | |
|---------|------|-------|
| Ku-'oli | once | tetu. |
| 1SG-buy | rice | DEM |
- ‘I bought that rice.’

(Esser 1964, Martens 1988abc)

Like the bare predicate construction in Javanese (SUBSECTION 3.4.2), the UV verb is unmarked, but takes an agreement prefix with the actor. There are no remnant forms of the *-um-* infix, and *-in-* remains only in lexicalised nominal forms, e.g. *pinu'ai* ‘dried rice’ from *pu'ai* ‘dry in the sun’. Van den Berg (1995: 8) suggests that AV is typically used in relative clauses, cleft constructions and complement clauses and UV is used when the undergoer is definite and foregrounded in discourse. Hence, although this is a two-voice system, AV appears more marked in terms of morphology and distribution than UV.

Finally, there are reduced voice languages in which AV appears morphosyntactically basic. For example, in Kayan of Northern Sarawak, UV is variously marked via bare predicates, remnant morphology and a periphrastic construction, shown in (63):

- (63) *Kayan*
- a. **Periphrastic Undergoer Voice**
- | | | | | | |
|-------|-------|------|--------|-----|-------|
| En | na' | 'uk | basung | men | iha'. |
| UV.do | 3SG.2 | give | shift | to | 3SG.1 |
- ‘He gave the shift to him.’
- (Clayre 2014: 141)

Though bare predicates might suggest that UV is morphologically simpler, the use of periphrastic constructions like (63) would suggest that UV can be more marked than AV and would therefore support an analysis of AV as morphosyntactically basic, and Kayan as accusative.

This is also true of ‘no voice’ languages, such as Kenyah in Northern Sarawak, and Muna in Southeast Sulawesi. Kenyah languages are analysed as having no voice system, since there is no morphological means of mapping undergoer to subject. However, Rahmajanti (1995) suggests that there may be an alternation, signalled through word order:

- (64) *Lepo’ Ké Kenyah*
- a. **Actor Voice**
 Amai meli sapai m-aké.
 Father buy shirt for-1SG
 ‘Father bought a shirt for me.’
- b. **Undergoer Voice (?)**
 Aké amai meli ca sapai.
 1SG father buy one shirt
 ‘I was bought a shirt by father.’ (Rahmajanti 1995: 29)

It is not clear if there has been a change in subject in (64b), or only a topicalisation. In any case, this is similar to the Eastern Flores languages discussed in SUBSECTION 1.3.2.

Similarly, in Muna, a Muna-Buton language, basic clauses are actor voice and indicate the subject through person prefixes on the verb:

- (65) *Muna*
- a. **Basic (AV) clause**
 A-gholi kalei-no.
 1SG.REAL-buy banana-3SG.REAL
 ‘I bought his bananas.’ (van den Berg 1995: 5)

If there is a UV construction, it is marked through word order rather than verbal morphology:

- (66) *Muna*
- a. **Undergoer Voice (?)**
- | | | | |
|--------|-----|-----------------|-----------------|
| Kalei | ini | no-gholi-e | ina-ku. |
| banana | DEM | 3SG.REAL-buy-it | mother-1SG.POSS |
- ‘These bananas were bought by my mother.’ (van den Berg 1989)

However, there are some reflexes of *-um-* and *-in-* found in relative clauses:

- (67) *Muna Remnant Voice Morphology in Relative Clauses*
- a. **Head = Actor**
- | | | | |
|--------|-------------|-----|--------|
| Mie | k<um>alo-no | we | daoa. |
| person | <AV>go-3SG | LOC | market |
- ‘The person who went to the market.’
- b. **Head = Undergoer**
- | | |
|--------|-----------------|
| Kalei | ni-gholi-ku. |
| banana | UV-buy-1SG.POSS |
- ‘The bananas that I bought.’ (van den Berg 1989)

Hence, Muna and Kenyah have some reflexes of PAn morphology, but do not mark voice alternations morphologically like other languages in Middle Borneo and Sulawesi.

Thus, transitional languages in Middle Borneo and Sulawesi are subject to variation in their voice systems, at least in terms of the following morphosyntactic properties:

- (68) **Variation in Voice Systems**
- a. Number of alternations encoded
 - b. Presence or absence of case-marking/nominal particles
 - c. Pronominal system
 - d. Reflexes of **-um-* and **-in-*
 - e. Morphological/periphrastic voice constructions
 - f. Marking ‘voice’ phenomena through word order

3.4.3.2 Semantics

In the languages of Borneo and Sulawesi, there is also variation in terms of the semantic and discourse properties of the voice constructions, particularly in the status of AV. In some languages, AV clauses have the characteristics of active clauses. In Kimaragang, for example, *-um-* is said to indicate intransitivity, whilst *m-oN-*, the AV prefix, marks high transitivity (see SUBSECTION 2.4.1.2 on Kelabit *-em-* and *-um-*). In other languages, the AV construction is described as an antipassive. For example, Mead (1998) treats the *moN-* prefix as marking an antipassive in Bungku-Tolaki languages and Matti (1994) reaches similar conclusions for Mamasa of South Sulawesi.

AV clauses in Sulawesi languages tend to correlate with properties of low semantic transitivity (see S. Andersen & T. D. Andersen 2005: 261-270). For example, in Moronene, the undergoer tends to be indefinite/non-specific in AV, but definite in UV:

(69) *Moronene*

a. **Actor Voice**

Da-hoo nta mong-kea miano.
Be-3SG.ABS FUT AV-bite person
'It will bite someone.'

b. **Undergoer Voice**

Iso tealo kea-‘o yo wontu.
start pass bite-3SG.ABS ART mosquito
'Just then a mosquito passed by and bit him.'

(S. Andersen & T. D. Andersen 2005: 252)

This trend has parallels with Philippine-type languages (SUBSECTION 3.4.1.2). Indeed, AV tends to denote a whole process, whilst UV focuses on a significant act/action within a process. Secondly, AV is associated with durative/habitual aspect, whilst UV typically denotes a punctual event. Thirdly, AV often denotes a less volitional actor than UV. Fourthly, AV is favoured in irrealis contexts, particularly negation and, finally, UV

often has human, animate undergoers, whilst undergoers in AV are typically lower in animacy. For this reason, S. Andersen & T. D. Andersen (2005) suggest that AV emphasises the action, whilst UV emphasises the effect. Hence, UV tends to have properties of higher semantic transitivity (SUBSECTION 3.3.2).

Nonetheless, the AV prefix can also be used in semantically transitive contexts with specific and definite patients:

- (70) *Moronene*
- a. **Actor Voice**
- | | | | | |
|--------------|---------|--------|--------------|----------|
| Ka-i | po-nahu | arumai | ka-i | pong-ka. |
| Then-3SG.NOM | AV-cook | heard | then-3SG.NOM | AV-eat |
- ‘Then she cooked it and ate it.’
- (S. Andersen & T. D. Andersen 2005: 260)
- b.
- | | | | | | |
|-----|------|---------------------|---------|------|-------|
| Hai | hapa | ari-a-u | mo-‘ala | co’o | ana |
| at | what | finish-LOC-2SG.POSS | AV-take | 2SG | child |
-
- | | | | |
|-----------------|------|-----|----------|
| n-tina-‘ate | koie | yo | arataa? |
| LG-woman-little | that | ART | treasure |
- ‘Little girl, where did you get that treasure?’
- (S. Andersen & T. D. Andersen 2005: 260)

In (70a), the undergoer is realised via zero-anaphora, whilst in (70b) the undergoer is modified by both a demonstrative and an article. Both constructions are used to reflect the fact that the undergoer is topical and definite. Hence, though AV has semantic properties of the antipassive in some transitional languages, this is not equivalent to the definiteness restriction in prototypical Philippine-type languages (see SUBSECTION 3.4.1.2, S. Andersen & T. D. Andersen 2005: 251).

The same is true of West Coast Bajau in Sabah (cf. Miller 2007). In a corpus of eight narrative texts, almost all UV clauses contained both a specific actor and a specific undergoer. In contrast, the results for AV were split. Roughly half of the clauses had the characteristics of an active clause, with a specific actor and undergoer.

However, the other half had undergoers that were either syntactically oblique, non-specific or indefinite (Miller 2007: 227). Hence, Miller (2007) concludes that West Coast Bajau AV is unlike an antipassive.¹⁷⁴

3.4.3.3 Discourse

As for discourse tests, they support an analysis of UV clauses as basic, but give mixed results for AV, which sometimes appears active, and sometimes antipassive-like (cf. Quick 2005, S. Andersen & T. D. Andersen 2005). This can be illustrated for Pendau, Central Sulawesi. Quick (2005: 236) found that UV clauses are slightly more frequent in a corpus of narratives, with 243 UV clauses to 200 AV clauses. He also found that UV clauses overwhelmingly have the topicality patterns of active clauses using the metric of Referential Distance (RD):

Table 3.12 Referential Distance in Pendau (Quick 2005: 230-231)

RD	AV actor	AV undergoer	UV actor	UV undergoer
1-3 (High topicality)	80-90%	28-57%	82-96%	67-74%
>3 (Low topicality)	10-20%	43-72%	4-18%	26-33%

Both actors and undergoers tend to be high in topicality in UV. As for AV, the actor tends to have high topicality. However, the discourse status of the undergoer varies. In some texts/sentences, it has low topicality, which is typical of antipassives. However, in other contexts, the undergoer is highly topical. This suggests that AV

¹⁷⁴ Miller (2007: 18) analyses West Coast Bajau as Indonesian-type on the basis that it has several Indonesian-type properties. This includes a true passive and two symmetrical voice alternations. I include it under the heading of ‘transitional’ languages in order to facilitate comparison with other languages in Borneo and Sulawesi. It is an interesting comparison, as West Coast Bajau is said to resemble the languages of Sarawak more than those of Sabah, which tend to be more proto-typically Philippine-type.

sometimes has the discourse properties of an active clause and sometimes an antipassive, which supports the idea of transition.¹⁷⁵

Slightly different results are found for West Coast Bajau, where AV clauses are most frequent. Like other Sama-Bajau languages, West Coast Bajau has three voices: AV, UV and passive (Miller 2014). In a corpus of eight narrative texts, 52% of semantically transitive clauses were AV, 30.7% were UV and 17.3% were passive (Miller 2007: 226). This supports an analysis of AV as the basic clause type. However, when only transitive clauses with specific actor and undergoers are considered, the percentage of AV and UV clauses is roughly equal: 42.8% AV, 43.6% UV and 13.6% passive (19/140) (Miller 2007: 226). In other words, whilst almost all of the UV clauses contain two specific arguments, AV clauses often do not have a specific undergoer. Hence, discourse and semantic tests give conflicting results for the status of AV.

In summary, the languages of Middle Borneo and Sulawesi have a variety of different morphosyntactic features that are neither proto-typically Philippine-type nor proto-typically Indonesian-type. This suggests that a two-way typology may not be sufficient to capture the full extent of variation within Western Austronesian. Moreover, when the voice systems are compared at semantic and discourse levels, some interesting findings arise. UV tends to have the semantic and discourse characteristics of a basic transitive clause, like in Philippine-type languages. However, AV sometimes has the semantic and discourse characteristics of an antipassive, but sometimes the properties of an active. This can be summarised as follows:

¹⁷⁵ See S. Andersen & T. D. Andersen (2005: 256) for discourse measures in Moronene.

Table 3.13 Transitional Voice Systems

Level of analysis	AV transitivity	UV transitivity
Morphology	✓	✓
Syntax	✓	✓
Semantics	✓ (?)	✓
Discourse	✓ (?)	✓

Hence, transitional languages support the Aldridge (2011) view of alignment shift in that they provide evidence of an intermediate stage between Philippine-type and Indonesian-type languages, which will also be seen in Kelabit (SUBSECTION 3.5.3).

3.4.4 Summary

In this section, I analysed a selection of Western Austronesian languages using the methodology outlined in SUBSECTION 3.3. I argued that Philippine-type and Indonesian-type languages not only differ in their structural properties, but also in the semantic and discourse functions of the different voices. In Philippine-type languages, UV is analysed as basic according to semantic and discourse tests, which supports an analysis of these languages as ergative at discourse and semantic levels. In Indonesian-type languages, there is a greater degree of variation. Nonetheless, AV can be analysed as basic in languages like Standard Indonesian, which supports an accusative analysis. Hence, there is some evidence for an alignment shift when the voices are compared at multiple levels of structure.

In addition, I demonstrated that the languages of Sulawesi and Middle Borneo have a range of different morphosyntactic properties in their voice systems, some of which are Philippine-type, some of which are Indonesian-type and some of which are unique to the area. These languages also differ from Philippine-type and Indonesian-type languages at in their semantic and discourse properties. UV generally

retains characteristics of high discourse and semantic transitivity, but AV does not have the proto-typical characteristics of an antipassive. This suggests that classifying languages as either Philippine-type or Indonesian-type overlooks important aspects of variation within Austronesian voice systems that relate to wider theoretical and historical debates. I will now apply the methodology to the voice alternations in Kelabit and consider the implications that this has for the hypothesis of alignment change (Aldridge 2011, 2012).

3.5 Kelabit Voice

As discussed in SUBSECTION 2.2.1, the languages of Sarawak and Central Borneo lie genetically and geographically between Philippine-type and Indonesian-type languages (Hudson 1994, Clayre 2005, FIGURE 2.1). They consequently provide a unique opportunity to explore Aldridge’s (2011) theory of alignment shift. In this section, I analyse Kelabit voice using the methodology outlined in SUBSECTION 3.3 and compare the results with the findings in SUBSECTION 3.4.

3.5.1 Morphosyntax

Kelabit has three voices: AV, UV and IV (see CHAPTER 2). The alternations are marked morphologically and the voice markers are multifunctional, also encoding information about tense, aspect and mood (see SUBSECTION 2.4.1 for discussion of function and allomorphs). The system of voice markers is summarised in TABLE 3.14:

Table 3.14 Kelabit Voice Markers

	Actor Voice	Undergoer Voice	Instrumental Voice
Realis/Perfective	<i>neN-</i> (<i>ne-</i> <i>-um-</i>)	<i>-in-</i>	<i>peneN-</i>
Irrealis	<i>N-</i> (<i>-um-</i>)	<i>-en</i>	<i>peN-</i>

The alternations were illustrated in SUBSECTION 2.5.1 and are repeated below:

- (71) *Kelabit*
- a. **Actor Voice**
 La'ih sineh ne-nekul nubaq nedih ngen seduk.
 man DEM PFV-AV.spoon.up rice 3SG.POSS with spoon
 'That man spooned up his rice with a spoon.'
- b. **Undergoer Voice**
 Sikul lai'h sineh nubaq nedih ngen seduk.
 UV.PFV.spoon.up man DEM rice 3SG.POSS with spoon
 'That man ate his rice with a spoon.'
- c. **Instrumental Voice**
 Seduk pe-nekul la'ih sineh nubaq nedih.
 Spoon IV-spoon up man DEM rice 3SG.POSS
 'That man used a spoon to spoon up his rice.'
- (elicitation, fieldnotes)

The voices in (71) are equally morphologically marked – as summarised in TABLE 3.14 - and equally transitive, with two core arguments – *la'ih sineh* 'that man' and *nubaq nedih* 'his rice'. The arguments are expressed as NPs, rather than oblique PPs, and share a number of core argument properties in each of the three voice constructions (see SUBSECTION 2.5.1). Hence, the voice alternations are morphologically and syntactically symmetrical, like Cebuano and Javanese, and morphosyntax does not give any strong arguments for treating either AV or UV as basic.

Nonetheless, there are two ways in which the voices do not appear entirely syntactically symmetrical, much like Balinese. Firstly, FORM 2 pronouns are typically used in UV, and never in AV constructions, which is discussed further in CHAPTER 4. Secondly, AV and UV are subject to different word-order constraints, which is discussed in CHAPTER 5. For now, I focus on word order since it is pertinent to the discussion in this chapter. A common property of non-subject core arguments is that they appear directly following the verb (SUBSECTION 2.5.1.2):

(72) *Kelabit Word Order*

a. **Actor Voice**

Tesineh nedih [ne-nge-laak nubaq].
mother 3SG.POSS [PFV-AV-cook rice]
'Her mother cooked rice.'

b. [Ne-nge-laak nubaq] tesineh nedih.
PFV-AV-cook rice mother 3SG.POSS
'Her mother cooked rice.'

c. **Undergoer Voice**

[L<in>aak tesineh nedih] nubaq.¹⁷⁶
<UV.PFV>cook mother 3SG.POSS] rice
'Her mother cooked rice.'

d. Nubaq [l<in>aak tesineh nedih].
rice <UV.PFV>cook mother 3SG.POSS
'Her mother cooked rice.' (elicitation, fieldnotes)

In UV, the predicate and non-subject core argument always form a tight constituent and nothing – not even the subject – can intervene. In AV, however, the order verb-actor-undergoer is also possible (see SUBSECTION 5.5.1.2):

(73) *Kelabit Word Order*

a. **Actor Voice**

Ne-nge-laak tesineh nedih nubaq.
PFV-AV-cook mother 3SG.POSS rice
'Her mother cooked rice.'

b. **Undergoer Voice**

*L<in>aak nubaq tesineh nedih.
<UV.PFV>cook rice mother 3SG.POSS
For: 'Her mother cooked rice.' (elicitation, fieldnotes)

Hence, word order is more flexible in AV than UV as the post-verbal subject can appear both before and after the non-subject core argument. In this sense, AV clauses are like intransitive clauses, where a post-verbal subject can either appear clause-finally,

¹⁷⁶ Some speakers suggest that differences in word order may reflect differences in illocutionary force in that (72c) is interpreted as a question and (72d) as a statement. This remains to be further explored.

following any obliques, or intervene between the verb and the oblique (see SUBSECTION 5.4.2 for similar patterns in Balinese):

(74) *Kelabit Intransitive Clauses*

a. **Subject-Final**

Tudo [luun asuq] uih.
sit on stool 1SG.1
'I'm sitting on the stool.'

b. **Subject Post-verbal**

Tudo uih [luun asuq].
sit 1SG.1 on stool
'I sit on the stool.'

(elicitation, BAR18082014CH_01 00:59:09.950-00:59:14.550)

One possible interpretation is that the undergoer of an AV construction is less core than the actor of a UV construction, since the latter is obligatorily part of a VP, whilst the former can optionally appear in the post-subject position. This could support an analysis of Kelabit AV having developed from an antipassive, since it shares structural properties with intransitive clauses. An alternative is to treat the UV actor as incorporated, like the Quechua passive in SUBSECTION 3.2.1.1. However, UV does not have the semantic or discourse properties of a passive. In any case, the AV undergoer follows immediately after the predicate in the majority of cases, and speakers generally prefer verb-undergoer orders (SUBSECTION 5.5.1.2). Hence, morphology and syntax, though perhaps hinting at potential historical developments, tend to support a symmetrical analysis of voice alternations in Kelabit. This brings us to semantics and discourse.

3.5.2 Semantics

In the following two subsections, examples of Kelabit voice alternations are drawn from a traditional story, *Dayang Beladan*, which was recorded in November 2013 (see

APPENDIX 3). The story consists of 193 clauses, of which 136 contain a verbal predicate. This particular story and genre were chosen in order to be roughly comparable with previous studies relating to Philippine-type and Indonesian-type languages (see SUBSECTION 3.4.1.3 for Cebuano, and SUBSECTION 3.4.2.3 for Indonesian). Using Hopper & Thompson's (1980) transitivity parameters, the following differences in semantic transitivity were found. The clauses analysed include those where the predicate could theoretically refer to a semantically transitive event, even when only one argument was overtly encoded in the clause:¹⁷⁷

Table 3.15 Kelabit Semantic Transitivity

	Count	Mean Score	Standard Deviation	Min Score	Max Score
AV	51	6.63	2.13	3	10
UV	18	9.44	0.62	8	10

TABLE 3.15 suggests that UV clauses are higher in semantic transitivity than AV clauses. Indeed, AV clauses are on average 2.81 points lower on the transitivity scale from 0 – 10 and this is a statistically significant result ($t = -5.52, p < .000$). Moreover, AV clauses have a higher range of transitivity values: some AV clauses have the maximum semantic transitivity value, whilst others are very low on the scale. In contrast, UV clauses are consistently high in semantic transitivity. This is in keeping with an ergative analysis in which the UV clause is the basic transitive type. However, there is contradictory evidence in terms of discourse frequency in SUBSECTION 3.5.3.

¹⁷⁷ This does not include 8 examples of predicates with AV morphology that encode intransitive predicates (see TABLE 3.16). Importantly, there is still a significant difference in the mean transitivity of AV and UV clauses when only clauses with two overt participants are compared. In this context, the mean transitivity of AV is 7.77, compared with UV 9.47 ($t = -3.898, p < .000$). All figures are rounded to 2 decimal places.

The high semantic transitivity of UV clauses can be illustrated with examples from the text:

- (75) *Kelabit UV in foreground*
- Nalap neh pupuq.
 UV.PFV.fetch 3SG.2 hitting.implement
 ‘She fetched something to hit with.’
- Nukab neh bubpuq daan.
 UV.PFV.open 3SG.2 door hut
 ‘Opened the door to the hut.’
- Nalap neh dteh kayuh.
 UV.PFV.fetch 3SG.2 one stick
 ‘Picked up a piece of wood.’
- Nulin neh kuyad sineh.
 UV.PFV.throw 3SG.2 monkey DEM
 ‘And threw it at the monkey.’

(text, PDA10112013CH_01 00:08:45.260-00:08:53.220)

Much like other Western Austronesian languages, the UV clauses in (75) are foregrounded and express key actions in the storyline. They all pertain to a highly identifiable and topical actor, *Dayang Beladan*. The final clause also refers to a highly identifiable undergoer, *kuyad sineh* ‘that monkey’. Thus, the clauses are associated with telic, punctual action of an individuated actor on an individuated undergoer. Hence, UV scores highly in terms of Hopper & Thompson’s (1980) transitivity parameters.

The lower semantic transitivity of Kelabit AV might suggest that it has semantic correlations with the antipassive, like in Philippine-type languages (SUBSECTION 3.4.1.2). Many examples in the text do seem to have the semantics of antipassives, such as low individuation and identifiability of the undergoer (Cooreman

1994, Dryer 1990, Mithun 2000). However, other examples have the semantics of active clauses. The breakdown is shown in TABLE 3.16:

Table 3.16 *Kelabit Actor Voice Semantics*

		count	percentage
Antipassive-like	Intransitive	8	13.56%
	Transitive: zero, unidentifiable undergoer	17	28.81%
	Transitive: indefinite object	10	16.95%
	Transitive: new, identifiable undergoer	4	6.78%
	Total	39	66.10%
Active-like	Transitive: definite object	17	28.81%
	Transitive: zero, highly identifiable undergoer	3	5.08%
	Total	20	33.90%

The nasal prefix *N-* that marks AV constructions in Kelabit is used to form intransitive predicates such as *nalan* ‘walk’ (SUBSECTION 2.4.1.1.4). Aldridge (2012) argues that this is in keeping with an antipassive analysis, since antipassives are formally intransitive.¹⁷⁸ For verbs that could be syntactically transitive, 31/51 have an undergoer that is null and unidentifiable or indefinite or discourse new. This is illustrated by the following examples:

(76) *Kelabit AV as antipassive*

a. **Intransitive**

(*N-* + *dalan* ‘road’ → *nalan* ‘walk’)

Lem edteh edto Dayang Beladan nalan~nalan lem kebun
 one day Dayang Beladan REDUP~AV.walk in garden

nedih.

3SG.POSS

‘One day, Dayang Beladan was walking around in her garden.’

(text, PDA10112013CH_01 00:00:30.510-00:00:38.030)

¹⁷⁸ However, Kaufman (to appear) and Foley (2008) suggest this polysemy is rare.

b. **Transitive, zero unidentifiable object**

(*N-* + *dinger* → *ninger* ‘listen’)

Am ieh tidih.

NEG 3SG.1 PT=present

‘It [the stolen jaw harp] wasn’t there.’

[Adiq nieh ninger [no object] keyh].
so PT=3SG.1 AV.hear PT

‘So she listened.’

Nangey teh unih ih ngi ditaq.
over.there PT sound PT at high

‘And there was the sound of it coming from up high.’

(text, PDA10112013CH_01 00:04:59.890-00:05:04.180)

c. **Transitive, indefinite object**

(*N-* + *taruq* → *naruq* ‘do’)

Ken ngudeh teh ngaley sineh murih kuman pudo ba’ung neh?

Q why PT marten DEM often AV.eat ripe bananas DEM

‘Oh why does this yellow-throated marten keep eating those ripe bananas?’

[Dooq tuih naruq edteh ebpung].
good PT=1SG.1 AV.do one trap

‘I’d better make a trap.’

(text, PDA10112013CH_01 00:01:11.070-00:01:17.310)

d. **Transitive, identifiable new object**

(*N-* + *sipa* → *nipa* ‘pack’)

Rengaq idih ngaley sineh nipa uluh nedih keleyh...

when DEM marten DEM AV.pack head 3SG.POSS PT

‘As soon as the yellow-throated marten put his head [into the trap]...’

(text, PDA10112013CH_01 00:01:53.090-00:01:55.900)

In (76a), AV morphology is used to derive an intransitive predicate with no object. In (76b), no object is expressed because the object is generic and unidentifiable. Moreover, in both instances the action of the actor is more important than any effect it might have on the undergoer. This is characteristic of the antipassive (Cooreman 1994). In (76c) and (76d) there is an overt object encoded. However, the object is newly introduced into the discourse. In (76c), the object, *edteh ebpung* ‘a trap’, had

not been mentioned previously in the discourse and is introduced with the indefinite numeral, *edteh* ‘one’. In (76d), the object is more identifiable, in that it is expressed with a possessive pronoun *uluh nedih* ‘his head’. This perhaps follows from our real world knowledge that animals, like the yellow-throated marten, tend to have body parts. However, the marten’s head specifically had not previously been mentioned. Therefore, these instances, which collectively constitute roughly 66% of the cases of AV, seem to have the semantics of antipassives.

Nonetheless, there are 17 examples in this text alone in which the undergoer is given and encoded as definite. Indeed, there are a few cases of zero-anaphora where the object is both specific and highly identifiable from the immediate discourse (cf. Himmelmann 1999). These are illustrated in (77):

- (77) *Kelabit AV as transitive*
- a. **Transitive, definite object**
(N- + puwer → muwer ‘butcher’)
 Neh nieh muwer **ieh.**
 DEM PT=3SG.1 AV.butcher 3SG.1
 ‘And she butchered it [the yellow-throated marten].’
 (text, PDA10112013CH_01 00:02:13.960-00:02:16.340)
- b. **Transitive, zero identifiable object**
(N- + tekap → nekap ‘search’)
 Edteh teh lemulun raut ruding neh.
 one PT person play jaw.harp 3SG.2
 ‘Someone was playing her jaw harp.’
- Ni’er ruding, am teh ruding idih lem tidtuq nedih.
 AV.see jaw.harp NEG PT jaw.harp present in hand 3SG.POSS
 ‘She looked for the jaw harp but the jaw harp wasn’t in her hands.’
- [Nekap~nekap **[no object]** luun tanaq].
 REDUP~AV.search on ground
 ‘She looked everywhere.’

Am ieh tidih.
NEG 3SG.1 PT=present
'It wasn't to be found.'

(text, PDA10112013CH_01 00:04:47.090-00:04:59.010)

In (77a), the undergoer is expressed as a pronoun, which suggests high identifiability (cf. Cooreman 1994). Similarly, it is perfectly grammatical for undergoers in AV to be modified with possessive pronouns, demonstratives and definite markers. In (77b), the object is not expressed. However, it is understood that she is looking for the *ruding* 'jaw harp', which is topical in both the previous and subsequent discourse. Hence, it could be analysed as a case of definite null anaphora. Thus, although many AV clauses do seem to share cross-linguistic characteristics of antipassives with their Philippine-type equivalents, roughly 33% of the instances do not.

Consequently, although Kelabit UV appears high in semantic transitivity, Kelabit AV is not exactly the same as its Philippine-type equivalents since the constraint against definite undergoers is more of a tendency than a strict rule. Like Moronene, this supports an analysis of diachronic change, whereby Kelabit AV developed from an earlier antipassive, and retains some of the semantics, but has been reanalysed as an active clause-type. Hence, looking at semantic evidence reveals two interesting facts: firstly, that the alternations may not be as symmetrical as they appeared from the morphosyntax and secondly, that Kelabit voice alternations differ from the more conservative Philippine-type system. Thus, analysis of semantic transitivity suggests that UV is the basic clause type in Kelabit, but that Kelabit AV has moved away from the Philippine-type system, with some clauses retaining the semantics of an antipassive, and others developing the semantics of an active clause.

3.5.3 Discourse

In this section, the frequency and topicality of the voices in Kelabit are compared with studies of Cebuano and Indonesian, described in SUBSECTION 3.4. Of course, there are some differences, both in the time period in which the stories were documented (1960s-2013) and the medium in which they were first produced (oral versus written). However, the length and genre of the three pieces is roughly comparable and this will suffice for the purposes of this comparison.

The first comparison is discourse frequency. Only voice-marked and syntactically transitive clauses (i.e. those with two overt arguments) are compared. In Cebuano, UV clauses were the most frequent (see SUBSECTION 3.4.1.3). However, in both Indonesian and Kelabit, AV clauses are more frequent than UV (see SUBSECTION 3.4.2.3):

Table 3.17 Frequency of Voices

	Cebuano	Kelabit¹⁷⁹	Indonesian
Total voice marked clauses	32	50	51
Total AV	8	31	38
Total UV	24	17	13
Percentage AV	25%	62%	75%
Percentage UV	75%	34%	25%

In contrast to the evidence from semantics, discourse frequency groups Kelabit with Indonesian-type rather than Philippine-type since this particular measure suggests that AV clauses are more basic.

¹⁷⁹ Regardless of how they are compared, Kelabit AV clauses are considerably more frequent than UV. If all potentially transitive and voice-marked clauses are counted, then the difference is higher at 71.83% AV to 25.35% UV. However, since similarly ambiguous sentences were removed from the Indonesian and the Cebuano data, they are excluded above for ease of comparison. The total 50 in TABLE 3.17 includes 2 IV-marked clauses.

The most interesting result of all, however, relates to relative topicality of arguments. Following the metrics outlined in SUBSECTION 3.3, the topicality of arguments in Kelabit voices are displayed in TABLE 3.18.

Table 3.18 Topicality of Arguments in Kelabit

	Actor RD	Actor TP	Mean Actor	Undergoer RD	Undergoer TP	Mean Undergoer
AV	2.16	2.29	0.80	8.58	1.48	0.48
UV	2.82	3.82	0.89	7.41	1.35	0.46

TABLE 3.18 reveals that both AV and UV have exactly the same patterns of topicality for actor and undergoer. If we compare this with the expected patterns in TABLE 3.2, then these seem to be the patterns predicted for active and ergative clauses. Hence on a discourse level, at least in terms of topicality, both AV and UV appear to be transitive – suggesting some form of discourse symmetry. This is particularly interesting when compared with Cebuano and Indonesian:

Table 3.19 Cross-linguistic Topicality

	AV		UV	
	Mean Actor	Mean Undergoer	Mean Actor	Mean Undergoer
Cebuano	0.41	0.18	0.89	0.18
Kelabit	0.80	0.48	0.89	0.46
Indonesian	0.52	0.35	0.26	0.52

Cebuano has the expected patterns of an ergative language. UV clauses appear active with a highly topical actor and an undergoer with lower topicality. AV clauses, on the other hand, have the patterns of an antipassive, in that the actor is much lower in topicality. Indonesian, in contrast, has the expected patterns of an accusative language. AV clauses more or less have the topicality patterns of an active clause, and UV looks

more like a passive or inverse, as the undergoer is higher in topicality than the actor.¹⁸⁰ Thus, the topicality patterns support an analysis of Cebuano as ergative and Indonesian as accusative, on semantic and discourse grounds.

The Kelabit results, like Pendau, seem to support the idea of a transition between the two, as both AV and UV have the topicality patterns of a basic active clause. In fact, discourse evidence seems to suggest that the transition from ergative to accusative begins with the reanalysis of antipassive as active, as in Kelabit, and then with the reanalysis of ergative to passive, as in Indonesian. Hence comparing the voices on a discourse level supports Aldridge's (2011) view of Philippine-type and Indonesian-type languages undergoing a shift from ergative to accusative alignment, since both AV and UV have the topicality patterns of active/ergative clauses.

3.5.4 Summary

This section presented a detailed study of the morphosyntactic, semantic and discourse characteristics of Kelabit voice. From a morphosyntactic perspective, Kelabit appears to have symmetrical voice alternations, much like proto-typical Philippine-type and Indonesian-type languages (see SUBSECTION 1.3). Moreover, as shown in CHAPTER 2, Kelabit has some structural similarities with Philippine-type languages. These are summarised in TABLE 3.20:

¹⁸⁰ The topicality of the Indonesian AV actor is perhaps less than expected (see TABLE 3.2). This could reflect development from an antipassive-type clause.

Table 3.20 *Kelabit Structural Properties*

	Indonesian Type	Philippine Type	Kelabit
Symmetrical alternations	Y	Y	Y
True passive	Y	N	N
Applicative suffixes	Y	N	N
Micro roles with voices	N	Y	Y
Mood marking morphology	N	Y	Y
Case marking	N	Y	?

Kelabit shares five of the six parameters with Philippine-type languages, and only one definitively with Indonesian-type. It does not have a true morphological passive or applicative suffixes, but does have a voice for the instrument micro-role and mood-marking morphology (see TABLE 3.14). Moreover, as discussed in SUBSECTION 2.5.1.1, though there is no case-marking per se, the particles *neh* and *teh* share some similarities with Tagalog *ang*-marking. Hence, the sixth parameter is inconclusive. Thus, if we were comparing only Arka's (2002) structural properties, we might well classify Kelabit, like Lundayeh, as Philippine-type (cf. Clayre 2005).

However, when we compare voice alternations on multiple levels of structure, a more interesting picture emerges. In Kelabit, UV employs the PAn voice marker *-in-* and has many similarities with its Philippine-type equivalent. In particular, it is the most basic transitive clause-type on semantic grounds and has the topicality patterns of an active/ergative clause. Kelabit AV, in contrast, is much more like its Indonesian-type equivalent. It uses the innovative nasal prefix, except in *kuman* 'to eat', and has several properties of an active clause, such as being the most frequent in discourse. Nonetheless, some residue of antipassive semantics remains in the tendency towards less individuated and identifiable undergoers, also seen in Indonesian-type languages, though this is no longer an outright constraint, as in the Philippines. Consequently, Kelabit may be considered even more symmetrical than proto-typical

Philippine-type or Indonesian-type languages, as these show clear discourse and semantic motivations for treating one voice as more transitive, whilst Kelabit voices also appear symmetrical in terms of the discourse topicality of arguments. This can be summarised as follows:

Table 3.21 Voice in Kelabit

Level of analysis	AV transitivity	UV transitivity
Morphology	✓	✓
Syntax	✓	✓
Semantics	✓ (?)	✓
Discourse	✓	✓ (?)

Of course, these findings are only preliminary and are based on a single text and a single genre. There could almost certainly be improvements made in refining which methods are used to compare clauses on semantic and discourse levels and extending this sort of analysis to a wider range of languages and genres. Nonetheless, it has some important implications for the typology of Western Austronesian and the theory of alignment shift.

3.6. Conclusion

In conclusion, in this chapter I addressed two key questions: firstly, whether the two-way typology of Philippine-type and Indonesian-type is sufficient to capture the variation in Western Austronesian voice systems; and secondly, whether there is any evidence to support a transition from ergative to accusative alignment. In order to answer these questions, I analysed the voice system of Kelabit in terms of morphology, syntax, semantics and discourse, and compared it to voice systems in Philippine-type languages, Indonesian-type languages and transitional languages in Borneo and

Sulawesi. In particular, I explored how the different levels of comparison provided arguments for treating different voices as basic.

Applying a fine-grained, parametric approach to voice in Kelabit produced some interesting results, as shown in SUBSECTION 3.5.4. Firstly, it showed the importance of considering semantics and discourse when analysing Western Austronesian alignment, since analysing the morphosyntax of the voice alternations alone fails to reveal asymmetries at other levels of structure. Secondly, it showed that voices can be treated as more or less basic within a single language, depending on the basis of comparison. Finally, it revealed that Western Austronesian languages vary in the morphosyntactic, semantic and discourse properties associated with each voice. Indeed, there is not only variation between Philippine-type and Indonesian-type languages, as might be expected given the prevailing two-way typology, but also within languages that ostensibly belong in the same typological category. Hence, in answer to the first question, a binary distinction between Philippine-type and Indonesian-type seems too narrow to capture the range of variation within Austronesian voice systems.

This leads to the question of alignment. In Philippine-type languages, it is UV that tends to be basic, whilst AV has the semantic and discourse properties of an antipassive. This supports an analysis of the alignment as ergative, at least on semantic and discourse levels. In Indonesian-type languages, there is a greater degree of variation but at least some languages have semantic and discourse properties associated with accusative alignment, where AV is active, and UV passive. As for Kelabit, and a number of transitional languages in Sulawesi and Borneo, UV has the discourse-semantic characteristics of an ergative clause. However, AV is more like its

active Indonesian-type equivalent, despite a tendency towards using AV with non-individuated undergoers, just like antipassives.

What this seems to imply is that an earlier semantic or discourse antipassive is in the process of being reanalysed as an active clause and developing the appropriate semantic interpretations and discourse features as a result. It perhaps also implies that this sort of large-scale typological change begins at a discourse level, with a change in frequency (see Du Bois (1987) for similar discussions). Thus, this study supports Aldridge's (2011) proposal that Austronesian languages are undergoing a shift from ergative to accusative by providing some evidence of an intermediate stage at which the semantics of ergativity remain in UV, whilst the actor voice has developed discourse, and to a lesser extent, semantic properties of active/transitive clauses. However, this shift is not morphosyntactic – as in the canonical understanding of alignment - but rather occurs at a discourse and semantic level.

Consequently, a better way of thinking of the typology of Western Austronesian voice is to consider the degrees of symmetry at the levels of morphology, syntax, semantics and discourse. As outlined in SUBSECTION 3.3, this can be used to position a language on the alignment spectrum from ergative to accusative (see also Arka 2002). In doing so, we can recognise at least three important groups:

- (78) *A New Typology of Western Austronesian Voice*
- a. **Discourse ergative-type**
UV has discourse/semantic characteristics of ergative/transitive
AV has discourse/semantic characteristics of antipassive
 - b. **Discourse intermediate-type**
UV has discourse/semantic characteristics of ergative/transitive
AV has discourse/semantic characteristics of active/transitive
 - c. **Discourse accusative-type**
AV has discourse/semantic characteristics of active/transitive
UV has discourse/semantic characteristics of passive

The discourse intermediate languages have the greatest degree of symmetry, since they can be symmetrical even beyond the level of morphosyntax. Extending this sort of parametric approach to other Western Austronesian languages may well reveal further distinctions that need to be captured in order to better understand the extent of variation and the mechanisms of diachronic change (see CHAPTER 6). In the next chapter, I address the validity of claiming that Kelabit cannot be classified as either Philippine-type or Indonesian-type by exploring another structural property that is said to differ between the two major classes: the pronoun systems.

Chapter 4

Pronominal Systems

4.1 Introduction

In the previous chapter, I argued that Kelabit voice alternations – like many Western Austronesian languages – can be considered morphosyntactically symmetrical, in the sense that both actor voice (AV) and undergoer voice (UV) are equally morphologically marked and contain two core arguments (see SUBSECTION 1.3). I then compared the Kelabit voice alternations at morphological, syntactic, semantic and discourse levels with Philippine-type languages, Indonesian-type languages and transitional languages in Central Borneo and parts of Sulawesi. Unlike Philippine-type languages, which could be considered ‘ergative’ on semantic and discourse grounds, Kelabit AV did not have the characteristics of an antipassive. Unlike Indonesian-type languages, some of which could be considered ‘accusative’ on semantic and discourse grounds, Kelabit UV did not have the characteristics of a passive. Consequently, I concluded that the two-way typology of Philippine-type and Indonesian-type is not sufficient to capture the syntactic variation within Western Austronesian and that a group of languages, including Kelabit, had to be recognised that differed not only in observable surface differences, but also at a deeper level in terms of their alignment systems.

In this chapter, I present further evidence for the idea that Western Austronesian languages have undergone a change from ergative to accusative alignment, and the idea that the two-way typology cannot capture the full extent of variation, by exploring the pronominal system in Kelabit. As seen in SUBSECTION 2.4.2.8, Kelabit has two sets of basic pronouns and their distribution depends partly on the voice construction. Related forms in the Philippines and Northern Borneo have previously been analysed as marking different cases or relating to the function of the pronouns within the voice system (Clayre 2005, Soriente 2013). Accordingly, the pronouns have been used to support analyses of Philippine-type languages as either symmetrical or ergative (SUBSECTION 4.2.1). However, the Kelabit pronouns do not fit nicely within either a case-based analysis or a voice-based analysis and seem to represent an asymmetry in the Kelabit voice system (SUBSECTION 4.2). Consequently, this chapter explores an alternative idea that the difference between the pronouns may be largely prosodic rather than syntactic in nature.

This leads to an exploration of whether the variant pronouns are prosodically weak or clitic forms and a comparison of Kelabit clitics with the widespread clitic phenomena in other Western Austronesian languages (SUBSECTION 4.3.2). As argued in SUBSECTION 1.3.1, clitic phenomena constitute another observable difference between Philippine-type and Indonesian-type languages. Ultimately, it transpires that Kelabit does have clitic pronouns, but clitics that differ from proto-typical Philippine-type and Indonesian-type systems. Hence, the analysis of pronominal systems supports the conclusion that a two-way typology of Western Austronesian is not sufficient, both in terms of the differences in case-marking and in terms of the differences in clitic systems.

The chapter is structured as follows. SUBSECTION 4.2 surveys Kelabit pronouns and illustrates the limitations of case-based and voice-based analyses; SUBSECTION 4.3 defines the notion of ‘clitics’ and provides an overview of pronominal clitic phenomena in Western Austronesian; SUBSECTION 4.4 presents the methodology used to establish whether the Kelabit pronouns function as clitics; SUBSECTION 4.5 presents the results; and SUBSECTION 4.6 discusses the implications for the syntax-prosody interface and Western Austronesian typology.

4.2 Kelabit Pronouns

The pronominal system in Kelabit is highly complex and demonstrates singular, dual, paucal and plural number distinctions; inclusive and exclusive oppositions and an impersonal pronoun *narih* (SUBSECTION 2.4.2.8). The basic pronouns were illustrated in TABLE 2.11 and are repeated in TABLE 4.1:

Table 4.1 Kelabit FORM 1 pronouns

	1.INCL	1.EXCL	2	3
SINGULAR		uuh	iko	ieh
DUAL	kiteh	kediwéh	meduwéh	diwéh
PAUCAL	teluh	keteluh	meteluh	deteluh
PLURAL	tauh	kamih	muyuh	ideh

In addition, there is a reduced paradigm of variant forms in the 1SG, 2SG, 3SG and 3PL, which have very little morphologically or phonologically in common with their FORM 1 counterparts, apart from the third plural form:

Table 4.2 Kelabit FORM 2 pronouns

	FORM 1	FORM 2
1SG	uuh	kuh
2SG	iko	muh
3SG	ieh	neh
3PL	ideh	deh

The FORM 2 pronouns in TABLE 4.2 can be considered a paradigm as they share behavioural characteristics. For example, they combine with the preposition *ngen* to mark obliques (SUBSECTION 2.4.2.8):¹⁸¹

Table 4.3 Kelabit Oblique Pronouns

	FORM 1	FORM 2 (clitic)
1SG	ngen uuh	ngekuh
2SG	ngen iko	ngemuh
3SG	ngen ieh	ngeneh
3PL	ngen ideh	ngedeh

Clitic oblique pronouns are only possible with the reduced set of variant pronouns in TABLE 4.2. They do not occur for any of the remaining pronouns in the paradigm in

TABLE 4.1:

- (1) a. *ngen kiteh* ‘to 1DU.INCL’
 **ngekiteh*
- b. *ngen teluh* ‘to 1PAU.INCL’
 **ngeteluh*
- c. *ngen tauh* ‘to 1PL.INCL’
 **ngetauh*

¹⁸¹ The same pattern is observed in combination with the verb/particle *ken* producing forms such as *kekuh* ‘I say’, *kemuh* ‘you say’, *keneh* ‘he says’ and *kedeh* ‘they say’.

Moreover, variant forms in these four person and number combinations are common in the languages of Borneo, including Penan Benalui, Punan Tubu', Kenyah Òma Lóngh and Kenyah Lebu' Kulit (Soriente 2013). Consequently, the FORM 2 pronouns in TABLE 4.2 form a separate paradigm. The rest of this section concerns how best to analyse the variant pronominal forms.

4.2.1 Previous analyses

4.2.1.1 Case-based Analyses

Pronouns in Northern Borneo and the Philippines are commonly thought to represent different cases (Lobel 2013, Kroeger 2005, Billings & Kaufman 2004 among others). Indeed, case distinctions are reconstructed for earlier stages in Austronesian prehistory, such as Proto-Southwest Sabah in Lobel (2013):

Table 4.4 Proto-Southwest Sabah (Lobel 2013: 103)

	NOM	GEN	OBL
1SG	*aku	*=ku	*d[i]-ak(əi)(n?)
2SG	*(əi)-ka[w], *=kə	*=mu, *=nu	*d[i]-iyun
3SG	*[s]iə	*=yə, *=nə, *nyə	*di[si]ə
1DU.INCL	*[k]itə	*=tə	*d[i]-at(əi)(n?)
1PL.INCL	*[ki]ta-kau	*=ta-kau	*di-ta-kau
1PL.EXCL	*ə-kai	*=mai	*d[i]-am(əi)(n?)
2PL	*ə-kau, *=kau	*=muyu[n]	*d[i]-amuyu[n]
3PL	*[s]idə	*=[ni-]də	*di[si]idə

Typically, the pronouns that correspond to Kelabit FORM 1 are labelled NOM, and the pronouns that correspond to Kelabit FORM 2 are labelled GEN. These labels represent the syntactic function of the pronoun and can equate to a symmetrical analysis of the voice system, whereby NOM marks subjects and GEN marks non-subject core arguments (cf. SUBSECTION 1.4.1). For example, consider Kimaragang, a Dusunic language spoken in Sabah:

- (2) *Kimaragang Dusun*
- a. **Actor Voice**
 Mangalapak okuh do niyuw.
 AV.TR.split 1SG.NOM GEN coconut
 ‘I will split a coconut/some coconuts.’
- b. **Undergoer Voice**
 Lapak-on kuh it niyuw.
 split-UV 1SG.GEN NOM coconut
 ‘I will split the coconut(s).’
- c. **Dative Voice**
 Lapak-an kuh do niyuw it wogok.
 split-DV 1SG.GEN GEN coconut NOM pig
 ‘I will split some coconuts for the pigs.’
- d. **Instrumental Voice**
 Nokuroh.tu n-i-lapak nuh do niyuw inoh
 why PST-IV-split 2SG.GEN GEN coconut MED.NOM

 dangol kuh?
 knife 1SG.GEN
 ‘Why did you use my bush knife to split coconuts?’

(Kroeger 2005)

In the AV clause in (2a), the actor is the NOM subject and the undergoer the GEN non-subject. In the UV clause in (2b), the actor is the GEN non-subject and the undergoer the NOM subject. Hence, the alternations are syntactically symmetrical. The actor non-subject is labelled GEN in non-actor voices, such as (2b), (2c) and (2d), since the same form also marks possession, as shown in *dangol kuh* ‘my bush knife’ in (2d).

Based on the obvious similarities of the GEN forms in Proto-Southwest Sabah, Kimaragang and many other Austronesian languages to the Kelabit pronouns in TABLE 4.2, one hypothesis is that FORM 2 pronouns are a reduced set of genitive forms, whilst the FORM 1 pronouns represent nominative case. However, it will be shown in SUBSECTION 4.2.2 that Kelabit FORM 2 pronouns do not have exactly the same

distribution as GEN pronouns in languages like Kimaragang and hence this analysis is not adopted.

4.2.1.2 Voice-based Analyses

A second analysis that has been proposed particularly for the languages of Borneo, is that the form of the pronouns relates to their function within the voice system (Clayre 2005, Soriente 2013, Miller 2007). In contrast to a case-based analysis, the form of the pronoun does not represent the syntactic function alone. Instead, it reflects a combination of semantic role and syntactic function and is explicitly linked to the voice-construction (cf. Soriente 2013: 181).

The voice-based analysis of pronouns is most clearly articulated in Clayre (1991, 2005) for Lundayeh – a closely-related language to Kelabit (SUBSECTION 2.2.1).¹⁸² Rather than assigning case labels to the pronouns, the voice-based analysis states that FORM 1 is used for subjects, FORM 2 for actor semantic roles that are not in subject function and FORM 3 for non-actor, non-subjects (cf. SUBSECTION 1.4.1 and 2.5.1):

Table 4.5 Lundayeh Pronoun Sets (cf. Clayre 2005: 24)

	Subject	Non-Subject	
	FORM 1 (all roles)	FORM 2 (actor)	FORM 3 (non-actor roles)
1SG	uuh	kuh	nekuh
2SG	iko	mu	nemu
3SG	ieh	neh	neneh
1PL.INCL	tau	tau	netau
1PL.EXCL	kai	kai	nekai
2PL	muyuh	muyuh	nemuyuh
3PL	ideh	deh	nedeuh

¹⁸² See Boutin (1988: 60) for similar discussion in relation to Sabahan languages.

This account follows from the following alternations in Clayre (2005). In each case, the semantic roles of actor (A) or non-actor (N-A) and the grammatical function of subject (S) or non-subject (N-S) is specified:

- (3) *Lundayeh*
- a. **Actor Voice**
- | | | | | |
|-------|----------|---------|-----|---------|
| Iko | nguit | neneh | amé | nekuh. |
| 2SG.1 | AV.bring | 3SG.3 | go | 1SG.3 |
| A/S | | N-A/N-S | | N-A/N-S |
- ‘You bring him to me.’ (Clayre 2005: 25)
- b. **Undergoer Voice**
- | | | | | |
|-------------|-------|-------|------|---------|
| Inapung | kuh | ieh | rat | neneh. |
| UV.PFV.hide | 1SG.2 | 3SG.1 | from | 3SG.3 |
| | A/N-S | N-A/S | | N-A/N-S |
- ‘I hid it from him.’ (Clayre 2005: 25)
- c. **Instrumental Voice**
- | | | | | |
|--------|-------|---------|-------|-------|
| Pimeli | kuh | lal | usin | inih. |
| IV.buy | 1SG.2 | hen | money | this |
| | A/N-S | N-A/N-S | | N-A/S |
- ‘I’ll use this money to buy the hen.’ (Clayre 2005: 21)

On the basis of (3), the contexts in which FORM 1, FORM 2 and FORM 3 are used can be summarised as follows:

Table 4.6 Summarising Lundayeh Pronouns

Role	Actor Voice			Undergoer Voice		
	Actor	Undergoer	Other	Actor	Undergoer	Other
Form	FORM 1	FORM 3	FORM 3	FORM 2	FORM 1	FORM 3

This might support an asymmetrical analysis of voice alternations in Lundayeh. FORM 3 is typically used for obliques, such as the malefactive *rat neneh* in (3b). Hence, AV in Lundayeh appears to have a single core argument and an oblique. UV, in contrast, has two distinct forms, suggesting two core arguments with distinct cases.

Consequently, the voice-based analysis is easily translatable into an ergative analysis of voice in Lundayeh:

Table 4.7 An Ergative Analysis of Lundayeh

	Actor Voice			Undergoer Voice		
Role	Actor	Undergoer	Other	Actor	Undergoer	Other
Case	ABS	OBL	OBL	ERG	ABS	OBL

As discussed in SUBSECTION 3.4.1, ergative analyses have been posited for a number of Western Austronesian languages on the basis of semantics and discourse (see Aldridge 2011, Gerdts 1988, Gault 1999, Nolasco 2005 etc.). The pronominal system in Lundayeh supports an account of morphological ergativity as well. In other words, a voice-based analysis could be reinterpreted as a variant of traditional case-based analyses that assumes a morphologically asymmetrical picture of Austronesian voice, rather than symmetrical alternations.

4.2.2 Extending the analysis to Kelabit

The previous analyses of Austronesian pronouns present two hypotheses to explore, namely that Kelabit FORM 2 pronouns are genitive case-marked or that Kelabit FORM 2 pronouns represent actors in non-subject functions of non-actor voices. A voice-based analysis can be applied to Kelabit, as shown in (4):

- (4) *Kelabit*
- a. **Actor voice**
- | | | | | |
|-----|-------|----------|--------|-----|
| Nih | uih | natek | bubpuq | ih. |
| DEM | 1SG.1 | AV.close | door | PT |
- A/S
- ‘I’m closing the door.’
- (elicitation, BAR17102013CH_01 00:51:09.585-00:51:12.558)

b. **Undergoer voice**

Senatek	kuh	neh	bubpuq	ih.
UV.PFV.close	1SG.2	PT	door	PT
	A/N-S			

‘I already closed the door.’

(elicitation, BAR17102013CH_01 00:53:02.138-00:53:04.854)

c. **Instrumental voice**

Enun	pe-natek	kuh	bubpuq	ih?
what	IV-close	1SG.2	door	PT
	A/N-S			

‘What do I use to close the door?’

(elicitation, BAR17102013CH_01 01:01:13.792-01:01:16.702)

In (4), actor roles in subject function are encoded using FORM 1, whilst actor roles in non-subject functions are encoded with FORM 2. This appears to support a voice-based model. However, there are several ways in which the Kelabit pronouns behave differently from the Lundayeh pronouns in TABLE 4.5. Firstly, FORM 1 can also be used for non-subject functions in actor voice:

(5) *Kelabit*

a. **Actor Voice**

Ieh	ni'er	uih.
3SG.1	AV.see	1SG.1
A/S	N-A/N-S	

‘He sees me.’

(elicitation, fieldnotes)

If we compare (5a) with the Lundayeh equivalent in (3a), it is apparent that Lundayeh uses FORM 3 for non-actor non-subjects, whilst Kelabit uses FORM 1.

Secondly, wherever FORM 2 is found, FORM 1 is also possible as an alternative.

Thus, FORM 1 can alternate with FORM 2 for actor non-subjects in undergoer voice:¹⁸³

¹⁸³ The use of the particle *teh* seems to be obligatory in contexts like (6b). This remains to be further explored.

(6) *Kelabit Differential Marking*

a. **Undergoer Voice, FORM 2**

Seni'er	kuh	ieh.
UV.PFV.see	1SG.2	3SG.1
	A/N-S	N-A/S

'I saw him.'

b. **Undergoer Voice, FORM 1**

Seni'er	uih	tieh (teh+ieh).
UV.PFV.see	1SG.1	PT=3SG.1
	A/N-S	N-A/S

'I saw him.'

(elicitation, fieldnotes)

Moreover, FORM 1 and FORM 2 alternate as ways of marking possession, alongside dedicated possessive pronouns *kudih* and *duih* (SUBSECTION 2.4.2.8.1):

(7) **Possession**

a. rumaq uih
house 1SG.1
'my house'

b. rumaq kuh
house 1SG.2
'my house'

c. rumaq kudih
house 1SG.POSS
'my house'

d. duih rumaq
1SG.POSS house
'my house'

(elicitation, fieldnotes)

Finally, FORM 1 and FORM 2 also alternate with a subset of ambitransitive verbs that take an experiencer argument, such as *keliq* 'know', *dooq pian* 'like', *dooq ileh* 'be able to', *kelupan* 'forget' and *sekenan* 'remember', and predicates with the accidental prefix *ne-* or the abilitative prefix *ke-* (SUBSECTION 2.4.1.1):

- (8) **Experiencer Verbs**
- a. Na'am uih keliq.
NEG 1SG.1 know
'I don't know.'
- b. Na'am keliq kuh.¹⁸⁴
NEG know 1SG.2
'I don't know.' (elicitation, fieldnotes)
- (9) **Accidental Prefix**
- a. Ne-bilaq uih bigan ih.
ACCID-break 1SG.1 plate PT
'I accidentally broke the plate.'
- b. Ne-bilaq kuh neh bigan ih.
ACCID-break 1SG.2 PT plate PT
'I accidentally broke the plate.' (elicitation, fieldnotes)

Hence, FORM 2 is not a unique marker of actor non-subjects, but functions as a form of differential marking. The differences between Kelabit and Lundayeh are summarised in TABLE 4.8:

Table 4.8 Comparing Lundayeh and Kelabit

	Actor voice			Undergoer Voice			Possessor
	Actor	Undergoer	Other	Actor	Undergoer	Other	
Lun-dayeh	FORM 1	FORM 3	FORM 3	FORM 2	FORM 1	FORM 3	FORM 2
Kelabit	FORM 1	FORM 1	FORM 3	FORM 1/ FORM 2	FORM 1	FORM 3	FORM 1/ FORM 2 possessives

Unlike the seemingly ergative system of Lundayeh, the distribution of Kelabit pronouns in TABLE 4.8 suggests a system where core arguments (FORM 1) are

¹⁸⁴ It is possible that these predicates are nominal or precatatorial (see SUBSECTION 2.4.2). The word-order difference between (8a) and (8b) is discussed in SUBSECTION 4.3.3.

distinguished from obliques (FORM 3) and FORM 2 simply functions in certain contexts as an alternative to FORM 1.¹⁸⁵

Consequently, neither a case-based analysis nor a voice-based analysis can account for the variant pronouns in TABLE 4.2. A case-based analysis is ruled out by the fact that the same pronoun is used for multiple syntactic functions, as in (5), and the same function is represented by multiple pronoun forms, as in (6) to (9). A genitive analysis of FORM 2 is not supported in Kelabit since other forms are also used for possession, as shown in (7). Equally, an ergative analysis is ruled out by the fact that the non-subject in the actor voice clause in (5) is marked with the core FORM 1 pronoun rather than the oblique FORM 3. Similarly, the actor non-subject of the undergoer voice is not exclusively marked by a distinct form that could be analysed as ergative, such as FORM 2, but can also be marked in the same way as actor subjects and non-actor non-subjects with FORM 1. Thus, the syntactic function does not predict which form of the pronoun will be used.

Along the same lines, a voice-based analysis is ruled out by the fact that the voice construction does not predict which pronoun form will be used either. Although FORM 2 is restricted to non-actor voices, non-actor voices do not restrict themselves to FORM 2 actors, as shown in (6b). Moreover, alternations are also found in intransitive constructions like those in (8) and accidental clauses like (9) that have no voice-marking whatsoever. Thus voice cannot be the decisive distinction between the variant pronouns in TABLE 4.2. Consequently, although the Kelabit pronouns may well have developed diachronically from genitive or ergative case marked forms, such analyses no longer reflect the synchronic structure of the language.

¹⁸⁵ An interesting avenue for future research would be to explore the frequency and distribution of the variant pronouns in the naturalistic corpus (see APPENDIX 1). This may help to clarify why one or other of the pronouns is used in contexts where both are grammatical (cf. SUBSECTION 6.4).

4.2.3 A Prosodic Distinction

In SUBSECTION 4.2.2, I illustrated that Kelabit FORM 1 and FORM 2 pronouns alternate with one another. However, it is not the case that they freely alternate and thus we can maintain that there is a distinction between them beyond their different forms, even if this distinction is not analysed as case or in relation to the voice system.

The first important difference is distribution. FORM 1 is much more widely distributed than FORM 2, which is common in the languages of Borneo with reduced FORM 2 paradigms (cf. Soriente 2013). FORM 1 can appear in all the positions of full NPs, including before and after the verb, as shown in (5) to (9). It also appears in equative clauses, unlike FORM 2:

(10) *Kelabit Pronoun Distribution*

a. **FORM 1**

Uih nih.
1SG.1 DEM
'It's me.'

b. **FORM 2**

*Kuh nih.
1SG.2 DEM
For: 'It's me.'

(elicitation, fieldnotes)

In fact, FORM 2 is restricted to the contexts outlined in SUBSECTION 4.2.2. It cannot alternate with FORM 1 in any other contexts, such as marking pronominal subjects or non-actor, non-subject functions:

(11) *Restrictions on Kelabit FORM 2*

a. **FORM 2 as non-actor subject**

*Seni'er ieh kuh.
UV.PFV.see 3SG.1 1SG.2
A/N-S N-A/S

For: 'He saw me.'

b. **FORM 2 as actor subject**

*Kuh ni'er ieh.
1SG.2 AV.see 3SG.1
A/S N-A/N-S
For: 'He saw me.'

c. **FORM 2 as non-actor non-subject**

*Ieh ni'er kuh.
3SG.1 AV.see 1SG.2
A/S N-A/N-S
For: 'He saw me.' (elicitation, fieldnotes)

This suggests a possible prosodic explanation, as the property of having 'special syntax', or a distribution distinct from NPs, is a property widely associated with clitic forms (Zwicky 1977, SUBSECTION 4.3.1).

A second difference is that FORM 2 can never appear sentence-initially and typically follows a verb, noun or preposition, as shown in (12):

(12) *Kelabit FORM 2 word order*

a. **FORM 2 sentence-intially**

*Kuh seni'er ieh.
1SG.2 UV.see 3SG.1
A/N-S N-A/S
For: 'I saw him.'

b. **FORM 2 pre-verbally**

*Na'am kuh keliq.
NEG 1SG.2 know
'I don't know.' (elicitation, fieldnotes)

Although the word order in (12b) is exactly the order used in (8a) with FORM 1 pronouns, it is not grammatical for FORM 2, which follow the verb, as in (8b).¹⁸⁶

¹⁸⁶ This pattern was confirmed by a number of speakers. Nonetheless, there is one counter-example in the corpus. This could represent dialect differences or age-variation and would be an interesting avenue for future research.

A constraint against sentence-initial position is not unlike the behaviour of the Dutch clitic pronoun *ie*, which can never appear sentence-initially, unlike the strong pronoun form *hij*:

- (13) *Dutch Clitics*
- a. **Weak Pronoun Pre-verbally**
 *[Ie] komt morgen.
 3SG come.3SG.NON.PST tomorrow
 ‘He will come tomorrow.’
- b. **Strong Pronoun pre-verbally**
 [Hij] komt morgen.
 3SG come.3SG.NON.PST tomorrow
 ‘He will come tomorrow.’
- c. **Weak Pronoun post-verbally**
 Komt [ie] morgen?
 come.3SG.NON.PST 3SG tomorrow
 ‘Will he come tomorrow?’
- d. **Strong Pronoun post-verbally**
 Komt [hij] morgen?
 come.3SG.NON.PAST 3SG tomorrow
 ‘Will he come tomorrow?’ (van der Leeuw 1997: 52)

In the post-verbal position in Dutch, the weak and strong forms are free to alternate. However, only the strong pronoun form is grammatical sentence-initially. Hence, the distribution of FORM 2 and the fact that it cannot occur sentence-initially fits with the idea that FORM 2 could represent a prosodically weak form in the sense of Cardinaletti & Starke (1999).

Finally, it seems that FORM 2 cannot be focused, whereas FORM 1 can be highlighted in argument focus:

- (14) *Kelabit Focus*
- Q. Senuruq iih tieh ngelaak ngen tauh?
 UV.PFV.order who PT=3SG AV.cook for 1PL.INCL
 ‘Who ordered her to cook for us?’
- A1: Senuruq uih tieh.
 UV.PFV.order 1SG.1 PT=3SG
 ‘I ordered her.’
- A2: *Senuruq kuh tieh.
 UV.PFV.order 1SG.2 PT=3SG
 ‘I ordered her.’
- (elicitation, fieldnotes)

Again, this is a property of prosodically weak elements. For example, Chung (2003: 551) describes a similar restriction in Chamorro. Thus, some of the differences in distribution between FORM 1 and FORM 2 suggest that FORM 2 may be prosodically weaker than FORM 1. Consequently, the rest of the chapter explores the prosody of the Kelabit pronouns and how this compares to other Western Austronesian languages.

4.2.4 Summary

In this section, I discussed a set of variant pronouns in Kelabit that do not have many phonological similarities with their basic counterparts. I outlined two previous accounts of pronouns in the languages of Borneo that treat cognate forms as marking genitive case and/or marking actors in non-actor voices. However, I determined that the Kelabit FORM 2 pronouns cannot be accounted for by a case-based analysis or a voice-based analysis. Finally, I discussed some additional distributional differences between FORM 1 and FORM 2, including the fact that FORM 2 cannot appear in initial position or in focus contexts, whilst FORM 1 has a wide distribution in line with full NPs. This prompted the proposal that FORM 2 pronouns are prosodically weaker clitic forms. In the following section, I define clitics in the context of this study and summarise pronominal clitic phenomena in other Austronesian languages.

4.3 Clitics in the Literature

Clitics have been widely discussed in the literature from both a cross-linguistic perspective and in theoretical phonology, morphology and syntax (Zwicky 1977, Nevis et al. 1994, Gerlach & Grijzenhout 2001, Everett 1996, Klavans 1985, Kaisse 1985, Spencer 1991, Anderson 1992, Nespor 1993 among others). The main areas of debate include where clitics fit in the prosodic hierarchy (Nespor & Vogel 1986); whether ‘clitic’ is needed as a separate morphological category from ‘word’ and ‘affix’ (Anderson 1992); and how to analyse the irregular syntax of clitics which do not always seem to follow X-bar theoretic rules (Bögel et al 2010).

Clitics are normally seen as unstressed, mono-syllabic forms of functional categories such as pronouns, auxiliaries, determiners, question particles and negation (Gerlach & Grijzenhout 2001, Austin 2004).¹⁸⁷ They are typically defined in contrast to independent words and dependent affixes (Zwicky 1977, Zwicky & Pullum 1983). On the one hand, they differ from independent words in that they cannot stand on their own and are dependent upon a prosodic host. On the other hand, they differ from affixes in that they are less selective of their hosts, which can belong to many different word classes.¹⁸⁸ Take, for example, the possessive ‘s in English which attaches to the NP that it modifies, but can take words of any syntactic class as its prosodic host (cf. Klavans 1985):

- (15) *English Possessive Clitic*
a. **Prosodic host = N**
[_{NP} the man]’s hat

¹⁸⁷ Of course, this may not always be the case and depends somewhat on the definition of clitics. In Kananavu, for example, there are multisyllabic clitics that receive stress in certain environments (Elizabeth Zeitoun, p.c.).

¹⁸⁸ See Zwicky (1977) and Zwicky & Pullum (1983) for a list of phonological, morphological and syntactic factors to distinguish between affixes, clitics and words. These are tendencies and it is possible to find exceptions for most of the factors explored (cf. van der Leeuw 1997). They nonetheless support the idea of clitics falling in-between words and affixes.

- b. **Prosodic host = V**
[_{NP} the man who ran]’s hat
- c. **Prosodic host = P**
[_{NP} the man I got this for]’s hat
- d. **Prosodic host = Adj**
[_{NP} the man whose hair was brown]’s hat

In contrast, affixes, such as the regular past tense inflection *-ed* in English, are morphosyntactically dependent on stems of particular word classes, in this case verbs:

- (16) *English Past Tense Affix*
- a. **Prosodic host = V**
walk-ed
 - b. **Prosodic host = N**
*child-ed
 - c. **Prosodic host = P**
*for-ed
 - d. **Prosodic host = Adj**
*happy-ed

Consequently, the main differences between clitics, affixes and words can be summarised in TABLE 4.9, following van der Leeuw (1997):

Table 4.9 Clitics, Words and Affixes

	Words	Clitics	Affixes
Morphosyntax	independent	independent	dependent
Prosody	independent	dependent	dependent

Words are both prosodically and morphosyntactically independent, whilst affixes are both morphosyntactically and prosodically dependent on their stem.¹⁸⁹ Clitics fall in the middle as they are morphosyntactically independent like words, but prosodically dependent like affixes in that they lack inherent stress (Zwicky 1977, van der Leeuw 1997: 22). Thus, prosodic structure is key to defining clitics and I adopt a prosodic methodology for analysing the variant pronouns (SUBSECTION 4.4).

4.3.1 Clitic Subtypes

Clitic phenomena can be subcategorised according to position and type. In terms of position, Zwicky (1977) distinguishes between proclitics, enclitics and endoclitics. Clitics that precede their host are known as proclitics and those that follow their host are known as enclitics. Endoclitics and mesoclitics, where the clitic occurs within the prosodic word rather than at its left or right edge, are considerably rarer. Nonetheless, there are a number of reported examples, such as Brazilian Portuguese, which creates two prosodic words after the insertion of the clitic between the stem and the affix:¹⁹⁰

- (17) *Brazilian Portuguese Endoclitics*
- | | | | | |
|----|------------------------|---|---------------------------|-------------------------|
| a. | (daría) _{Pw} | → | (dár-lho) _{Pw} | -(ía) _{Pw} |
| | 1SG.would.give | | give-3SG.OBJ.to.3SG | -1SG.would |
| | ‘I would give.’ | | ‘I would give it to him.’ | |
| | | | | |
| b. | (trarás) _{Pw} | → | (trá-lo) _{Pw} | -(ás) _{Pw} |
| | 2SG.will.bring | | bring-3SG.OBJ | -2SG.will |
| | ‘You will bring.’ | | ‘You will bring it.’ | (van der Leeuw 1997:27) |

¹⁸⁹ Van der Leeuw (1997) suggests that affixes like the Pashto perfective prefix *wə-* can be inherently stressed, and therefore independent on the prosodic structure of the host to which it attaches. However, typically affixes form a single prosodic word with their stems.

¹⁹⁰ Other examples include the Estonian emphatic particle *-ki* and infixes Hua pronoun clitics (Zwicky 1977).

As Klavans (1985: 97) discusses, the domains of phonology and syntax are independent in terms of clitic position. Thus, a clitic which is syntactically dependent on the previous word is not necessarily prosodically enclitic, as discussed in SUBSECTION 4.6.

In terms of type, a distinction is often drawn between simple clitics, special clitics and bound word clitics (Zwicky 1977). A simple clitic is a phonologically reduced form of a full lexical or functional item. Apart from its prosodic dependency, the simple clitic is syntactically identical to its non-clitic variant. A common example is English *'ll*, a phonologically reduced form of the auxiliary *will* (cf. Austin 2004):

- (18) *English*
- a. I will see you tomorrow.
 - b. [I]'ll see you tomorrow.
 - c. [The doctor]'ll see you tomorrow.
 - d. [The doctor in room three]'ll see you now.

In (18b), (18c) and (18d) the full form of (18a) could also have been used. Thus, the simple clitic has the same syntactic distribution as the full form. This sort of cliticization is often linked to stylistic factors and is common in fast and casual speech (Zwicky 1977: 5).

In contrast, special clitics are ‘weak’ forms of ‘strong’ functional items. For example, many languages are said to have a contrast between strong and weak pronouns, including French, Dutch, Egyptian Arabic and Warlpiri (cf. Zwicky 1977, SUBSECTION 4.2.3). Strong forms are so-called since they can be stressed and form a constituent of their own, unlike weak forms (van der Leeuw 1997: 2). Weak forms often have a different distribution to either strong forms or lexical items. For example, consider Italian object pronouns:

- (19) *Italian*
- a. **Strong Pronoun**
 Elena torna, e trova proprio me a casa.
 Elena return.3SG and find.3SG certainly 1SG.ACC LOC house
 ‘Elena returns and finds me at home.’
- b. **Weak proclitic**
 Elena torna, e mi= trova a casa.
 Elena return.3SG and 1SG.ACC find.3SG LOC house
 ‘Elena returns and finds me at home.’
- c. **Weak enclitic**
 Elena torna a trovar =mi a casa.
 Elena return.3SG to find. 1SG.ACC LOC house
 ‘Elena returns to find me at home.’ (Austin 2004)

In (19a), the strong pronoun is used when an adverb separates the verb and the object. However, a weak proclitic is used with finite verbs, as in (19b), and a weak enclitic with non-finite verbs, as in (19c). These are subject to what Zwicky (1977: 4) terms ‘special syntax’ as their distribution contrasts with that of other paradigmatically related forms. Moreover, though simple clitics are often phonologically related to the full lexical items, as in (19), weak and strong pronouns can be phonologically unrelated, like the Dutch pronouns in (13). According to Zwicky (1977), special clitics are used when the pronouns are unstressed, whilst strong forms are used when the pronoun needs to be accentuated for information structural reasons.

Finally, a bound word clitic is one that attaches to words and phrases of different morphosyntactic classes, unlike the Italian pronouns which cliticise to verbs or English *'ll* which attaches to the subject NP. They typically have no non-clitic alternative and often appear directly after a first word, syntactic constituent or prosodic unit. For this reason, they have come to be known as second-position clitics or Wackernagal clitics, following Wackernagel’s law (Wackernagel 1892). One example is Serbo-Croatian:

- (20) *Serbo-Croatian Clitics*
- a. **Host = first word**
 Taj [joj ga je] čovek poklonio.
 that 3SG.F 3SG.N AUX man present
 ‘That man presented her with it.’
- b. **Host = first constituent**
 Taj čovek [joj ga je] poklonio.
 that man 3SG.F 3SG.N AUX present
 ‘That man presented her with it.’ (Bögel 2010)

In (20), the clitic cluster *joj ga je* attaches in second position, whether this is following the first word, as in (20a), or following the first constituent, as in (20b). The constituent *taj čovek* cannot normally be discontinuous in Serbo-Croatian, which has been suggested as evidence for the fact that second-position should be defined prosodically rather than syntactically (Lowe 2011: 364). Nonetheless, whether the definition of second-position should be syntactic, prosodic or a combination of both remains a matter of some debate (Diesing et al 2009, Chung 2003).

In short, the differences between the three types of pronouns can be summarised in TABLE 4.10:

Table 4.10 *Simple Clitics, Special Clitics and Bound Words*

	Simple Clitic	Special Clitic	Bound Word Clitic
Corresponding non-clitic form	Full lexical or functional item	Strong functional form	No non-clitic variant
Relationship to non-clitic variant	Phonologically reduced form	No phonological relationship	n/a
Distribution	Same as full form	Distinct from strong form	Second-position

Historically, it is possible to think of the different clitic types as reflecting different stages of development in a grammaticalisation pathway from word to clitic to affix

(cf. van Gelderen 2011). Indeed, the distinction between bound-word and special clitics is not dissimilar to Klavans' (1983) distinction between lexical and post-lexical clitics. In Klavans' model, lexical clitics, such as the Romance pronouns, select a host of a particular word-class. Post-lexical clitics, in contrast, are second-position clitics which are added enclitically to whichever phrase appears initially. This, Klavans (1983) argues, reflects the fact that lexical clitics are in the process of becoming affixes and are therefore attached in the lexicon, whilst bound word clitics are attached post-lexically. Uriagereka (1995) also distinguishes between weak clitics and strong clitics in Romance: weak clitics have more in common with affixes, whilst strong clitics have more in common with independent words.¹⁹¹

Hence, clitics can appear in different positions and be of different types, which may well represent different stages of development or degrees of grammaticalisation. This becomes particularly useful as a way of categorising the development of clitic phenomena in Austronesian, as we will see in the following section.

4.3.2 Austronesian Clitic Phenomena

Western Austronesian pronouns, and particularly the so-called genitive pronouns, are often analysed as clitics (Himmelmann 2005a, Chung 2003, Tsukida 2005, Rubino 2005). For example, Lobel's (2013) reconstruction analyses the genitive pronouns of Proto-Southwest Sabah as clitics in SUBSECTION 4.2.1.1. Indeed, Ross (2006, 2015) suggests that clitic pronouns were a feature of Proto-Austronesian (PAn) which he reconstructs with a single set of bound enclitic pronouns, serving both genitive and nominative functions. By the time of Proto-Malayo Polynesian, he suggests that the

¹⁹¹ See Cardinaletti & Starke (1995) for similar arguments in Germanic.

enclitic pronouns inherited from PAN were exclusively genitive and that a new set of nominative clitics had been derived from old free-standing pronouns:

Table 4.11 Proto-Austronesian and Proto-Malayo Polynesian Clitic Pronouns (Ross 2006)

	Proto-Austronesian	Proto-Malayo Polynesian	
		GEN	NOM
1SG	*=ku	*=ku	*=(h)aku
2SG	*=Su	*=mu	*=ka(hu)

Given their historical provenance and widespread distribution in Austronesian today, Himmelmann (2005a: 131) seems justified in suggesting that clitics are most likely attested in all Western Austronesian languages.

Though most Western Austronesian languages can be said to have pronominal clitics, not all Austronesian pronominal clitics are of the same type. In fact, clitic phenomena are another means often used to distinguish between Philippine-type and Indonesian-type languages (cf. Himmelmann 2005a, Brickell 2014, SUBSECTION 1.3.1). For example, Himmelmann (2005a) suggests that second-position enclitics are a key feature of Philippine-type languages, whilst proclitic actors are characteristic of Indonesian-type. Lee & Billings (2005) suggest that Austronesian clitic phenomena can be subdivided into what they term Wackernagel and verb-adjacent clitics. This is similar to the distinction between bound-word clitics and special clitics outlined in SUBSECTION 4.3.1, since Wackernagel clitics always attach to the first constituent, whilst verb-adjacent clitics attach to the verb, regardless of its position.

Billings & Kaufman (2004) surveyed several languages of Taiwan, the Philippines and Sulawesi. They suggest that more conservative languages in the North have Wackernagel clitics, whilst the more innovative languages in the South, especially in Sulawesi, have verb-adjacent clitics, either enclitic or proclitic.

According to Lee & Billings (2005), a number of languages from the Philippines are transitional between Wackernagel and verb-adjacent clitic positions. Finally, many languages in Indonesia have actor pronouns that are proclitic to the verb in UV (see SUBSECTION 1.3.1, 3.4.2.1). These are sometimes considered prefixes and may therefore represent the endpoint of grammaticalization.

Austronesian Wackernagel clitics can be illustrated from Seediq, an Atayalic language spoken in Taiwan. Holmer & Billings (2014) suggest that clitics attach to the first constituent, providing it is a syntactic head (see also Starosta 2009e). This includes verbs, negation, tense/aspect markers, subordinators and interrogatives:

- (21) *Seediq clitics*
- a. **Prosodic host = negation**
 Ini=ku imah sino.
 NEG=1SG.NOM AV.CNG.drink wine
 ‘I don’t drink wine.’
- b. **Prosodic host = tense/aspect marker**
 Wada=mu puq-un ka damac.
 PST=1SG.GEN eat-UV NOM food
 ‘I ate the food.’
- c. **Prosodic host = subordinator**
 [...] Ado=ku m-beyax t<m>alang yaku.
 because=1SG.NOM AV-strong <AV>run 1SG
 ‘[...] Because I am good at running.’
- d. **Prosodic host = interrogative**
 Ye=ku ini huwa m-ekan tmaku hini?
 Q=1SG.NOM NEG how.CNG AV-eat tobacco here
 ‘Is it ok if I smoke here?’
(Holmer 2005: 190)

Since the clitic pronouns do not select a host of a particular category and occur in second position in (21), they can be classified as Wackernagel or bound-word clitics. As can be seen in (21a) and (21b), this applies equally to NOM and GEN pronouns.

The languages of the Philippines are also commonly described as having Wackernagel clitics (Billings & Kaufman 2004, Kroeger 1998c). However, recent accounts have suggested that languages like Tagalog actually show a mixture of Wackernagel and verb-adjacent uses. In many cases it is difficult to decide between the two analyses since clauses tend to be verb-initial, making the immediate post-verbal position also the second-position that is relevant for Wackernagel clitics (SUBSECTION 5.4). Nonetheless, Wackernagel patterns do occur in the contexts in which they can be distinguished from verb-adjacent clitics:

- (22) *Tagalog clitics*
- a. **Prosodic host = interrogative**
 Bakit =ka hindi s<um>a-sagot?
 why 2SG.NOM NEG <AV>IRR-answer
 ‘Why aren’t you answering?’
- b. **Prosodic host = negation**
 Bakit hindi =ka s<um>a-sagot?
 why NEG 2SG.NOM <AV>IRR-answer
 ‘Why aren’t you answering?’ (Lee & Billings 2005: 242)

In (22a), the clitic is positioned between an initial question word and a negative. It is not adjacent to the verb and consequently unambiguously in a Wackernagel position. The order in (22b) is also attested and could be described as a Wackernagel pattern, since the clitic does not follow the verb but rather attaches to the negative marker. However, the syntax does not preclude other analyses, since the pronoun could also be syntactically proclitic to the verb. Indeed, this order is sometimes seen as the mid-stage in the transition from Wackernagel enclitics to Indonesian-type proclitic actors (Wolff 1996, see below). Moreover, corpus studies show an overall preference for verb-adjacent order (Lee & Billings 2005) and theoretical accounts have been proposed that treat Tagalog clitics as verb-adjacent (Billings & Konopasky 2002).

Consequently, although Tagalog can be said to display Wackernagel patterns in the contexts in which they can be distinguished, it has an overall preference for verb-adjacent ordering, which could act as a trigger for reanalysis.

Some languages in Southeast Mindanao in the Philippines have verb-adjacent rather than Wackernagel clitics (cf. Lee & Billings 2005). In other words, the position of the clitic is defined relative to the verb rather than an initial position. This can be illustrated for Tagakaulo:

- (23) *Tagakaulo*
 a. **NOM and GEN clitics**
 Ananga' wala' da= mu= kami= kilala?
 why NEG PFV 2SG.GEN 1SG.NOM UV.recognise
 'Why do you not recognise us?' (Guili 1978: 76)

Unlike Tagalog in (22), in Tagakaulo the NOM and GEN clitics are verb-adjacent proclitics. They immediately precede the verb, rather than being enclitic to the initial interrogative. Nonetheless, they maintain the restriction against clitics in initial position (Lee & Billings 2005: 252). Consequently, the languages described in Lee & Billings (2005) are considered transitional between Wackernagel clitics and verb-adjacent clitics.

In parts of Sulawesi, clitic phenomena are unambiguously verb-adjacent. Interestingly, unlike Seediq, Tagalog and Tagakaulo, NOM and GEN clitics behave differently. For example, consider the Kaili-Pamona language Kulawi. The so-called GEN pronouns in Kulawi attach as verb-adjacent enclitics in realis mood and proclitics in irrealis mood (Billings & Kaufman 2004). In contrast, NOM pronouns can be either enclitics or free forms. This is illustrated in (24):

- (24) *Kulawi Clitics*
- a. **GEN clitic**
 Moma i-epe=ku ka=rata=mu.
 NEG REAL-hear=1SG.GEN NOM=come-2SG.GEN
 ‘I haven’t heard about your arrival.’
- b. **NOM clitic**
 Moma=’a t<um>ai.
 NEG=1SG.NOM <AV>come.here
 ‘I didn’t come here.’ (Billings & Kaufman 2004: 21)

In (24a), unlike Tagalog in (22), the GEN clitic is enclitic to the verb, rather than the fronted negation *moma*. In (24b), in contrast, the NOM pronoun is enclitic to the negation. Hence, NOM and GEN clitics have a different distribution in Kulawi and GEN clitics can be described as verb-adjacent in that they always follow the verb, regardless of what comes in clause-initial position. The restriction against clitics in initial position is largely absent in Sulawesi (Lee & Billings 2005).

Finally, in some Indonesian-type languages, there are pronominal prefixes or proclitics that indicate a first or second person actor in undergoer voice, the typical function of a GEN pronoun (SUBSECTION 4.2). These are typically in complementary distribution with the UV prefix for third person actors and sometimes appear cognate to the GEN pronouns in Philippine and Sulawesi languages. Consider the following examples from Indonesian and Javanese:

- (25) *Indonesian*
- a. **Free-standing 1SG pronoun in AV**
 Aku membaca buku.
 1SG AV.read book
 ‘I read books.’
- b. **Clitic 1SG pronoun in UV**
 Buku ku=baca.
 book 1SG=read
 ‘The book was read by me.’ (adapted from Musgrave 2002: 38)

- (26) *Javanese*
- a. **Clitic 2SG pronoun in UV**
 Surat wis mbòk=kirim.
 letter PFV 2SG=send
 ‘The letter has been sent by you.’
- b. **Clitic 1SG pronoun in UV**
 Surat wis tak=kirim.
 letter PFV 1SG=send
 ‘The letter has been sent by me.’ (elicitation, fieldnotes)

Clitics such as those in (25) and (26) are variously treated as prefixes and proclitics (Himmelmann 2005a: 132). Indeed, Austin (2004: 13) suggests that there is some evidence in Sasak dialects for a historical change from special clitics to affixes.¹⁹² However, pronouns representing subjects – or NOM pronouns – as in (25) are non-clitic forms. Thus, the range of clitic phenomena in Western Austronesian does seem to illustrate a process of grammaticalisation from more word-like clitics to weak clitics to affixes, at least in the GEN pronouns.

A number of works have considered possible stages of development from Wackernagel clitics in the more conservative Philippine-type languages to verb-adjacent clitics and finally affixes in the more innovative Indonesian-type languages (van den Berg 1996, Wolff 1996, Zobel 2002, Billings & Kaufman 2004). Billings & Kaufman (2004), following Wolff (1996), suggest the following potential stages of reanalysis:

¹⁹² See Austin (2011) for further discussion of differences between Sasak dialects with respect to clitic phenomena. Note that some dialects such as Ngenó-ngené have systems similar to the languages of Sulawesi where proclitics are used in irrealis mood, whilst other dialects such as Menó-mené have verb-adjacent clitics for U, but Wackernagel enclitics for S/A (Austin p.c.).

Table 4.12 *The Re-analysis of Austronesian Clitics*

Stage of Development	Clitic position		Clitic type
1.	NEG=GEN.PRON	UV.verb	Wackernagel
2.	NEG	GEN.PRON=UV.verb	Verb-adjacent
3.		GEN.PRON-UV.verb	Verbal prefix

They argue, following Himmelmann (1996), that the change from Stage 1 to Stage 2 is triggered by the fact that the realis form of UV is marked with a prefix in many languages, whilst the irrealis form is unmarked, leaving a slot for pronominal proclitics. Specifically, Billings & Kaufman (2004) argue that the change is triggered by the fact that the Proto-Austronesian **-in-* perfective/realis undergoer voice marker has become a prefix *ni-* or *i-* in the verb-adjacent languages that they discuss. In any case, clitics in Austronesian are not all of one type but rather can be classified as second-position, verb-adjacent or affix-like according to their position in a process of grammaticalization.

4.3.3 Possible Kelabit clitic patterns

Given the previous discussion, there are three parameters that seem to vary across Western Austronesian clitic phenomena: the clitic position, the clitic type and whether NOM and GEN pronouns behave the same or differently. From the syntax of Kelabit, we might conclude that the pronouns are enclitic, since they always follow the verb, noun or preposition that they form a syntactic phrase with:

- (27) *Kelabit Clitic Syntax*
- a. [Senatek kuh]_{VP} neh bubpuq ih.
 UV.PFV.close 1SG.2 PT door PT
 ‘I closed the door.’
 (elicitation, BAR17102013CH_01 00:53:02.138-00:53:04.854)

- b. Edteh laak ieh [ruyung kuh]_{PP}.
 one year 3SG.I with 1SG.2
 ‘He was with me for one year.’
 (text, BAR04092014CH_02 00:00:56.850-00:00:58.410)

Similarly, they combine as enclitics with prepositions and particles:

- (28) a. *ngekuh* ‘to me’ → ngen + kuh
 b. *kekuh* ‘I say’ → ken + kuh

However, as discussed in SUBSECTION 4.3.1, syntactically enclitic pronouns are not always prosodically enclitic.

In terms of type, SUBSECTION 4.3.2 demonstrated that distinguishing between Wackernagel clitics and verb-adjacent clitics in verb-initial languages depends on contexts in which second position is distinct from the immediate post-verbal position. These include fronted adverbs and negation. Wackernagel clitics in Tagalog follow the fronted adverbial phrase, as shown in (29), or the negation as shown in (22):

- (29) *Tagalog*
 a. **Fronted Adverb**
 [Bukas ng gabi nang alas.otso] =siya aalis.
 tomorrow GEN night ADV eight.o’clock 3SG.NOM FUT.AV-leave
 ‘It’s tomorrow night at eight that he’s leaving’ (Kroeger 1998c)

In Kelabit, the UV clauses where FORM 2 pronouns occur are typically verb-initial (SUBSECTION 5.5.1.3). However, it is possible to find fronted adverbs pre-verbally, either separated into a distinct intonational phrase and delimited by a pause, or within the same intonational phrase. In the second instance, the pronoun does not encliticise to the adverb – as it does in (29) – but follows the verb instead:

- (30) *Kelabit*
 a. **Fronted Adverb**
 [Iyuk~iyuk] niding =neh sineh keyh.
 REDUP~grow UV.PFV.lift 3SG.2 DEM PT
 ‘Gradually he lifted that same one.’
 (text, BAR03082014CH_01 00:00:44.550-00:00:46.560)

Hence, we might conclude that the Kelabit FORM 2 pronouns are verb-adjacent.

Finally, much like Kulawi in (24) and the Indonesian equivalents in (25), FORM 1 and FORM 2 pronouns have a different distribution (SUBSECTION 4.2.3). This can be seen particularly in the context of negation, as shown in (8) and repeated below:

- (31) *Kelabit Negation*
 a. **FORM 1**
 Na’am uih keliq.
 NEG 1SG.1 know
 ‘I don’t know.’
 b. **FORM 2**
 Na’am keliq kuh.
 NEG know 1SG.2
 ‘I don’t know.’ (elicitation, fieldnotes)

In (31b), the FORM 2 pronoun is enclitic to the verb rather than the negative, which further supports an analysis of Kelabit FORM 2 pronouns as verb-adjacent. In (31a), however, the FORM 1 pronoun follows the negative rather than the verb. This is the position that full NPs are most likely to occur in following *na’am* ‘NEG’, as discussed in SUBSECTION 2.5.1. Consequently, the data seems to support an analysis of FORM 1 pronouns as free forms rather than clitics. Alternatively, the order in (31a) could suggest an analysis of FORM 1 pronouns as Wackernagel clitics, attaching in second position. Whichever analysis provides a better account of the data, Kelabit appears to follow a similar pattern to Kulawi and have NOM and GEN pronouns that differ in their distribution and clitic status.

4.3.4 Summary

In this section, I defined clitics as elements that are prosodically weak and require a prosodic host to attach to. I observed that clitics can appear in different positions and tend towards three different types. How they are classified depends largely on their distribution and whether this is defined according to the distribution of a corresponding full lexical item (SIMPLE CLITICS); relative to a specific lexical host (SPECIAL CLITICS) or relative to a post-lexically defined first position (BOUND-WORD CLITICS). Moreover, I argued that Austronesian clitic phenomena are spread along a grammaticalization cline from strong clitics to weak clitics to affixes. The outer points on this continuum have been treated as defining characteristics of Philippine-type (WACKERNAGEL ENCLITICS) and Indonesian-type languages (PROCLITIC ACTORS). The Kelabit data seems to support a view of FORM 2 pronouns as intermediate verb-adjacent clitics. In order to explore this hypothesis, I now present a methodology for identifying whether or not the Kelabit pronouns are clitics and whether they attach proclitically or enclitically.

4.4 Methodology for Identifying Clitics

In the previous section, I showed that clitic pronouns in Austronesian are split between Wackernagel and verb-adjacent clitics. I then proposed that Kelabit FORM 2 pronouns are verb-adjacent clitics and form a syntactic unit with the element to their left. However, I defined clitics as elements with a prosodic dependence, following van der Leeuw (1997). Therefore, we need to adopt a prosodic methodology in order to analyse whether or not the Kelabit pronouns are indeed clitics and whether they attach proclitically or enclitically. This section sketches a few relevant aspects of prosodic theory in order to develop a prosodic means of testing the hypothesis.

4.4.1 Prosody

Prosody is typically defined as having two main linguistic purposes: highlighting and defining boundaries (cf. Grice & Baumann 2007). Highlighting involves giving prominence to a particular part of the utterance, whereas boundary marking involves delimiting prosodic constituents. Both make use of the prosodic features pitch, loudness, length and vowel quality. These can also be analysed in terms of their acoustic correlates (Grice & Baumann 2007, Cruttenden 1997):

Table 4.13 Prosodic Features and Acoustic Correlates

Prosodic feature	Acoustic correlate
Pitch	Fundamental Frequency (F ₀)
Loudness	Intensity
Length	Duration
Vowel Quality	Spectral quality/formants

In terms of prosodic constituents, prosodic phonology typically recognises a prosodic hierarchy, where the largest constituent is the utterance and the smallest constituent the syllable (van der Leeuw 1997: 9, Selkirk 1980):¹⁹³

- (32) **The Prosodic Hierarchy**
- a. Phonological Utterance
 - b. Intonational Phrase
 - c. Phonological Phrase
 - d. Prosodic Word
 - e. Foot
 - f. Syllable

Prosodic constituents loosely correspond to syntactic units, such as clauses, phrases and terminal nodes. For example, Intonational Phrases are typically uttered under a

¹⁹³ Nb. various terms are used to denote hierarchical prosodic structure in the literature (cf. Hirst & Di Cristo 1998: 35-36). For example, Intonational Phrases are variously known as Intonation Units, Breath Groups and Intonation Phrases (cf. Fox 2000, Cruttenden 1997, Pierrehumbert 1980).

single intonation contour and often correspond to syntactic clauses or units of information (Chafe 1987). Similarly, prosodic words typically correspond to a terminal node at c-structure (Lowe 2011: 367). This is where rules of stress and phonological processes such as vowel harmony apply. However, the correspondence is not exact and a number of studies have revealed mismatches between prosody and syntax (see SUBSECTION 4.6).

The prosodic hierarchy is also assumed to be subject to the Strict Layer Hypothesis which states that a prosodic level consists exclusively of units of the prosodic level directly below it (Selkirk 1984). The Strict Layer Hypothesis can be summarised by the following principles (Selkirk 1995):

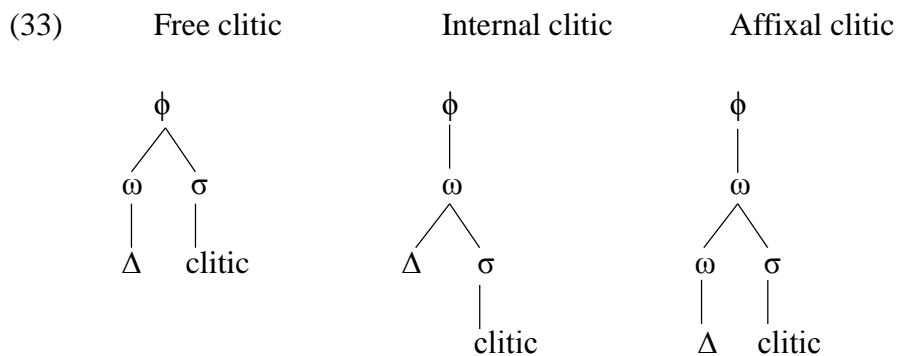
Table 4.14 The Strict Layer Hypothesis

1. Layeredness	A prosodic constituent cannot be dominated by a constituent that is lower in the hierarchy
2. Headedness	A prosodic constituent must dominate at least one constituent of the next lowest level
3. Exhaustivity	A prosodic constituent cannot dominate a constituent that is more than one level lower in the prosodic hierarchy
4. Nonrecursivity	A prosodic constituent cannot be dominated by a constituent of the same level.

In other words, the Strict Layer Hypothesis states that intonational phrases (IPs) are made up exclusively of prosodic phrases (iPs), and prosodic phrases are made up exclusively of prosodic words (Pws) and so on. The question becomes: how do we deal with clitics in this context?

Some accounts assume a separate clitic group as part of the prosodic hierarchy (cf. Nespor & Vogel 1986, Hayes 1989). However, more recent accounts typically agree that adding a separate layer is neither necessary nor independently motivated

(cf. Gerlach & Grijzenhout 2001). Some treat clitics as being phrasal affixes (Anderson 1992). Others argue that clitics can either attach at the level of prosodic word or to the prosodic phrase in order to explain phonological similarities and differences between clitics and affixes (Selkirk 1995, Berendsen 1986, Peperkamp 1996). Indeed, Selkirk proposes a prosodic typology of clitics depending on the level at which they attach (Δ = host, σ = syllable, ω = word, ϕ = phrase):



As schematised in (33), free clitics attach directly to the phonological phrase. Affixal clitics attach to the prosodic word but a prosodic word boundary exists between the host word and clitic. In contrast, an internal clitic is one which attaches to the prosodic word without a prosodic boundary between host and clitic. The concept of free clitics is used to explain patterns such as final-devoicing in Dutch, which does not occur under affixation but does occur under cliticization:

- (34) *Dutch*
- a. **Affixation**
*geef*er ‘giver’
 /xe:v/_{stem} + -/ər/_{affix} → [xe:vər]
- b. **Cliticization**
geef ‘r een *kat* ‘give her a cat’
 /xe:v/_{host} + /ər/_{clitic} → [xe:fərənkat]

(Gerlach & Grijzenhout 2001: 6)

Since phonological processes such as final devoicing usually occur within the domain of the prosodic word, Gerlach & Grijzenhout (2001) suggests that clitics in Dutch are free clitics and do not form part of the prosodic word of the host like affixes, but attach to form a prosodic phrase. However, this seems to go against the Strict Layer Hypothesis.

For this reason, I adopt the view that clitics form a single prosodic word with their host, following van der Leeuw (1997) and Booij (1996). As van der Leeuw (1997: 2) states, the prosodic word is the natural host for clitics since it is the domain of stress, which is exactly what clitics are defined as lacking (see SUBSECTION 4.3). In any case, whether clitics form part of a clitic group, prosodic word or prosodic phrase with their hosts, we would expect to find boundary effects. Consequently, this section will now consider how the boundaries between constituents are marked and how we can use these effects to test the hypothesis outlined in SUBSECTION 4.3.3.

4.4.2 Prosodic boundary marking

High level boundaries, such as those between utterances and intonational phrases, are most commonly marked with a pause (Cruttenden 1997: 30). However, pauses cannot help to identify smaller prosodic constituents, such as the prosodic word, and even higher prosodic boundaries can be recognised without overt silent pauses (Himmelman & Ladd 2008: 252). In these cases, there are two other significant determiners for prosodic boundaries: syllable lengthening and pitch reset.

The first major determiner is pre-boundary or final lengthening. Numerous studies have shown that syllables tend to undergo final-lengthening at prosodic boundaries (cf. Wightman et al. 1992, Himmelman & Ladd 2008: 247). This applies not only for intonational phrases but also smaller units, including the prosodic word

(Xu 2011, Beckman & Edwards 1990, White & Turk 2010). Final-lengthening is distinct from accentual lengthening since it is not typically accompanied by increased intensity, unlike stress (Grice & Baumann 2007: 31). Moreover, syllable lengthening has been argued to be a cross-linguistic tendency (Grice & Baumann 2007: 31, Cruttenden 1997: 33) and has been identified in a number of languages, including Spanish, French and Italian (Hirst & Di Cristo 1998) and non-Indo-European languages, such as Kayardild (Round 2012).

The second major determiner is pitch reset. Himmelmann & Ladd (2008: 252) argue that a change in pitch often occurs at the start of a new prosodic unit. This may also be accompanied by a pitch fall over the syllable(s) prior to the boundary edge. As Cruttenden (1997: 34) argues, pitch change on unstressed syllables tends to indicate boundaries rather than highlighting. Indeed, Cruttenden (1997: 163) talks of declination as a possible prosodic universal (cf. Hirst & Di Cristo 1998: 19) and falling tone over intonation units has been reported for Tagalog (Cruttenden 1997: 160). Although the effects are likely to be larger with higher prosodic boundaries, it is possible that some effects might be found at lower level boundaries as well.¹⁹⁴ Hence, the key acoustic parameters that can be used to identify prosodic boundaries are duration and F_0 .¹⁹⁵

¹⁹⁴ Nb. Xu (2011) suggests that F_0 effects are only found at prosodic phrase and intonational phrase boundaries. This is supported by the Kelabit findings in SUBSECTION 4.5.

¹⁹⁵ Other factors that could be relevant include post-boundary strengthening of consonants (cf. Cho 2004); word-stress (Iivonen 1998) and changes in intensity or vowel quality, such as the use of creaky voice (cf. Himmelmann & Ladd 2008). However, these factors have not yet been shown to reoccur cross-linguistically in the same way as pitch changes and final-lengthening and are consequently left for future research when both the cross-linguistic situation and Kelabit prosody are better understood.

4.4.3 Formulating predictions

Given the discussion of prosody in SUBSECTION 4.1, we can now formulate two important sets of predictions that allow us to test the hypothesis that Kelabit FORM 2 pronouns are prosodically weaker than FORM 2. I defined clitics in SUBSECTION 4.4.1 as forming a single prosodic word with their host. Consequently, if we consider the syllable immediately preceding the pronoun, the pronoun syllable itself and the syllable immediately following the pronoun, then we make the following predictions:

- (35) a. If the pronoun is enclitic, we expect to find a single boundary after the pronoun syllable.
- b. If the pronoun is proclitic, we expect to find a single boundary before the pronoun syllable.
- c. If the pronoun is free-standing, we expect to find boundaries either side of the pronoun syllable.

This can be represented schematically as follows where σ_{pre} represents the preceding syllable, $\sigma_{pronoun}$ represents the pronoun syllable, σ_{post} represents the following syllable and | represents a boundary:

- (36) *Predictions*
- a. **Enclitic pronoun**
 σ_{pre} $\sigma_{pronoun}$ | σ_{post}
- b. **Proclitic pronoun**
 σ_{pre} | $\sigma_{pronoun}$ σ_{post}
- c. **Free pronoun**
 σ_{pre} | $\sigma_{pronoun}$ | σ_{post}

Secondly, based on the cross-linguistic evidence summarised in SUBSECTION 4.4.2, I predict that Kelabit will show the following word boundary effects:

- (37) **Boundary Effects**
- a. The duration of the pre-boundary syllable will be longer than those preceding and following it.
 - b. The mean F_0 of the pre-boundary syllable will differ significantly from the post boundary syllable where the pitch is reset.

Thus, there are two steps involved in testing the hypothesis that Kelabit FORM 2 pronouns are clitics. The first step is to test whether the acoustic properties of duration and F_0 mark prosodic word boundaries in Kelabit, and the second step is to use any significant determiners of word boundaries to test for the number of boundaries surrounding the pronoun syllable.

4.4.4 Experimental design

In order to test the hypotheses and explore the predictions in SUBSECTION 4.4.3, it is necessary to collect a sample of Kelabit pronouns in use. It is difficult to find a sufficient number of examples to quantitatively analyse duration and mean F_0 effects using the naturalistic corpus.¹⁹⁶ Moreover, it is difficult to control for all of the potentially relevant prosodic factors that might affect duration and pitch, including speaker attitudes/emotions and the way that information is structured in a given example (cf. Himmelmann & Ladd 2008: 260). For this reason, an experimental approach was adopted in order to elicit examples of both pronouns with the same immediate context.

Given the constraints of a field setting, Himmelmann (2006b: 169) suggests that working with four to ten speakers is a good basis for detailed prosodic analysis. In this study, I elicited multiple test sentences from five speakers in order to ensure a

¹⁹⁶ See Himmelmann & Ladd (2008) for discussion of the relative merits of prosodic research using experimental data and naturalistic corpora.

large enough data set for statistical analysis. Following Himmelmann (2006b: 168), the segmental context of the test sentences (excluding the variables) was kept as similar as possible. Only 1SG pronouns were used. This is partly to limit the experiment to a feasible number of sentences to record in the field, and partly because all of the FORM 2 pronouns in TABLE 4.2 appear to behave identically (SUBSECTION 4.2.1). Future research could explore the differences between 1SG, 2SG, 3SG and 3PL. In particular, 3PL may show less of a distinction between *ideh* and *deh* as the forms are so similar in phonological terms, and 2SG *iko* is realised as *ko* in certain contexts which provides another variant to analyse (see SUBSECTION 4.5.4).¹⁹⁷

The test sentences used in the experiment varied along the following syntactic parameters in order to represent different contexts in which the pronouns appear:

Table 4.15 Variables in Prosody Study

Pronoun Form	Transitivity/Voice	Pronoun Function	Pronoun Position
FORM 1	Intransitive	S	Pre-verb
FORM 2	Transitive AV	A	Post-verb
	Transitive UV	U	Post-object

The variants are defined as follows:

(38) **Pronoun Form**

- a. FORM 1: *uih*
- b. FORM 2: *kuh*

(39) **Transitivity/Voice**

- a. Intransitive: A predicate with a single argument S
- b. Transitive AV: A predicate with two arguments: A is the subject and U is the non-subject core argument.
- c. Transitive UV: A predicate with two arguments: U is the subject and A is the non-subject core argument.

¹⁹⁷ Future research could also explore information structure factors, such as focus and prominence, which have been shown to affect prosody cross-linguistically (Xu 2011).

- (40) **Pronoun Function**
- a. S: The sole argument of an intransitive predicate
 - b. A: The actor argument of a transitive predicate
 - c. U: The undergoer argument of a transitive predicate

- (41) **Pronoun Position**
- a. Pre-verb: — V (XP)
 - b. Post-verb: V — (XP)
 - c. Post-object: V XP —

Some combinations of variables are ungrammatical by definition, e.g. the function A or U with an intransitive predicate. Others are not found in the Kelabit corpus (see SUBSECTION 4.2.2). Consequently, test sentences were developed to represent nine syntactic contexts. Examples of each are schematised in (42):

(42) *Contexts used in Prosody Experiment*

a. **Context 1**

FORM 1 Intransitive S in post-verbal position

V_{Intr} S
uih

Dooq pian uih kuman buaq nuk inih.
good like 1SG.1 AV.eat fruit REL DEM
'I like to eat those fruits.'

b. **Context 2**

FORM 2 Intransitive S in post-verbal position

V_{Intr} S
kuh

Dooq pian kuh kuman buaq nuk inih.
good like 1SG.2 AV.eat fruit REL DEM
'I like to eat those fruits.'

c. **Context 3**

FORM 1 Transitive UV A in post-verbal position

V_{UV} $A_{\text{N-S}}$ U_{S}
uih

Senuruq uih tieh nge-laak ngen tauh.
UV.PFV.order 1SG.1 PT=3SG.1 AV-cook for 1PL.INCL
'I asked her to cook for us.'

d. **Context 4**

FORM 2 Transitive UV A in post-verbal position

V_{UV} A_{N-S} U_S
kuh

Senuruq kuh tieh nge-laak ngen tauh.
UV.PFV.order 1SG.2 PT=3SG.1 AV-cook for 1PL.INCL
'I asked her to cook for us.'

e. **Context 5**

FORM 1 Transitive AV A in post-verbal position

V_{AV} A_S U_{N-S}
uih

Pu'un~pu'un ne-kuman uih [edteh buaq kabere].
First~REDUP PFV-AV.eat 1SG.1 one fruit pineapple
'First, I ate a pineapple.'

f. **Context 6**

FORM 1 Transitive AV U in post-verbal position

A_S V_{AV} U_{N-S}
uih

Ieh ne-nuruq uih nge-laak ngen tauh.
3SG.1 PFV-AV.order 1SG.1 AV-cook for 1PL.INCL
'She asked me to cook for us.'

g. **Context 7**

FORM 1 Transitive AV A in pre-verbal position

A_S V_{AV} U_{N-S}
uih

Uih ne-nuruq ieh nge-laak ngen tauh.
1SG.1 PFV-AV.order 3SG.1 AV-cook for 1PL.INCL
'I asked her to cook for us.'

h. **Context 8**

FORM 1 Transitive AV A in post-object position

V_{AV} U_{N-S} A_S
uih

Pu'un~pu'un ne-kuman [edteh buaq kabere] uih.
REDUP~first PFV-AV.eat one fruit pineapple 1SG.1
'First, I ate a pineapple.'

which pronominal subjects and objects were possible and test contexts were built using these predicates in consultation with a native speaker. The final list of examples was then verified for naturalness and piloted with one male and one female speaker before approaching participants to record. The breakdown of test sentences per context is given in TABLE 4.16:

Table 4.16 Test sentences per context

Context	Number of Examples
1. FORM 1 Intransitive S in post-verbal position	3
2. FORM 2 Intransitive S in post-verbal position	3
3. FORM 1 Transitive UV A in post-verbal position	4
4. FORM 2 Transitive UV A in post-verbal position	4
5. FORM 1 Transitive AV A in post-verbal position	2
6. FORM 1 Transitive AV U in post-verbal position	3
7. FORM 1 Transitive AV A in pre-verbal position	3
8. FORM 1 Transitive AV A in post-object position	1
9. FORM 1 Transitive UV U in post-object position	3

4.4.5 Procedure for Data Collection

The 26 paragraphs were presented to five participants: two men and three women. Multiple speakers were recorded as prosodic features are known to vary between individuals and men and women are known to have different pitches (Himmelman & Ladd 2008: 265). The participants' ages ranged from 42 to 60 and all were living in Bario at the time of the experiment. Four of the five spoke exclusively with a Northern dialect of Kelabit and one alternated between Northern and Southern pronunciations (see SUBSECTION 2.2.2). Participants were selected partly on account of availability but also to limit potential differences due to age or dialect. The circumstances in which they were produced were as similar as possible, following Himmelman (2006b: 169).

The participants were given the paragraphs in a randomised order as a written document with instructions to read each paragraph aloud with a short pause between

them (see APPENDIX 2).¹⁹⁸ Following Himmelmann & Ladd (2008: 265), reading was seen as a good method of eliciting the intonation of the example sentences without influencing the participants by having them repeat after the fieldworker or a native speaker research assistant. Each set of paragraphs was repeated twice by each speaker. This resulted in a total of 260 paragraph tokens: 26 paragraphs by five speakers by two repetitions.

In order to accurately analyse duration and pitch effects, it is necessary to ensure that these are not affected by hesitation or disfluency, which can also produce lengthening (cf. Cruttenden 1997). In order to prevent disfluency, the examples were printed in a large, clear font. Speakers were also given time to read through the sentences and familiarise themselves with the spelling system used. For this study, I used the spelling system outlined in SUBSECTION 2.3.1. Participants were given the opportunity to ask questions and repeat sentences. Similarly, in order to ensure that the recordings were of a sufficient quality for acoustic analysis, speakers were isolated from background noise as far as possible. They were recorded using an Audio Technica PRO70 Lavalier Microphone to further limit the effects of background noise and wind.

In a field setting, it is not always possible to limit these completely. Hence, the randomised list of paragraphs included each test sentence twice so that if one instance was affected by background noise or disfluency, the other could be selected for analysis. Ultimately, I selected one test sentence per context per speaker to include in the analysis. Any sentences where there was obvious disfluency or too much background noise were excluded. Otherwise, I selected one of the test sentences at

¹⁹⁸ The participants were not aware that the purpose of the experiment was to investigate the pronouns. Upon completion, I explained the specific focus and why many of the test paragraphs appeared almost identical.

random. Only one of the two examples was selected – even if both were of equal quality – so as not to bias the results towards the patterns of a particular speaker in case of significant interspeaker variation. Hence the final analysis involves 130 tokens.

4.4.6 Data Analysis

In order to establish whether the two pronouns in Kelabit behave like clitics, the example sentences selected were coded into syllables using Praat. Syllables were identified in Praat using auditory cues, the waveform of the utterance and changes in the formant values. However, as Fox (2000: 13) suggests, the process of segmentation is by no means straightforward. In this experiment, the main difficulty involved identifying where voiceless consonants began and ended, particularly when followed by another voiceless consonant. In this instance, the boundary was simply placed in the middle of the two consonants or in relation to visual changes in intensity or formants. Where voiceless consonants followed a pause, such as at the beginning of the intonation unit, the decision was made to mark the start of the consonant using visual changes in intensity or formants. Finally, it was sometimes difficult to establish whether a particular word was monosyllabic or bisyllabic. In this experiment, the following items were treated as mono-syllabic on the basis of their acoustic properties:

- (44) **Monosyllabic Items**
- a. pronouns, e.g. *uih* and *ieh*
 - b. particle + pronouns, e.g. *nuih* and *tieh*
 - c. the numeral *edteh* ‘one’, since the initial schwa was elided in speech.

Every decision taken in the segmentation was applied as consistently as possible and all annotation was conducted by the author.

After the syllable boundaries had been coded, measurements for duration and mean F_0 were taken for each syllable using the Praat Script Prosody Pro (Xu 2013). The pronoun syllable and the syllables immediately preceding and following were selected and coded for later analysis according to their position and context variables (SUBSECTION 4.4.3). The remaining syllables in contexts 3 and 4 were coded according to whether they represented a single syllable word, a non-final syllable in a multi-syllable word or a final-syllable in a multi syllable word.

To test whether duration and pitch changes are associated with boundary syllables in Kelabit, the remaining syllables were analysed using a one-way independent ANOVA. This explored the effect of syllable position – i.e. final or non-final – on the dependent variables duration and mean F_0 . Syllables that appeared before a pause and/or before an intonation phrase boundary were excluded from the analysis as they could show lengthening due to hesitation or position at a higher level boundary. Similarly, syllables for the particles *neh* and *teh* were not coded, as they show many of the characteristics of clitics discussed in SUBSECTION 4.2 and 4.3. Ultimately, analysis revealed significant differences in syllable duration, but not in pitch (SUBSECTION 4.5.1).

To test whether the FORM 1 and FORM 2 pronouns differ in their prosody, the pronoun syllable and the syllables immediately preceding and following were analysed. Firstly, they were analysed in the contexts in which only the form of the pronoun differs: contexts 3 and 4, and contexts 1 and 2. Since the test sentences were repeated by the same speakers, a repeated measures two-way ANOVA test was used to explore the main effects of syllable position and pronoun form and whether there is a significant interaction between the two in terms of syllable duration. Evidence for boundaries can be taken from the main effect of position on the dependent variable(s)

and evidence for whether or not the two pronouns differ in this respect can be taken from the interaction between position and pronoun type.

Finally, the prosodic behaviour of the FORM 1 pronoun was analysed across all of the contexts in which it occurs. A one-way ANOVA was conducted on the duration of the pronoun syllable, looking at the dependent variables of voice, function and position. This allows us to explore whether FORM 1 pronouns have a consistent prosodic behaviour, as is argued for NOM clitics in Philippine-type languages, or differ according to context, like the NOM pronouns in Kulawi.

4.4.7 Summary

In this section, I explored the basics of prosodic phonology and used these to build a prosodic definition of words and clitics. I discussed that boundaries between prosodic constituents are often marked by lengthening and pitch change and used these facts to formulate a set of predictions based on the assumption that clitics form part of the prosodic word of their host. I then outlined a method for testing these predictions which involved eliciting examples of the pronouns in different contexts. These were collected from multiple speakers and coded into syllables in order to measure the acoustic properties of the pronouns and how they compare to the immediate context. This allows us to examine whether or not there is evidence for word boundaries and thus whether the pronouns act as proclitics, enclitics or independent words. The following section summarises the results of the study.

4.5 Results

4.5.1 Word-boundary effects in Kelabit

The first hypothesis to explore was that word boundaries in Kelabit are marked by final-lengthening and pitch reset. The results suggest that final-syllables are significantly longer than non-final syllables. Comparing duration using an independent t-test, final-syllables are significantly longer than non-final syllables, $t(322) = -10.72$, $p < 0.01$. This represents a large effect size, $r = .51$. The mean lengths of non-final vs final syllables are shown in TABLE 4.17 and the boxplots in FIGURE 4.1:

Table 4.17 Word-final Lengthening in Kelabit

	Number	Mean duration	Standard deviation	standard error
non-final	197	168.24	70.61	5.03
final	127	258.67	69.60	6.17

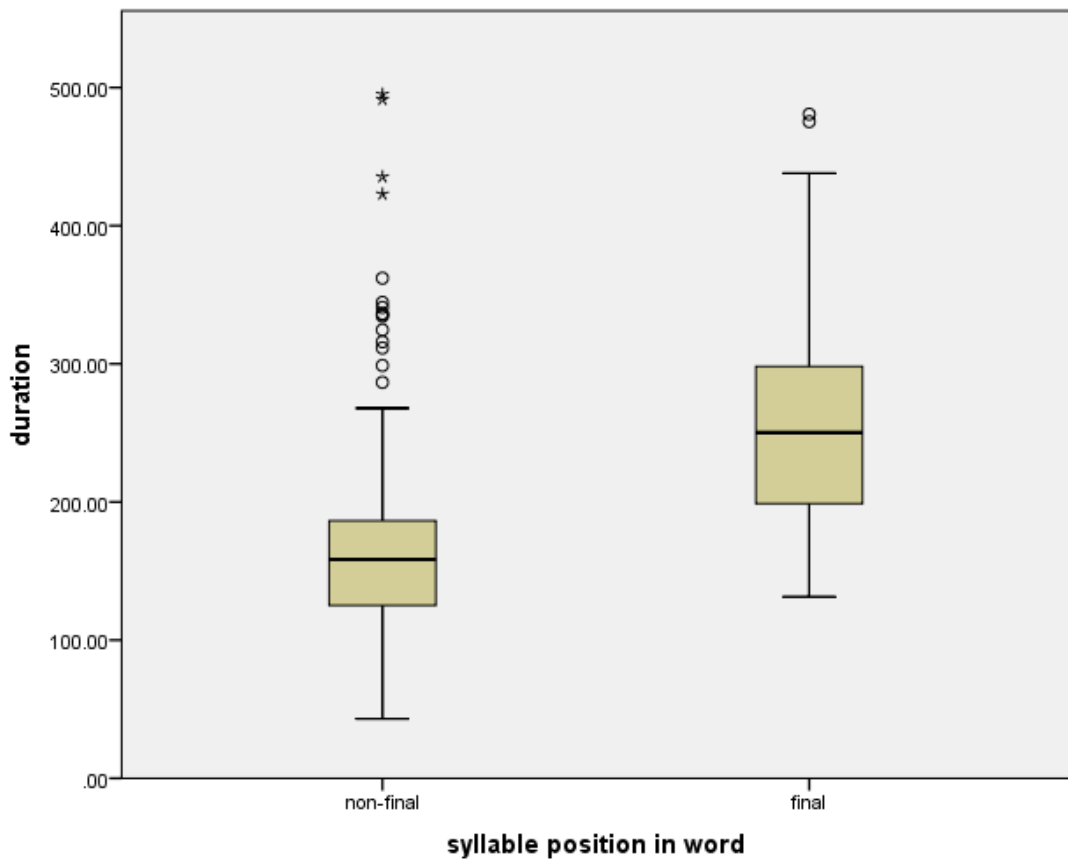


Figure 4.1 Boxplot of Duration by Syllable Position

To rule out the possibility that lengthening effects are only seen in multisyllabic words, the effect of syllable position on duration was also analysed when syllables were subcategorised into mono-syllabic words, non-final syllables of multisyllabic words, and final syllables of multisyllabic words. A one-way independent ANOVA also reveals a significant effect of syllable position on duration when single syllable words are distinguished from other final-syllables, $F(2, 321) = 60.5$, $p < 0.01$, $\omega^2 = 0.27$.

Planned contrasts revealed that both final syllables of a multi-syllabic word and final syllables in a single syllable word are significantly longer than non-final syllables, $t(321) = 10.53$, $p < 0.01$, $r = 0.51$. Final syllables of a multi-syllabic word are significantly longer than non-final syllables, $t(321) = 10.12$, $p < 0.01$. This is a

fairly large effect, $r = 0.49$. Similarly, single syllable words are also significantly longer than non-final syllables, $t(321) = 6.67$, $p < 0.01$, though this is a medium effect, $r = 0.35$. Interestingly, planned contrasts also revealed a significant contrast between single syllable words and final syllables, $t(321) = 2.18$, $P < 0.05$. Though this is a very small effect, $r = 0.12$, it could be potentially important in the analysis of Kelabit pronouns since they are monosyllabic.¹⁹⁹

The descriptive statistics for the three syllable types are shown in TABLE 4.18 and FIGURE 4.2:

Table 4.18. Syllable Duration

	Number	Mean duration	Standard deviation	Standard error
Multiple-syllable non-final	197	168.24	70.61	5.03
Single-syllable final	56	238.67	62.21	8.31
Multiple-syllable final	71	265.70	73.13	8.68

¹⁹⁹ Nb. when using either a Gabriel or Hochberg's GT2 post hoc test, which are designed to cope with sample sizes that are different, the contrast between single word syllables and final syllables is not significant.

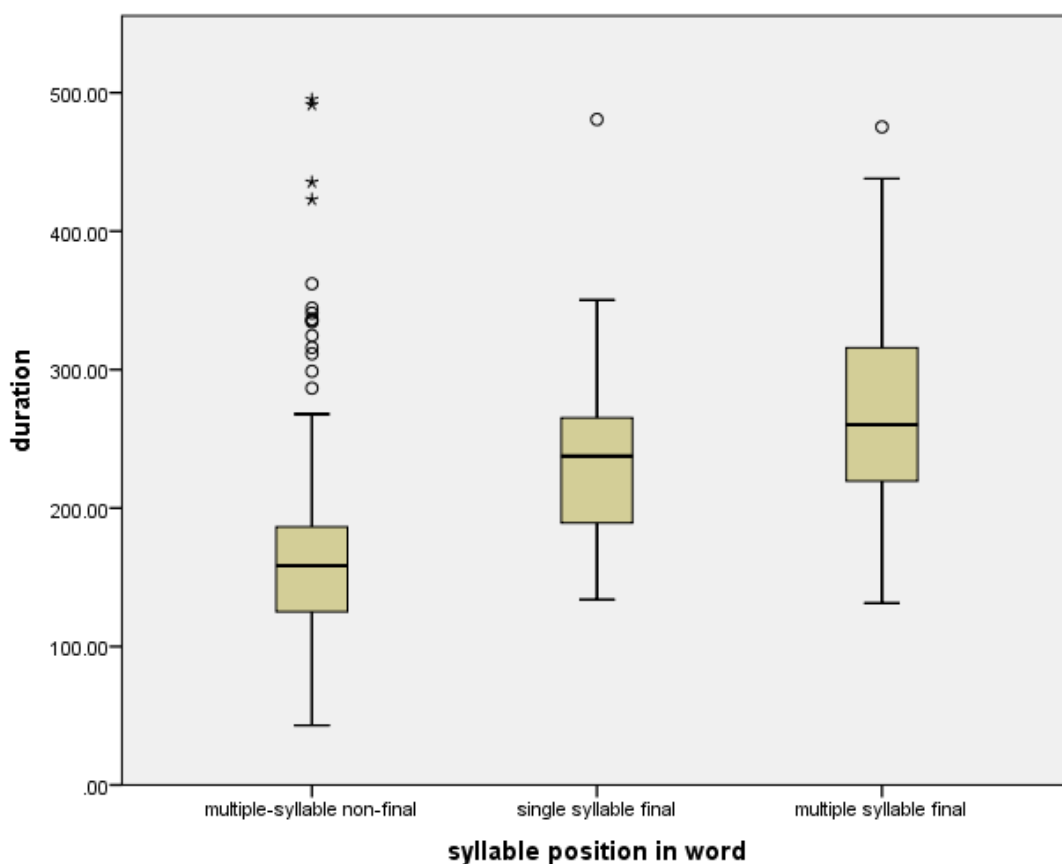


Figure 4.2 Boxplot of Duration by Syllable Position, separating Single Syllable Words

In terms of pitch, however, there are no significant effects of syllable position, regardless of whether two groups or three groups are compared. An independent t-test of mean F_0 in final versus non-final syllables revealed no significant effects, $t(305) = 1.273$, $p > 0.05$. The effect size is also minimal, $r = .07$. Similarly, a one-way independent ANOVA of mean F_0 in non-final syllables, monosyllabic words and final syllables of multisyllabic words, reveals no significant effect, $F(2, 304) = 1.02$, $p > 0.05$ and indeed no effect size, $\omega^2 = 0.00$. The descriptive statistics are summarised in TABLES 4.19 and 4.20 and FIGURE 4.3 and 4.4:

Table 4.19 Word-final Pitch effects in Kelabit

	Number	Mean F ₀	Standard deviation	Standard error
Non-final	180	143.94	68.24	5.09
Final	127	144.58	55.90	4.96

Table 4.20 Mean F₀

	Number	Mean F ₀	Standard deviation	Standard error
Multiple-syllable non-final	180	153.94	68.24	5.09
Single-syllable final	56	140.50	52.95	7.08
Multiple-syllable final	71	147.80	58.30	6.92

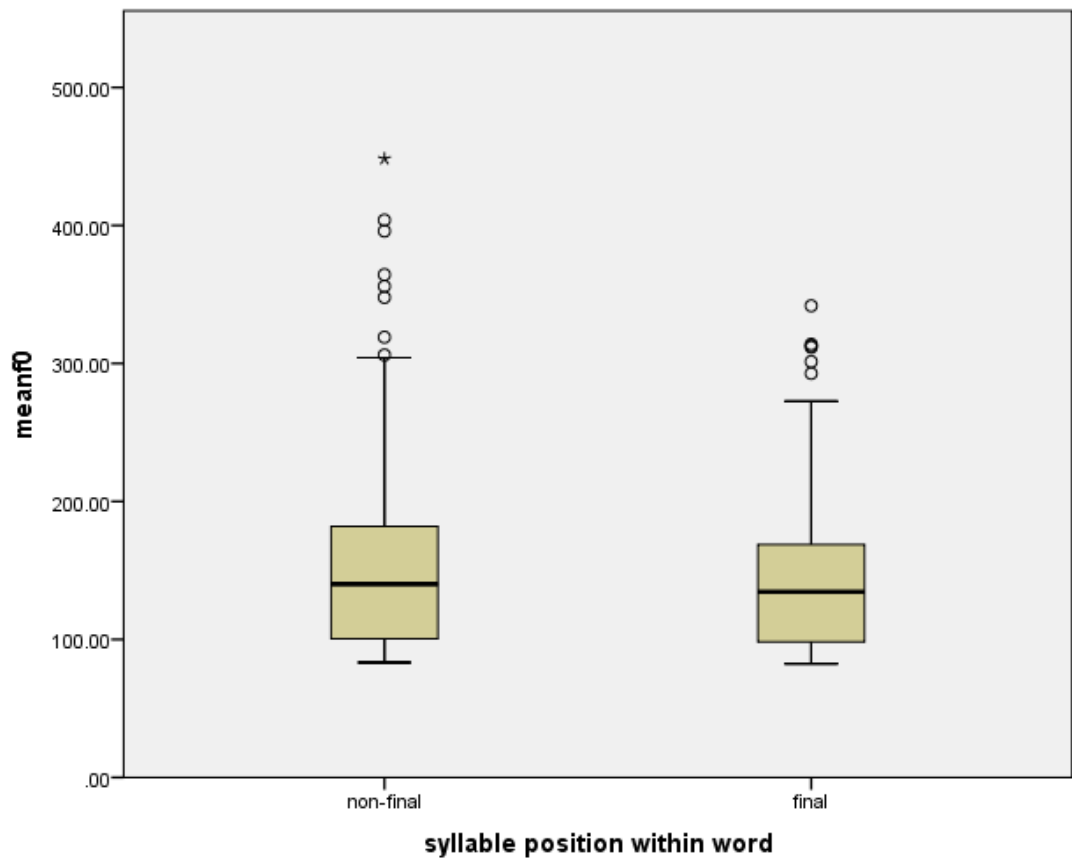


Figure 4.3 Boxplot of mean F₀ by Syllable Position

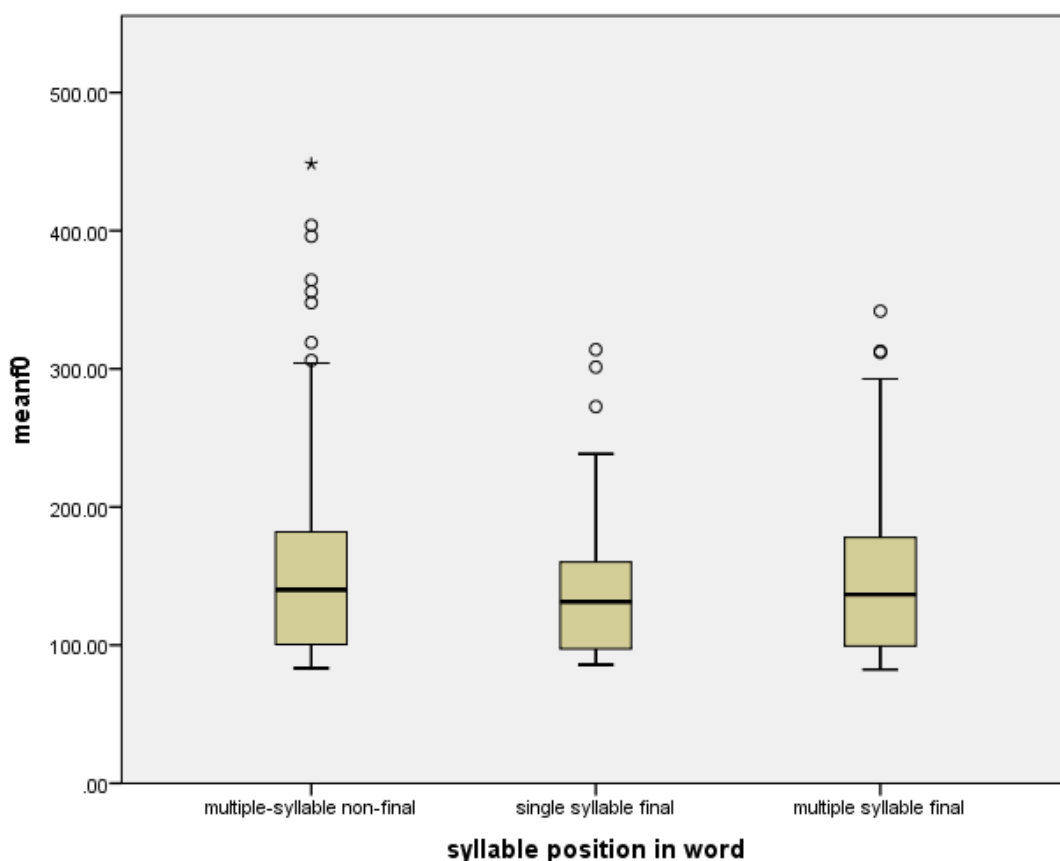


Figure 4.4 Boxplot of mean F_0 by Syllable Position, separating Single Syllable Words

Consequently, I conclude that word-boundaries in Kelabit are demarcated by final-lengthening but not pitch reset. Pre-boundary or word-final syllables are significantly longer than non-final syllables but have no significant pitch differences. For this reason, we will investigate the clitic status of Kelabit pronouns using length as a marker of word boundaries.

4.5.2 Comparing FORM 1 and FORM 2

4.5.2.1 Transitive Predicates

In order to explore the prosody of Kelabit pronouns, the pre-pronoun, pronoun and post-pronoun syllables were compared in the 40 test sentences in Context 3 and

Context 4. These represent the main contexts in which both pronouns are found, namely as non-subject actors in UV transitive clauses (SUBSECTION 4.2). In all of the example sentences in these contexts, the post-pronoun syllable is a mono-syllabic word. Hence, we would expect the following patterns for enclitics, proclitics and free pronouns:

- (45) a. **Enclitic pronoun**
- | | | | |
|----------------|--------------------|--|-----------------|
| σ_{pre} | $\sigma_{pronoun}$ | | σ_{post} |
| <i>short</i> | <i>long</i> | | <i>long</i> |
- b. **Proclitic pronoun**
- | | | | |
|----------------|--|--------------------|-----------------|
| σ_{pre} | | $\sigma_{pronoun}$ | σ_{post} |
| <i>long</i> | | <i>short</i> | <i>long</i> |
- c. **Free pronoun**
- | | | | | |
|----------------|--|--------------------|--|-----------------|
| σ_{pre} | | $\sigma_{pronoun}$ | | σ_{post} |
| <i>long</i> | | <i>long</i> | | <i>long</i> |

The results, given in TABLE 4.21, TABLE 4.22 and FIGURE 4.5, show that the average duration of syllables is long – short – long in both contexts. This suggests that syllable length varies according to syllable position, but not according to which pronoun is used. In other words, it suggests that both pronouns are proclitics.

Table 4.21 Raw Duration (ms) of Syllables in Context 3 & 4

speaker	sentence	Pre-pronoun		Pronoun		Following	
		uih	kuh	uih	kuh	uih	kuh
1	1	264.15	305.52	228.70	186.34	289.44	340.05*
	2	291.07	247.45	173.83	139.22	491.43	435.40
	3	165.23	203.82	511.16*	177.65	n/a	354.33*
	4	186.14	222.13	202.96	148.73	336.06	480.74*
2	1	301.80	291.56	166.22	188.42	250.18	331.39
	2	255.20	344.96	139.56	164.78	251.22	286.59
	3	248.61	218.13	214.81	175.48	345.45*	185.42
	4	272.15	223.92	186.73	180.55	262.04	310.95
3	1	341.41	297.39	387.98	219.31	314.18	268.25
	2	316.32	256.37	95.05	141.74	311.47	211.90
	3	196.09	202.64	272.87*	355.26*	n/a	n/a
	4	141.82	215.68	159.59	153.29	418.44*	517.18*
4	1	280.20	320.42	173.82	144.15	247.25	246.38
	2	319.47	223.10	115.08	128.55	221.63	217.89
	3	157.78	200.63	211.63*	206.90	n/a	261.31
	4	186.61	183.86	159.89	169.06	335.67	257.90
5	1	308.11	311.12	181.35	140.58	243.94	174.65
	2	350.36	234.01	117.13	116.48	495.16	422.87
	3	223.93	218.70	212.75	187.18	245.81	149.47
	4	159.07	178.83	228.78	121.73	226.43	198.65

*followed by a pause

Table 4.22 Mean Duration (ms) of syllables in Context 3 & 4

		Number	Mean duration	Standard deviation	Standard error
pre-pronoun syllable	uih	20	248.28	66.61	14.89
	kuh	20	245.01	49.30	11.02
pronoun syllable	uih	20	206.99	95.61	21.38
	kuh	20	172.27	51.33	11.48
post-pronoun syllable	uih	17	310.73	86.07	20.88
	kuh	19	297.44	106.04	24.33

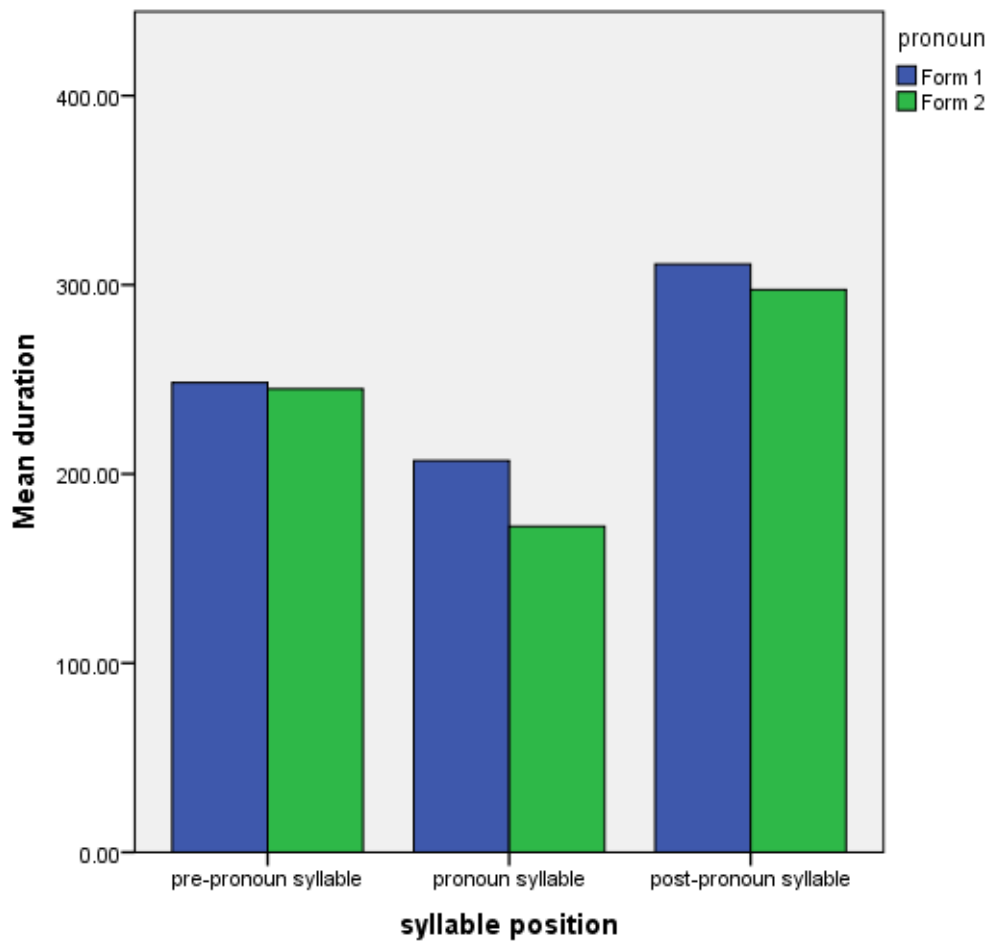


Figure 4.5 Mean Duration of Syllables in Context 3 & 4

When compared with the mean duration measurements for non-final syllables (168.24 ms) and final-syllables (mono-syllabic 238.67 ms and multi-syllabic 265.70 ms) in SUBSECTION 4.5.1, the pre-pronoun and post-pronoun syllables in TABLE 4.22 and FIGURE 4.5 seem to have the characteristics of final (long) syllables in both contexts.²⁰⁰ In contrast, the pronouns have the characteristics of non-final (short) syllables. Moreover, though FORM 1 pronouns are generally longer than FORM 2, this is not statistically significant (see SUBSECTION 4.5.2.1.1) and perhaps simply reflects

²⁰⁰ In fact, the post-pronoun syllable appears even longer than the pre-pronoun syllable. It is possible that this results from the post-pronoun syllable occurring at the edge of a prosodic phrase in the example sentences used and could be controlled for in future research.

the different phonological make-up of the two pronouns. Hence, both pronouns appear to have the pattern of proclitics represented in (45b).

Before looking at the results of the statistical analysis, it is worth reflecting on characteristics of the raw data. As can be seen from TABLE 4.21, there are quite a few cases where test syllables are followed by pauses – either within the data set or immediately following. It is difficult to interpret what these pauses mean as we wouldn't expect pauses following a proclitic element. On the one hand, they may reflect prosodic phrase boundaries following the post-pronoun syllable, or alternative prosodic structures following the pronoun (see SUBSECTION 4.6.1 for discussion of Kelabit pronouns as potentially undetermined between proclitic and enclitic and note that the pronouns do appear lengthened in these instances). On the other hand, they could signal disfluency or hesitation – both of which affect syllable duration. A possible source of disfluency comes from the written form. For example, consider sentence 3, which prompts a large number of pauses in TABLE 4.21. The test sentences with the FORM 1 pronoun *uih* and the FORM 2 pronoun *kuh* are repeated below as they were written in the stimulus materials:

(46) a. **Context 3, Sentence 3**

Seni'er uih teh ieh tudo sebuleng.

UV.PFV.see 1SG.1 PT 3SG.1 sit alone

'I saw him sat alone.'

(e.g. experiment, BAR21082014CH_10 00:08:55.683-00:08:58.585)

b. **Context 4, Sentence 3**

Seni'er kuh ieh tudo sebuleng.

UV.PFV.see 1SG.2 3SG.1 sit alone

'I saw him sat alone.'

(e.g. experiment, BAR21082014CH_10 00:06:59.808-00:07:01.888)

If we compare (46a) and (46b), the segmental context differs in that (46a) includes the particle *teh*, since it would be ungrammatical or at least unnatural without it

(SUBSECTION 4.2.2). The particle is optional in (46b) and is therefore omitted in this particular sentence. Moreover, the particle *teh* is written as a free-standing word, although it is most natural to pronounce *teh ieh* as [tiyəh]. This could have prompted disfluency, if speakers were trying to decide whether to read the sentence as it was written or as it is more commonly pronounced. Alternatively, there may be interesting prosodic phrasing differences depending on whether *teh* is pronounced as a separate word/syllable or as *tieh*, forming a monosyllabic prosodic word with the pronoun.²⁰¹ (46b) is also the only instance where the two pronouns could potentially form a clitic cluster. It is possible that there are two options in a sentence like (46b): either to treat FORM 1 as the host for the FORM 2 clitic – or to form a clitic cluster with both pronouns attached to the following host word. These may both be options available to all speakers or represent different stages in a language change process. In any case, all of these factors could potentially have affected the length of the syllable.

Similarly, the length of the syllables may also have been affected by accenting, in addition to the position. This seems to be the case for speaker 1 for the post-pronoun syllable of sentence 2:

- (47) **Context 3, Sentence 2**
 a. [Senibu uih dooq~dooq] [neh latiq tauh].
 UV.PFV.plant 1SG.1 REDUP~good PT farm 1PL.INCL
 V N ADV N
 ‘I planted my farm well.’
 (experiment, BAR18082014CH_03 00:01:34.660-00:01:37.740)

This is perhaps linked to the fact that the word immediately following the pronoun in (47) is an adverb, rather than the undergoer argument, as in the rest of the sentences. The particular word order may lend itself to emphasising or prosodically highlighting

²⁰¹ Speakers varied as to whether they read *teh ieh* as one syllable or two.

the adverb (see SUBSECTION 5.5.3). This could well explain additional lengthening on the first syllable of *dooq-dooq* that might not have been predicted.

In summary, it is important to bear in mind that the syllable durations in the test sample may also have been affected by accentual lengthening, lengthening due to higher prosodic boundaries and/or lengthening due to hesitation. Indeed, when I remove the four cases of pronoun syllables immediately before pauses in the test set – as the four cases most likely to be affected by disfluency – then the differences between FORM 1 and FORM 2 all but disappear, as shown in TABLE 4.23.

Table 4.23 Mean duration of syllables in Context 3 & 4 – possible errors removed

		Number	Mean duration	Standard deviation	Standard error
pre-pronoun syllable	uih	20	248.28	66.61	14.89
	kuh	20	245.01	49.30	11.02
pronoun syllable	uih	17	184.95	65.46	15.88
	kuh	19	162.64	28.69	6.58
post-pronoun syllable	uih	17	310.73	86.07	20.88
	kuh	19	297.44	106.04	24.33

4.5.2.1.1 Statistical Analysis

Conducting a two-way repeated measures ANOVA removes cases of missing data such that an equal sample size is compared across each category. As such, the problematic data discussed above were not included in the analysis. Mauchly's test revealed that the assumption of sphericity held. The results find a significant main effect of syllable position on the duration of the syllables, $F(2, 32) = 16.69, p < 0.01$. Contrasts reveal that the pre-pronoun syllable is significantly longer than the pronoun syllable $F(1, 16) = 30.96, p < 0.01, r = 0.66$ and the post-pronoun syllable is also significantly longer than the pronoun syllable, $F(1, 16) = 25.99, p < 0.01, r = 0.62$.

Both of these effect sizes are large.²⁰² However, there is no significant effect of pronoun type, $F(1, 16) = 2.42$, $p > 0.05$ and importantly no significant interaction between position and pronoun type $F(2, 32) = 0.37$, $p > 0.05$. This indicates that both pronouns are significantly shorter than either the preceding or the following syllable but that the prosody or syllable duration of the test sentences does not differ depending on whether *uih* or *kuh* is articulated. Hence, statistical analysis supports the conclusion in SUBSECTION 4.5.2.1 that both pronouns may be proclitics in this context.

4.5.2.2 Intransitive predicates

To test if the same results are found in different contexts where both FORM 1 and FORM 2 pronouns alternate, the same experiment was conducted for context 1 and 2, where the pronouns are used with intransitive predicates. In this case, the post-pronoun syllable is not a single-syllable word (as was true of the previous experiment) but rather the initial syllable of the following word. This can be illustrated in (48):

- (48) **Context 2, Sentence 3**
 a. Kadiq keliq kuh malem neh nuk midih sineh.
 but know 1SG.2 before PT REL INTR.present DEM
 ‘But I’ve known that for ages.’
 (experiment, BAR18082014CH_03 00:01:47.380-00:01:50.340)

In (48), the pre-pronoun syllable *liq* and post-pronoun syllable *ma* are underlined. Unlike *dooq* ‘good’ in context 3, which might be considered a mono-syllabic word in the post-pronoun position, *ma* is simply the first syllable in a multi-syllabic word *malem* ‘before’.²⁰³ Hence, we expect the post-pronoun syllable in these examples to

²⁰² Nb. the post-pronoun syllable is not significantly longer than the pre-pronoun syllable, $F(1,16) = 3.09$, $p > 0.05$, $r = 0.16$.

²⁰³ It is not clear if the reduplicated form *dooq-dooq* is one prosodic word or two. This could also be amended if the experiment was repeated.

be short rather than long. Consequently, if the pronouns are proclitics in these contexts then we expect them to be a similar length to the following syllable. This can be schematised as follows:

- (49) a. **Proclitic pronoun in Context 1 & 2**
- | | | | |
|-----------------------|--|---------------------------|------------------------|
| σ_{pre} | | σ_{pronoun} | σ_{post} |
| <i>long</i> | | <i>short</i> | <i>short</i> |

Importantly, (49) is exactly the pattern that we find, as seen in TABLE 4.24 and FIGURE 4.6:

Table 4.24 Mean Duration of Syllables in Context 1 & 2

		Number	Mean duration	Standard deviation	Standard error
pre-pronoun	uih	13 ²⁰⁴	247.67	68.13	18.90
syllable	kuh	13	253.46	67.03	18.59
pronoun syllable	uih	13	190.01	57.32	15.90
	kuh	13	170.22	55.94	15.51
post-pronoun	uih	13	180.68	33.85	9.39
syllable	kuh	13	170.04	38.68	10.73

²⁰⁴ Two cases are removed from the data set as the pronoun is either preceded or followed by a pause.

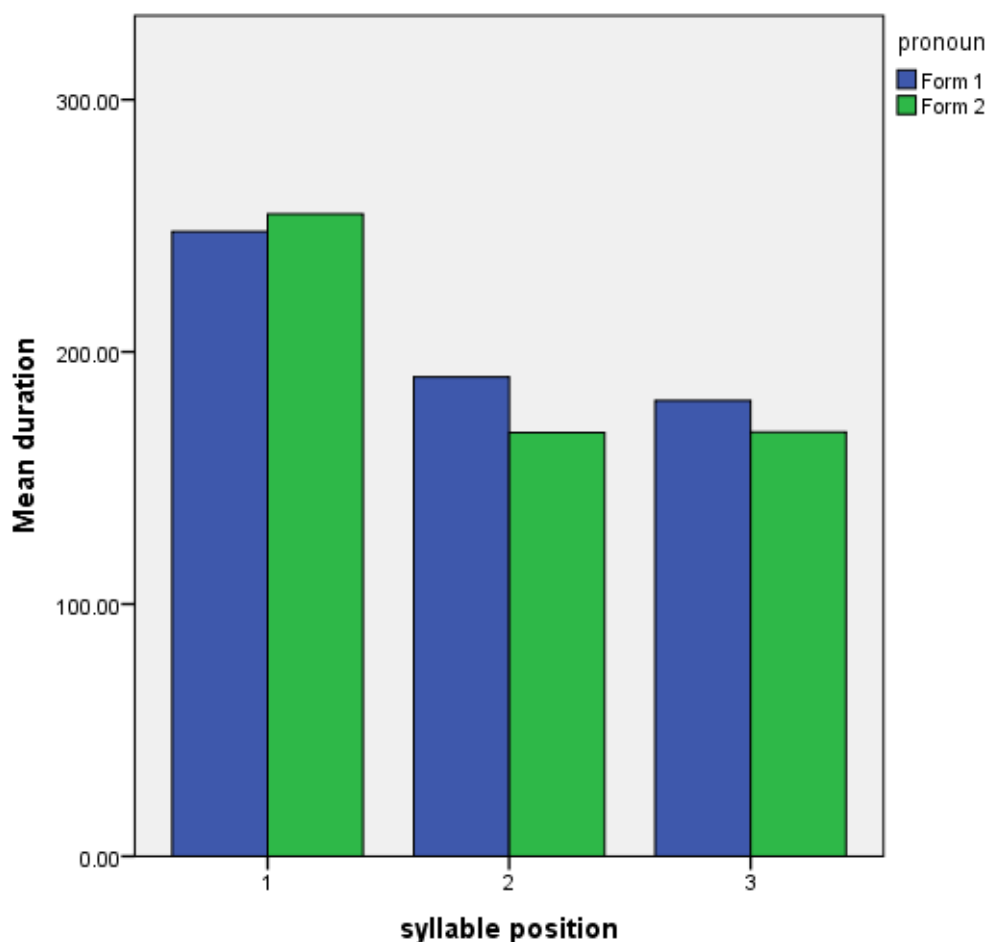


Figure 4.6 Mean Duration of Syllables in Context 1 & 2

Hence, the two pronouns do not seem to differ in their prosodic behaviour in these contexts either and the results support an analysis of both pronouns as proclitics.

4.5.2.2.1 Statistical Analysis

Much as in SUBSECTION 4.5.2.1.1, this is confirmed in the statistical analysis. Mauchly's test of sphericity is non-significant meaning that the assumption of sphericity can be held. There is a significant effect of syllable position, $F(2,24) = 8.208, p < 0.01$, but no significant effect of pronoun type and no significant interaction between syllable position and pronoun type. In both cases, contrasts reveal that the pre-pronoun syllable is significantly longer than the post-pronoun syllable, $F(1, 12) =$

9.608, $p < 0.01$ and the pronoun syllable, $F(1,12) = 11.51$, $p < 0.01$. However, there is no significant difference between the pronoun and post-pronoun syllable, $F(1,12) = 0.82$, $p > 0.05$. Hence, we find exactly the pattern predicted in (49). Consequently, the two pronouns appear to be prosodically proclitic in both contexts in which the two pronouns alternate. This leads to the question of whether FORM 1 is also proclitic in the other environments in which it occurs, which I address in SUBSECTION 4.5.3.

4.5.3 Comparing FORM 1 in all Contexts

In the previous sections, I established that both FORM 1 and FORM 2 pronouns appear to be prosodically proclitic in the contexts analysed so far. This is an interesting result given the distribution of the pronouns in SUBSECTION 4.2.3. If the pronouns are proclitics, then how do we explain the fact that they have the same distribution as full NPs? Equally, if they behave in the same way as FORM 2 clitics then why is it that they can occur utterance-initially when FORM 2 pronouns cannot?²⁰⁵ Moreover, how do we explain that they precede the negation and can be the focused element, in contrast to FORM 2?

It seems that there are perhaps two possible answers. Firstly, the FORM 1 clitics could be simple clitics rather than special verb-adjacent or Wackernagel clitics. Simple clitics, by definition, have the same distribution as non-clitic elements. The most likely candidate, in this case, for its non-clitic counterpart would be the emphatic pronoun *keduih* (SUBSECTION 2.4.2.8.3).

Another possibility is that FORM 1 pronouns are in the process of degrammaticalising from clitic to free word.²⁰⁶ Degrammaticalisation, though

²⁰⁵ Note that it is not unexpected that FORM 1 pronouns would occur initially given a proclitic analysis. Rather, it is strange that FORM 2 pronouns wouldn't.

²⁰⁶ Or that previously free-standing words are grammaticalising into clitics on analogy with FORM 2. I am inclined to assume the direction of change is from clitic to word (i.e. degrammaticalisation) given

unexpected from the grammaticalisation pathway in SUBSECTION 4.3.2, has been reported in the literature (Haspelmath 2004, Campbell 2001). It is seen, for example, in the development of the genitive *-s* in English and Swedish which is claimed to develop into a clitic from a case marking affix (cf. Norde 2006). If the pronoun is becoming a free-standing item, then the fact that it has the same distribution as full NPs is no longer surprising. Equally, this could explain why the FORM 1 pronouns are, on average, longer than the FORM 2 pronouns.

To test whether the FORM 1 pronoun is always proclitic or sometimes differs in its prosodic behaviour, we can compare the duration of the pronoun syllable in all the contexts in which it occurs. If FORM 1 pronouns are always clitics, then we expect there to be no significant effect of context on syllable duration. If, however, the pronouns are sometimes clitics and sometimes free-standing, then we expect the pronouns to be significantly longer when they are free-standing, since we would expect them to be lengthened like non-pronominal mono-syllabic words. Conducting a one-way ANOVA reveals a significant effect of context on the duration of the pronoun syllable, $F(8,121) = 4.113$, $p < 0.01$. Contrasts further reveal that the pronouns in contexts 7-9 are significantly longer than those in context 1-6 ($t = 4.916$, $p < 0.01$). However, within these groups there is no significant difference in the duration of the pronoun syllable. Hence, we might conclude that the pronouns in contexts 1-6 are clitics, whilst the pronouns in contexts 7-9 are free-standing forms. The descriptive statistics can be seen in TABLE 4.25.²⁰⁷

that cognate forms in the more conservative Philippine-type systems are said to be clitics, whilst cognate forms in the more innovative Indonesian-type languages are said to be free-standing (SUBSECTION 4.3.2).

²⁰⁷ The figures in TABLE 4.25 include all of the raw data. As discussed in SUBSECTIONS 4.5.2.1 and 4.5.2.2, some of these pronouns are followed by pauses, which could affect the duration of the syllable. However, even if cases of pronouns followed by pauses are removed from the data set, there is still a significant effect of context of syllable duration, $F(8,112) = 5.075$, $p < 0.01$, and contrasts still reveal a significant difference between contexts 1-6 and contexts 7-9, $t = 5.735$, $p < 0.01$.

Table 4.25 Mean Duration of Pronouns by Context

Context	Number	Mean Duration	Standard Deviation	Standard Error
1	15	211.53	79.87	20.62
2	15	173.43	56.46	14.57
3	20	206.99	95.61	21.38
4	20	172.27	51.33	11.48
5	10	156.42	39.83	12.60
6	15	186.19	52.37	13.52
7	15	261.09	73.80	19.05
8	5	273.55	116.23	30.01
9	15	275.34	145.44	65.04

If FORM 1 can be both clitic and free-standing, the question arises of what determines the prosodic status of the pronoun. In other words, what do contexts 1-6 have in common, that contexts 7-9 do not have? In order to investigate this, I split each context into the variables described in SUBSECTION 4.4.3, namely, pronoun form, voice, pronoun function and pronoun position. When the four variables are added into the model, there is a significant effect of position, $F(2, 121) = 6.24$, $p < 0.01$, but no significant effects for pronoun type (as seen in the previous sections), pronoun function or voice. The mean duration according to position, function and voice are shown in TABLES 4.26, 4.27 and 4.28:

Table 4.26 Mean Duration according to Position

Position	Number	Mean Duration	Standard Deviation	Standard Error
Post-verbal	95	186.49	68.71	7.05
Post-object	10	261.09	73.80	19.06
Preverbal	15	274.00	120.03	26.84

Equally, in some cases, the pronoun in contexts 7-8 was realised as *nuih* (neh+uih) or *tuih* (teh+uih). Clearly, the extra consonant would affect the duration of the pronoun. However, even when these cases are removed, there is still a significant effect of context on syllable duration, $F(8,100) = 2.900$, $p < 0.01$, and contrasts still reveal a significant difference between contexts 1-6 and 7-9.

Table 27. Mean Duration according to Function²⁰⁸

Function	Number	Mean Duration	Standard Deviation	Standard Error
Subject	75	223.13	94.09	10.87
Non-subject	55	188.69	71.28	9.61

Table 4.28 Mean Duration according to Voice

Voice	Number	Mean Duration	Standard Deviation	Standard Error
Intransitive	30	192.48	70.67	12.90
Transitive	45	214.45	83.99	12.52
AV Transitive	55	212.51	96.40	13.00
UV				

Planned contrasts reveal that the pronouns in the immediately post-verbal position are significantly different from both the pronouns in the preverbal position and those in the post-object position. These are not, however, significantly different from one another. In other words, there appears to be a clitic position immediately following the verb (cf. Billings 2005 for similar suggestions in relation to Tagalog). Pronouns in the immediately post-verbal position are clitics, regardless of whether the clause is AV or UV and regardless of whether the pronoun fulfils a subject or non-subject function.

This has two implications: firstly, it supports the analysis of Kelabit pronouns as verb-adjacent, rather than second-position clitics, as proposed in SUBSECTION 4.3.3. Secondly, it suggests that the FORM 1 pronouns are sometimes clitics and sometimes free-standing, like the Kulawi equivalents in SUBSECTION 4.3.2. This seems to be the

²⁰⁸ The difference between the mean durations for subjects and non-subject functions reflects the fact that in the three contexts in which the FORM 1 pronoun is free-standing, it is subject. However, this is not a significant factor on its own, since it is also possible to find clitic subjects in AV, i.e. in context 5. A similar explanation holds for the difference between duration in intransitive clauses vs the two transitive types. Of the three contexts in which FORM 1 is non-clitic, two occur in AV and one in UV, but none in intransitive clauses. It is presumably possible to find non-clitic pronouns in intransitive clauses, since it would also be grammatical for the subject pronouns to occur pre-verbally (see CHAPTER 5). This could be tested in future research.

most important distinction between FORM 1 and FORM 2. That is, it is not the case that FORM 2 is a clitic and FORM 1 is not, as hypothesised in SUBSECTION 4.2.3. Neither is it the case that FORM 1 pronouns are second-position enclitics and FORM 2 pronouns verb-adjacent, as could have been concluded on the basis of the syntactic behaviour (SUBSECTION 4.3.3). Rather, prosodically the FORM 2 pronouns are always clitics and restricted to the post-verbal clitic position, whilst the FORM 1 pronouns can be realised as free-standing words in other positions in the clause.

4.5.4 Summary

In summary, this section has demonstrated that Kelabit has word-final syllable lengthening and used this prosodic cue to investigate the behaviour of the two pronouns. The first test involved measuring duration in the two pronouns and their immediately surrounding syllables in contexts (both transitive and intransitive) in which only the pronoun differed. The results suggest that both pronouns behave as prosodic proclitics in these environments. The second test involved comparing the duration of the pronoun in all nine of the test contexts explored. This revealed that whilst FORM 2 pronouns are always prosodically clitic, FORM 1 pronouns can also be free-standing words. Moreover, it revealed that the key factor affecting clitic status seems to be position, rather than the pronoun form and function, or the voice of the clause in which it occurred. Thus, it can be concluded that Kelabit has a clitic position immediately following the predicate in which the pronoun attaches proclitically to the following prosodic word. The following section discusses the implications of these results in light of the Austronesian clitic phenomena discussed in SUBSECTION 4.3.2.

4.6 Discussion

In the final section, I explore the implications of the finding that Kelabit pronominal clitics are proclitic when attached to a verbal host. Firstly, I discuss the finding in relation to the prosody-syntax interface. Secondly, I discuss the implications of both analyses for the typology of Austronesian clitics and historical development.

4.6.1 Prosody-syntax mismatch

That the pronouns behave prosodically as proclitics is interesting given the fact that they form syntactic units with the words immediately preceding them. This was shown in SUBSECTION 4.3.3 and can be seen in (27), repeated as (50):

- (50) *Kelabit*
- a. [Senatek kuh]_{VP} neh bubpuq ih.
UV.PFV.close 1SG.2 PT door PT
'I closed the door.'
(elicitation, BAR17102013CH_01 00:53:02.138-00:53:04.854)
- b. Edteh laak ieh [ruyung kuh]_{PP}.
one year 3SG1 with 1SG.2
'he was with me for one year.'
(text, BAR04092014CH_02 00:00:56.850-00:00:58.410)

Hence for Kelabit pronouns there seems to be a mismatch between syntactic and prosodic constituents.

That the phonology and the syntax can form different constituents is increasingly recognised cross-linguistically.²⁰⁹ For example, a prosody-syntax mismatch can be seen in Kwakwala, where an enclitic, such as the possessor *is*, can be prosodically attached to the left, whilst functioning syntactically to the right:

²⁰⁹ See also Payne (1983) on clitics in the Yagua language of Peru.

- (51) *Kwakwala* (Wakashan)
- a. Nanaqəsil-ida iʔgəl'wat-I əlewinux^wa-s=is mestuw-i la-xa
 guides-ART expert-ART hunter-INS=his harpoon-ART PREP-ART
- mig^wat-i.
 seal-ART
- ‘An expert hunter guides the seal with his harpoon.’
 (Anderson 1992: 18)

Indeed, similar patterns are also attested in other Austronesian languages, such as Central Bontok, where phrase-marking clitics can attach enclitically to the previous word but modify morphosyntactically the following one:

- (52) *Central Bontok*
- a. In-manok nan babái=[_{PP} S nan masdem].
 AV-chicken PT woman=LOC PT night
- ‘The woman performs a chicken sacrifice [_{PP} in the evening].’
 (Reid 1970: 23)

Hence, a prosody-syntax mismatch is not, in itself, surprising.

However, given a proclitic analysis, it becomes difficult to account for examples like (8b), repeated as (53), where the pronoun appears utterance finally:²¹⁰

- (53) *Kelabit*
- a. **FORM 2**
 Na'am keliq kuh.
 NEG know 1SG.2
- ‘I don't know.’ (elicitation, fieldnotes)

In these instances, there is nothing to which the pronoun can attach proclitically. Therefore, they are problematic under a proclitic analysis. Similarly, a proclitic analysis does not help us to explain why the FORM 2 pronouns are ungrammatical

²¹⁰ It is not known how often such examples do occur in naturalistic data without a clausal complement.

utterance initially (SUBSECTION 4.2.3) and why the indirect pronoun *ngekuh* seems to involve enclitic attachment.

One possibility is that the pronouns are clitics that attach either enclitically or proclitically. This is in keeping with the particles *teh* and *neh*, which are strong candidates for behaving as prosodic clitics in Kelabit (SUBSECTION 2.4.2.14.1). They often seem to act as proclitics in that they combine with pronouns:

- (54) a. *teh* + *ieh* → *tieh*
 b. *neh* + *uih* → *nuih*

However, it is also possible to find the particles utterance-finally (see SUBSECTION 2.4.2.14.1):

- (55) **Particles Utterance-finally**
 a. Kapeh~kapeh peh dooq teh.
 REDUP~how PT good PT
 ‘However it comes is good.’ (elicitation, fieldnotes)

Thus, it may well be the case that all clitics in Kelabit are underdetermined as either proclitic or enclitic. Indeed, it may be the case that clitics in other Austronesian languages – which are typically analysed from a syntactic rather than prosodic perspective – also demonstrate both proclitic and enclitic attachment properties.

Beyond prompting further research into whether or not the clitic pronouns and particles can be enclitic in given environments, we must also address the question of why the clitics cannot occur utterance-initially if they are – at least sometimes – to be analysed as proclitics. Here, we may draw a comparison with Slovenian clitics, which Marušič (2008) analyses as being Wackernagel-type clitics that are unspecified, i.e.

either enclitics or proclitics.²¹¹ Though the pronouns are often enclitic, like the related Wackernagel clitics in Serbo-Croatian (see SUBSECTION 4.3.1), they can also be used proclitically:

- (56) *Slovenian*
- a. Kdo – za boûjo voljo – [ti je razbil avto]?
 who – for God’s sake – [you.DAT is ruin car]
 ‘Who, for God’s sake, ruined your car?’
 (Golden & Sheppard 2000)

In (56), the clitic cluster *ti je* appears directly after a pause and consequently must be proclitic in the same way that (53) is arguably forced to be analysed as enclitic.

In order to explain the complicated patterns of clitic placement in Slovenian, Marušič (2008) invokes both syntax and prosody. He demonstrates that clitic placement cannot be defined by syntax alone, since the clitic pronoun *je* can occur in different syntactic positions:

- (57) *Slovenian*
- a. Janez je mogoče ne mara.
 Janez her possibly NEG like
 ‘Janez possibly doesn’t like her.’
- b. Mogoče je Janez ne mara.
 possibly her Janez NEG like
 ‘Janez possibly doesn’t like her.’
- c. Ne mara je.
 NEG like her
 ‘He doesn’t like her.’
 (Marušič 2008)

²¹¹ Similar analyses have been presented for Macedonian, Czech and Old Czech (cf. Kosta & Zimmerling 2013).

The clitic can occur either before both the adverb and the negative, as in (57a); between the adverb and the negative, as in (57b), or after the negative, as in (57c). In contrast, the order of the adverb and the negative is fixed:

- (58) *Slovenian*
- a. Janez mogoče ne mara zelenjave.
 Janez possibly NEG like vegetables
 ‘Janez possibly doesn’t like vegetables.’
- b. *Janez ne mara mogoče zelenjave.
 Janez NEG like possibly vegetables
 For: ‘Janez possibly doesn’t like vegetables.’ (Marušič 2008)

Thus, Marušič (2008) concludes that a purely syntactic approach will not be able to explain clitic placement. Instead, he argues for a mixed approach, suggesting that clitics do not break syntactic constituents but are realised directly right-adjacent to the first intonation unit – i.e. following a pause. This can explain why the clitics can be proclitic in (56) but never occur utterance initially. Golden & Sheppard (2000) propose a somewhat different account but the important point is that a language can allow both proclitic and enclitic placement and still have restrictions against clitics in initial position.

It seems unlikely that Kelabit clitics follow the same pattern as Slovenian since Slovenian has a Wackernagel clitic system (cf. Kosta & Zimmerling 2013). Kelabit pronominal clitics, in contrast, have the properties of verb-adjacent clitics, filling a fixed position directly following the verb, noun or preposition that they form a syntactic constituent with. Nonetheless, the lesson to be learned from Slavic clitic systems is that adopting an approach to clitic placement that is based on both syntactic and prosodic features may help to explain the patterns that we find. Indeed, as seen in SUBSECTION 4.5.3, the clitic position seems to be determined syntactically directly

following the verb. Hence, the pronouns will always be syntactically enclitic. They may be prosodically enclitic as a last resort when there is no potential host for prosodic proclitics. Alternatively, prosodic attachment may be affected by whether the potential host is a predicate or not, as was shown to affect Slavic clitics in Diesling et al (2009).²¹² In any case, a key question for future research is whether the pronouns are also prosodically enclitic in particular environments and which factors seem to affect prosodic attachment.

4.6.2 Austronesian clitic typology

The results also have interesting implications for the place of Kelabit within the typology of Austronesian pronominal clitics presented in SUBSECTION 4.3.2. The typology can be loosely summarised as follows:

- (59) **Austronesian Clitic Typology**
- | | | |
|----|-----------------|-----------------------|
| a. | Philippine-type | Wackernagel clitics |
| b. | Transitional | Verb-adjacent clitics |
| c. | Indonesian-type | Verbal proclitics |

In SUBSECTION 4.3.2, I suggested that Philippine-type and Indonesian-type languages differ in their typical clitic systems. I also argued that a number of languages in the Southern Philippines and Sulawesi have clitic systems that are best described as transitional. Typically, such languages have verb-adjacent pronoun clitics, rather than second-position clitics. Moreover, in Sulawesi, the NOM and GEN clitics differed in

²¹² If a distinction between predicate hosts and non-predicate hosts were found, this could have a phonological explanation. Most UV predicates, apart from *kinan* ‘UV.R.eat’ – are formed via *-in/-en-* infixation to a bi-syllabic lexical root (SUBSECTION 2.4.1.2.3). Thus they are tri-syllabic, unlike most simple nouns and prepositions which – as lexical roots themselves – have two syllables. To my knowledge there very few four syllable words that contain suffixes in Kelabit (except *terepunen* ‘UV.IRR.keep’). Moreover, prefixes are much more common than suffixes (see SUBSECTION 2.4.1). Therefore, there may be phonological restrictions on enclitics in Kelabit. However, number of syllables cannot be the only factor since the pronouns attached proclitically even where the host was *kinan* in the prosodic study detailed in this chapter.

their behaviour and placement. However, a major distinction between the transitional clitic systems in Sulawesi and the Southern Philippines is that the transitional languages in the Philippines maintained a restriction against clause initial clitic placement. Given that Kelabit appears transitional in terms of its voice-system (CHAPTER 3), we might reasonably ask whether its pronominal clitic system is also transitional, and how it compares to the other languages described.

The results presented in SUBSECTION 4.5 suggest that the pronouns seem to fit best into a transitional clitic system, rather than the typically Philippine-type Wackernagel system or the typically Indonesian-type affix-like system. This is because, like Kulawi, FORM 1 and FORM 2 pronouns differ in their syntactic distribution and prosodic behaviour. FORM 2 is always a clitic and, as such, is realised verb-adjacently, like Indonesian-type clitics. However, unlike in Indonesian-type languages, the clitics follow their syntactic head. FORM 1 clitics seem to have some of the distributional characteristics of Philippine-type languages, in that the pronouns follow negative particles, rather than the predicate, as shown in (8a). However, unlike the typical Wackernagel enclitic systems, the pronouns are prosodically proclitic. Moreover, they are only proclitic in the immediately post-verbal position and can occur as free-standing words in other positions. Hence, it is possible to conclude that the FORM 1 pronouns are on their way to degrammaticalising to words, much like their cognates in Indonesian-type languages. Thus, Kelabit seems to share some characteristics of pronoun placement with the more conservative Philippine-type systems, but is moving towards an Indonesian-type system, both in the sense that the clitics attach proclitically, and in the sense that the FORM 1 pronouns can occur as non-clitics. Hence, the clitic system supports the conclusion from CHAPTER 3 that Kelabit is neither proto-typically Philippine-type nor Indonesian-type.

The study of clitics also has some implications for proposed diachronic changes in Austronesian. In contrast to other transitional systems, such as Kulawi, both Kelabit pronoun classes are syntactically enclitic and prosodically proclitic, regardless of whether undergoer voice realis or irrealis mood marking is employed. Equally, unlike Kulawi, Kelabit maintains the Southern Philippine restriction against clause initial clitics and has the UV realis infix and the irrealis suffix (SUBSECTION 2.4.1.2.3 and 2.4.1.3.1). Hence, the transition from Wackernagel to verb-adjacent clitic cannot be accounted for by the same path of development posited in Billings & Kaufman (2004) and Himmelmann (1996). This suggests that there may have been many independent changes in Austronesian syntax, and that closer analysis of clitic systems in Western Austronesian is warranted.

4.6.3 Summary

In this section, I discussed the implications of the previous section for the syntax/prosody interface and Austronesian typology. I argued that Kelabit clitics provide further evidence for the mismatch between syntax and prosody. Prosodically, the pronouns attach proclitically. However, there appears to be a syntactic clitic position directly following the host. This prompted the question of whether clitics can, in fact, be both enclitic and proclitic depending on context, which remains for future research. As for the place of Kelabit in Western Austronesian typology, I conclude that Kelabit clitics are best analysed as transitional rather than Wackernagel or verb-adjacent proclitics/prefixes. Therefore, the clitic system is another means by which the status of Kelabit as neither prototypically Philippine-type nor Indonesian-type becomes apparent.

4.7 Conclusion

In this chapter, I examined a set of variant pronouns. Cognate forms in related Austronesian languages have been analysed as representing NOM and GEN case. However, syntactic tests demonstrated that a case-based analysis would not accurately represent the pronouns in Kelabit, even though an ergative analysis (or voice-based analysis) could be extended to the pronouns in Lundayeh, one of the most closely related languages to Kelabit (SUBSECTION 2.2.1). Hence, Kelabit could also be described as transitional between Philippine-type and Indonesian-type on the basis of the form of the pronouns, in that they appear to have lost their strict case-marking function, although the FORM 2 is commonly used for actor non-subjects. Moreover, this supports the proposal from the previous chapter that Kelabit represents an intermediate stage in a transition from ergative to accusative, since the form which could be said to mark ergative in Lundayeh has developed into a form of differential marking.

I subsequently explored distributional rather than functional differences between the two pronouns, which led to the hypothesis that FORM 2 pronouns are prosodically weaker than FORM 1. Prosodic tests revealed that the two pronouns do differ in their prosody – but not in the sense that one is a strong pronoun and the other a weak pronominal clitic. Instead, the difference is that FORM 2 pronouns are always clitics, whilst FORM 1 pronouns can be realised as clitics in the post-verbal clitic position but also appear in other positions in the clause, where they stand as independent words. This also represents a transition from Philippine-type languages, where both NOM and GEN pronouns are typically analysed as Wackernagel enclitics. Moreover, it differs from Indonesian-type languages, since the pronouns are syntactically enclitic to their host, and NOM pronouns can sometimes be realised as

clitics. Consequently, clitic phenomena are another area in which more than two typological categories are needed to capture the different parameters of variation.

Two big questions remain. Firstly, what motivates the use of FORM 2 pronouns instead of FORM 1? Secondly, what determines whether FORM 1 pronouns are realised in the clitic position or as free-standing words elsewhere? This most likely ties in with the semantics and information structure of the voices, given that the FORM 2 pronouns are restricted to UV and certain experiential intransitive predicates. For now, I conclude that the system is demonstrably different from typical Philippine-type languages, both in case-marking and clitic status, and move in the next chapter to consider word order differences and information structure in more detail.

Chapter 5

Word Order

5.1 Introduction

In CHAPTER 4, I explored the fact that Western Austronesian languages not only differ in their voice systems, but also in their pronominal systems. Philippine-type languages tend to have systems of case-marking for both nouns and pronouns. The case distinctions are typically analysed as nominative and genitive, but are often understood as ergative/absolutive case systems. Both NOM and GEN pronouns are second-position enclitics. Indonesian-type languages, in contrast, do not have overt case-marking of nominal or pronominal arguments. Pronouns are typically free-standing. However, there are proclitic actor pronouns for 1SG and 2SG in undergoer voice. These appear cognate with GEN pronouns in other Western Austronesian languages. Hence, it seems reasonable to conclude that GEN pronouns are clitics, whilst NOM pronouns are non-clitic. As a result, Western Austronesian pronominal systems appear to make a transition from a system where the key difference between the pronouns is the case-marking, and both pronouns are prosodic clitics, to a system where the key difference between the pronouns is prosodic and cases are no longer morphologically distinguished.

In Kelabit, the Philippine-type case system has broken down. Although FORM 2 (GEN) pronouns are typically used for actors in non-actor voices, FORM 1 (NOM) pronouns exist as an alternative means of expression. However, they do not have the prosodic patterns typical of Indonesian-type languages, since both NOM and GEN pronouns can be clitics in the post-verbal position. Instead, the key difference is that NOM pronouns can also occur in other positions where they are free-standing. Consequently, both case-marking and clitic status constitute additional evidence that Kelabit is transitional between Philippine-type and Indonesian-type.

In this chapter, I explore one final case-study of variation in Western Austronesian, namely word order. As shown in SUBSECTION 1.3.1, Philippine-type and Indonesian-type languages differ in their basic word order. Philippine-type languages are said to be verb-initial, whilst Indonesian-type languages are verb-medial. However, as expected given the discussion in CHAPTERS 3 and 4, word order is not as clear cut as it might seem from this dichotomy. In fact, Austronesian languages also vary in how flexible word order is and the sorts of factors that motivate different word orders, including definiteness, animacy and information structure (SUBSECTION 5.4). Moreover, in a number of Western Austronesian languages word-order choices are affected by the voice construction (SUBSECTION 5.4). Consequently, this chapter explores word-order patterns in Kelabit and what this can tell us about the wider debates surrounding Kelabit voice, and the relationship between Kelabit and other Western Austronesian languages. Ultimately, word-order variation provides further support for the idea that Kelabit is transitional between Philippine-type and Indonesian-type, since UV seems to share characteristics with Philippine-type word order, whilst AV is similar to Indonesian-type equivalents (SUBSECTION 5.5.4).

The chapter is structured as follows. SUBSECTION 5.2 defines word order and introduces relevant aspects of word-order typology. SUBSECTION 5.3 presents a methodology for analysing word order. SUBSECTION 5.4 describes word-order variation in Western Austronesian and SUBSECTION 5.5 presents possible word orders and word-order constraints in Kelabit.

5.2 Word Order

Word Order is a generic term for the linear order in which words are arranged. Many word-order distinctions have been discussed in the literature, including the relative order of noun-possessor, noun-adjective, noun-determiner, noun-relative clause and noun-numeral (cf. Donohue 2007). However, more often than not, typologists and descriptive linguists are interested in the relative order of verb, subject and object phrases within the clause (Dryer 2013b). Hence, word order in this chapter refers to the phrasal order of the verb and its core arguments.

5.2.1 Greenberg and the Six-way Typology

Perhaps the best known typology of clausal word order dates back to Greenberg (1963) who drew on the earlier works of Behagel (1909/10) and Schmidt (1926) to describe a set of implicational universals using a sample of 30 languages. He described languages in terms of the basic order of subject (S), verb (V) and object (O).²¹³ There are six logically possible orders, which are illustrated in (1):

- (1) a. **SOV** (*Korean*, Koreanic)
 Kiho-ka saca-lil cha-ass-ta.
 Keeho-NOM lion-ACC kick-PST-IND
 ‘Keeho kicked the/a lion.’

²¹³ See SUBSECTION 5.3.2 for definition of basic word order.

- b. **SVO** (*Thai, Tai-Kadai*)
 Khon ní kàt mǎa tua nán.
 man this bite dog CLF that
 ‘This man bit that dog.’
- c. **VSO** (*Welsh, Celtic*)
 Lladdodd draig ddyn.
 kill.PST dragon man
 ‘A dragon killed a man.’
- d. **VOS** (*Malagasy, Austronesian*)
 Manasa ny lamba ny vehivavy.
 wash the clothes the woman
 ‘The woman is washing the clothes.’
- e. **OVS** (*Panare, Cariban*)
 Pi? kokampö unkĩ?
 child washes woman
 ‘The woman washes the child.’
- f. **OSV** (*Nadëb, Amazonian*)
 Samũy yi qa-wùh.
 howler-monkey people eat
 ‘People eat howler-monkeys.’ (Song 2011: 255)

Each of the orders in (1) is attested in the world’s languages. However, they do not occur with equal frequency. In fact, Greenberg’s (1963) sample only included SOV, SVO and VSO languages. Consequently, Greenberg (1963) proposed that there is a universal preference for orders in which the subject phrase precedes the object phrase. Subsequent studies report similar findings (e.g. Ruhlen 1975, Blake and Mallinson 1981, Tomlin 1986, Dryer 2013a).²¹⁴ Though the exact percentages differ, SOV and SVO word orders occur most frequently in the samples. Verb-initial orders are less frequent but widely attested and object-initial orders are relatively rare. Finally,

²¹⁴ Sampling methods are used in order to prevent distortion by analysing closely related languages (Dryer 1997)

some languages have no dominant order and are unclassified according to the six-way typology (see SUBSECTION 5.2.4). The results are summarised in TABLE 5.1:²¹⁵

Table 5.1 Distribution of Word Orders

Word Order	Greenberg (1963)	Ruhlen (1975)	Blake and Mallinson (1981)	Tomlin (1986)	Dryer (2013a)
SOV	43.0%	51.5%	41.0%	44.8%	41.0%
SVO	37.0%	35.6%	35.0%	41.8%	35.4%
VSO	20.0%	10.5%	9.0%	9.2%	6.9%
VOS	0.0%	2.1%	2.0%	3.0%	1.8%
OVS	0.0%	0.0%	1.0%	0.0%	0.8%
OSV	0.0%	0.2%	1.0%	0.0%	0.3%
Unclassified	0.0%	0.0%	11.0%	0.0%	13.7%
Sample Size	30	427	100	402	1377

5.2.2. Dryer and the VO/OV Typology

More recent typologies of word order, starting from Lehmann (1973) and Vennemann (1974), have moved away from the traditional six-way classification. Perhaps the best known example is Dryer (1997, 2013b) who proposed a typology in terms of two basic parameters, namely OV/VO and SV/VS.²¹⁶ This allows him to posit four major classes, shown in (2) in relation to the six-way typology (Dryer 2013b):

- (2)
- a. OV: SOV, OSV, OVS
 - b. VO: SVO, VSO, VOS
 - c. SV: SVO, SOV, OSV
 - d. VS: VSO, VOS, OVS

²¹⁵ Tomlin (1986) suggests an explanation for different frequencies in terms of three functional principles: *Theme First Principle* (TFP), *Animated First principle* (AFP) and *Verb-Object Bonding Principle* (VOB). SOV and SVO are argued to adhere to all three principles and hence are most frequent. VSO adheres only to the TFP and AFP and is therefore less frequent. VOS and OVS adhere only to the VOB and are therefore less frequent still and OSV adheres to no principles and is therefore dispreferred. However, this model cannot explain why SVO is statistically less frequent than SOV, as shown in Dryer (1989) etc.

²¹⁶ Dryer (2013b: 270) argues that the order of S and O is not significant. This is based on the fact that clauses with nominal S and O are rare cross-linguistically.

There are a number of advantages of Dryer's model. Firstly, it allows us to classify languages that remain unclassified by the Greenbergian approach. Many languages, such as Hanis Coos, are inconsistent with regards to the six-way typology and allow multiple orders of subject, object and verb:

Table 5.2 Word Order in Hanis Coos (Dryer 1983)

Word Order	Text frequency	Token number
SVO	38%	6
VOS	25%	4
VSO	19%	3
OVS	19%	3

Although the sample size is far too small to draw strong conclusions, the data in TABLE 5.2 would suggest that Hanis Coos has no basic order according to the six-way typology, since the most frequent order occurs only 38% of the time. However, if this is restated in terms of VO/OV and VS/SV then we can classify Hanis Coos as VO and VS, since these orders occur with a much greater frequency than the alternatives: VO (70%) vs OV (30%) and VS (77%) vs SV (23%) (Dryer 1997: 81). Moreover, some languages can be consistently classified with respect to one of the two parameters but not both (see Dryer 2013b). This applies to a number of Western Austronesian languages, which are consistently VO, but allow flexible positioning of the subject (SUBSECTION 5.4). Thus, though there remain languages with no dominant order using Dryer's typology, a greater number of languages can be classified when the two word-order parameters are handled separately.

Secondly, Dryer's classification is based both on clauses with two nominal arguments and clauses with only one nominal argument. The traditional typology relies on clauses with an overt nominal S and an overt nominal O, which is cross-linguistically rare in natural discourse. For example, Payne (1990: 220) analysed a

corpus of 1,526 Yagua clauses, of which only 49 contained two nominal arguments. Similar results are given for a range of genetically-unrelated languages in Du Bois (2003: 35) and were found for Kelabit in SUBSECTION 5.5. In contrast, the percentage of clauses containing at least one nominal argument is typically higher and allows for more accurate classification.

Thirdly, dividing word order into two parameters allows us to compare transitive and intransitive clauses, which often differ in their basic word order (see Dryer 2007). For example, in the Hokan language group of California transitive clauses are overwhelmingly SVO (Dryer 1997: 88). However, intransitive subjects tend to follow the verb and, as in many languages, are more frequent than transitive subjects (Dryer 1997: 90). The same could be said for languages like Spanish and Polish, where VS order is quite common in intransitive clauses (Dryer 1997: 87). Hence, describing the three languages as SVO is potentially misleading, despite this being the most frequent order in transitive clauses.

Finally, Dryer's typology makes a series of typological predictions that appear empirically correct. In particular, it predicts that verb-initial and verb-final languages should form a natural class, whereas verb-medial SVO and OVS languages should not. This would seem an equally natural grouping according to the six-way typology. However, OVS languages tend to pattern with SOV languages rather than SVO.²¹⁷ In fact, empirical evidence discussed in SUBSECTION 5.2.3 suggests that VO languages (VSO, VOS and SVO) form a natural class and OV languages (SOV, OVS and OSV) form a natural class. Hence, the typology does not lose any of the predictive power of

²¹⁷ There are few examples of basic OVS languages, but Hixkaryana, one of the best documented examples, has little in common with SVO. For example, it has postpositions and places manner adverbs after the verb. As shown in SUBSECTION 5.2.3.2, this is more similar to SOV than verb-final languages and hence in keeping with the OV typology (see Dryer 1997).

the traditional typology. Indeed, it distinguishes the more relevant VO/OV parameter from the less relevant SV/VS parameter (SUBSECTION 5.2.3).

Interestingly, when the two-way OV/VO typology is compared to the six-way typology of Greenberg (1963), both orders are roughly equal in their frequency. For example, of the 1,519 language sample in Dryer (2013c), 713 (46.9%) languages had dominant VO order, 705 languages had dominant OV (46.4%) order, and 101 (6.6%) had no dominant order.

5.2.3 Typological Correlations with Word Order

Whichever way word-order typology is presented, it has become a subject of much interest since a number of other typological properties have been shown to correlate with the basic order of S, V and O (see Dryer 1992). Though there are no exceptionless correlations, a number of regular patterns have been identified (Siewierska 1988: 8). In the following subsections, I focus on verb-initial and SVO languages as the vast majority of Western Austronesian languages fall into one or other of these categories (see Polinsky & Potsdam, to appear).

5.2.3.1 Verb-Initial Languages

Verb-initial languages, as discussed in SUBSECTION 5.2.1, constitute somewhere between 12-19% of the world's languages (cf. Clemens & Polinsky, to appear). They include VSO, VOS and alternating VSO/VOS languages and are found in a number of different geographical areas and language families. This includes: Afro-Asiatic and Nilo-Saharan languages in Africa; Celtic languages in Europe; Mayan and Oto-Manguean languages in Central America; Arawakan languages in South America; Salish and Wakashan languages in North America and of course Austronesian

languages in South East Asia and the Pacific (cf. Clemens & Polinsky, to appear). They share several typological properties. For example, they are typically prepositional and have preverbal auxiliaries and initial complementisers (see Dryer 2013b). Furthermore, Greenberg (1963) notes that all VSO languages have alternative possible word orders, most commonly SVO.

The strength of these correlations was tested in Dryer (1997), who conducted a quantitative analysis of the frequency of different typological properties in a sample of verb-initial languages. The results are given in TABLE 5.3:²¹⁸

Table 5.3 Typological Properties of Verb-initial Languages (Dryer 1997)

Typological Property	VSO	VOS
Prepositions	83.3%	100.0%
Noun-Genitive	88.9%	78.6%
Noun-Relative clause	100.0%	100.0%
Article-Noun	82.4%	100.0%
Numeral-Noun	79.2%	88.9%
Verb-PP	100.0%	100.0%
Negative-Verb	95.8%	100.0%
Auxiliary-Verb	72.7%	83.3%
Initial Q	63.2%	100.0%
Initial <i>wh</i>	79.2%	75.0%

Though the percentage of VOS languages with the properties listed were typically higher than the percentage of VSO, the differences are not statistically significant.

There are also additional structural properties beside word order that correlate with verb-initiality. For example, Clemens & Polinsky (to appear) discuss properties that are common in verb-initial languages, including a lack of non-finite verb forms, no verbal expressions with the meaning HAVE, and ergative alignment. Lastly, many researchers claim that verb-initial languages do not distinguish between nominal and

²¹⁸ The languages are from different genera, or genetic groupings with a similar time depth. The percentages refer to the number of genera containing languages with these features (see Dryer 1997: 76).

verbal syntactic categories (see Kaufman 2009) or that the distinction between the two categories is less clear cut than in other languages (Clemens & Polinsky, to appear). Hence, verb-initial languages share a number of characteristics.²¹⁹

5.2.3.2 SVO Languages

As discussed in SUBSECTION 5.2.2, Dryer's (1997) typology predicts that SVO languages should have features in common with verb-initial languages, since they are all VO. Dryer (1991) demonstrates that this is the case, as can be seen in TABLE 5.4:

220

Table 5.4 Comparing Verb-final, SVO and Verb-initial Languages (Dryer 1991)

Property	V-final	SVO	V-initial
Postpositional	96%	14%	9%
Relative clause-Noun	43%	1%	0%
Standard of comparison-adjective	82%	2%	0%
Predicate-copula	85%	26%	3%
Subordinate clause-subordinator	70%	6%	0%
Noun-plural word	100%	24%	1%
Adpositional phrase-verb	90%	1%	0%
Manner adverb-verb	91%	25%	1%
Verb-Tense/aspect aux verb	94%	21%	1%
Verb-negative auxiliary	88%	13%	0%
Genitive noun	89%	59%	28%
Sentence-Question particle	73%	30%	13%
<i>wh</i> not obligatorily initial	71%	42%	16%

For 10 of the 13 typological parameters, SVO languages are much more similar to verb-initial languages than to verb-final languages. Even for the final three parameters,

²¹⁹ Nb. despite typological similarities there are also structural differences between different verb-initial languages, some of which are discussed in SUBSECTION 5.4. This has led to a variety of theoretical accounts, including: right-hand specifiers, flat structure, V-raising, VP-raising and subject-lowering (see Potsdam 2009 and Clemens & Polinsky, to appear, for further discussion and references for different accounts of Austronesian)

²²⁰ Figures are averaged for language genera across six geographical areas to ensure that they are not genetically or geographically biased. Dryer typically uses language genera in his research. These are groups of genetically related languages such as the subgroups of Indo-European.

they are arguably intermediate between verb-final and verb-initial. Thus, Dryer (1991) concludes that the OV/VO parameter is fundamental for word-order typology and that verb-initial and SVO languages form a natural class.²²¹

5.2.4 Basic and Alternative Word Orders

So far, I have discussed languages with one fixed word order. However, many languages have multiple possible orders (cf. Bakker 1998, Siewierska 1998).²²² These languages are said to have ‘flexible’ word order and constitute a number of sub-types (Dryer 2013a). Firstly, non-configurational and discourse-configurational languages often permit all or most possible orders of S, V, and O (see Hale 1983, Dahlstrom 1991, É-Kiss 1995). For example, consider Hungarian.²²³

- (3) *Hungarian* (Uralic)
- a. **SVO**
Janos szereti Marit.
John loves Mary
‘It is John who loves Mary.’
 - b. **SOV**
Janos Marit ‘szereti.
John Mary loves
‘As for John and Mary, he loves her.’
 - c. **OSV**
Marit Janos ‘szereti.
Mary John loves
‘As for Mary and John, he loves her.’

²²¹ The idea that SVO languages do not differ in any meaningful way from VSO and VOS languages has been controversial (see Newmeyer 2004, 2005). For this reason, Dryer (2013b) repeated the study using a larger sample of languages and a greater number of typological parameters. In this study, there was a greater number of parameters for which SVO was intermediate between verb-final and verb-initial. Moreover, there was one parameter (the order of verb and negative particle) for which SVO languages more closely resembled verb-final languages. Nonetheless, in the vast majority of cases, SVO and verb-initial languages did pattern alike. Hence, although SVO and verb-initial languages are clearly different, there is sufficient evidence to posit a VO class.

²²² Such languages may not show the typological correlations discussed in SUBSECTION 5.2.3 (Mithun 1992).

²²³ The apostrophe in (3) indicates the placement of sentence stress.

- d. **OVS**
 ‘Marit szereti Janos.
 Mary loves John
 ‘It is Mary whom John loves.’
- e. **VOS**
 ‘Szereti Marit Janos.
 loves Mary John
 ‘John loves Mary.’
- f. **VSO**
 ‘Szereti Janos Marit.
 loves John Mary
 ‘John loves Mary.’ (É-Kiss 1981: 187)

All possible orders of S, V and O occur. Hence, word order is not constrained by grammatical function. Instead, other factors determine word order choice, as discussed in SUBSECTION 5.3.3.

In other cases, languages allow flexible positioning of one argument, and not of the other (see SUBSECTION 5.2.2). For example, Syrian Arabic has fixed VO order. However, the subject can either appear pre-verbally or post-verbally as both SVO and VSO orders are possible (Dryer 2013a). Finally, word order may differ depending on the clause-type. For example, in German SVO order predominates in main clauses without auxiliaries, but SOV order is found in subordinate clauses and clauses with auxiliaries:

- (4) *German*
- a. **No Auxiliary (SVO)**
 [Der Lehrer] **trink-t** [das Wasser].
 DEF.NOM.M teacher drink-PRS.3SG DEF.ACC.N water
 ‘The teacher is drinking water.’
- b. **Auxiliary (SOV)**
 [Der Lehrer] ha-t [das Wasser] **getrunken.**
 DEF.NOM.M teacher have-PRS.3SG DEF.ACC.N water drink.PST.PTCP
 ‘He didn’t know that I saw the man.’ (Dryer 2013a)

Hence, languages can differ in their word order flexibility and the factors that determine word order choice.

To better understand the nature of word-order flexibility, Siewierska (1998) analysed a sample of 171 languages from typologically diverse language families and different geographical areas. She aimed to understand how word order flexibility differed across languages and whether this correlated with morphological features, such as the presence or absence of case-marking and agreement. Her conclusion was that five categories of word order flexibility could be identified (Siewierska 1998: 504):

- (5) **Word Order Flexibility**
 - a. Rigid word order (only one dominant order)
 - b. Restricted word order (one dominant and one variant order)
 - c. Variable word order (one dominant and two variant orders)
 - d. Flexible word order (three or more variant orders)
 - e. Highly flexible word order (four or more variant orders)

It has been proposed in the literature that the degree of word order flexibility is linked to the morphological encoding of arguments. Those languages with case-marking or complex verbal agreement are said to allow highly flexible word order since grammatical functions can be determined on the basis of morphology. In contrast, those languages without case-marking may reflect grammatical functions configurationally through a fixed structural position. However, Siewierska (1998: 507-509) found that there is no direct link between morphological encoding and word order flexibility. Whilst rigid word order does correlate with the absence of verbal agreement or case-marking of nominal arguments, morphological encoding is not sufficient to predict highly flexible word order, since there are many languages with case that nonetheless have low degrees of word order flexibility.

As mentioned in SUBSECTION 5.2.3.1, Greenberg (1963) noted that the nature of variant word orders is often affected by the dominant word order. Siewierska (1998: 492) also found evidence to support this idea. She found that preferred and dispreferred variant orders as well as degree of flexibility differed depending on the basic word order (Siewierska 1998: 492).²²⁴ For verb-initial languages, the preferred variant order is overwhelmingly SVO. Indeed, 63% of VSO languages in Siewierska's sample had an SVO variant order, and 80% of VOS languages (Siewierska 1998: 494). Similarly, VSO languages are unlikely to have no variant orders in comparison with SOV and SVO. Hence, studies in word order do not necessarily restrict themselves to basic word orders but can also examine variant orders and the descriptive generalisations about word-order flexibility that can be made.

5.2.5 Summary

In this section, I discussed the fact that languages can have different word orders and different degrees of word order flexibility. I suggested that variation in word order can be approached from a typological perspective, be it the Greenberg (1963) six-way typology or the Dryer (1997, 2013b) OV/VO typology. Finally, I illustrated that the relative order of S, V and O can be correlated with a range of other typological properties, such as the use of prepositions or postpositions and the order of verbs and auxiliaries. The next section addresses the question of how to describe word order in a given language and how this can be applied to the comparison of Kelabit and other Western Austronesian languages.

²²⁴ See Siewierska (1998: 491-504) for some suggested pragmatic explanations.

5.3 Methodology for Studying Word Order

In the previous section, I narrowed the discussion of word order to the relative order of subject, object and verb within the clause. In this section, I introduce several important parameters for comparing word order in Kelabit with word order in other Western Austronesian languages. In particular, I follow Dryer (2007) in suggesting that word-order patterns can be compared across languages by answering the following questions:

- (6) **Analysing Word Order**
 - a. Which orders occur?
 - b. Which orders are more common/basic?
 - c. Which factors determine the choice between different orders?

In other words, it is necessary to describe the possible word orders, the basic word order and the contexts in which different word orders are likely to occur. Each of these requires its own methodological considerations.

5.3.1 Describing Possible Word Orders

The first step in comparing Kelabit to other Austronesian languages is describing possible word orders, not just in transitive clauses but also in intransitive clauses. As Dryer (1997) demonstrates, intransitive clauses and transitive clauses often show different word-order possibilities (SUBSECTION 5.2.2). Hence, word order variation in both clause types should be addressed. In order to describe possible word orders, a combination of elicitation and naturalistic discourse is required. This is because word orders can be difficult to elicit out of context in a language documentation corpus (cf. Grenoble 2007). However, certain word orders may be possible but quite infrequent in discourse. Moreover, it may be hard to collect the data needed to establish which

factors affect word order choice. Therefore, an approach that draws on both elicitation and data from a corpus collection is adopted.

5.3.2 Establishing Basic Word Order

Establishing basic word order has long been a central aim of studies seeking broad typological comparison (SUBSECTION 5.2.3). A variety of ways of identifying basic word order have been proposed (cf. Song 2011, 2012, Siewierska 1988, Dryer 2007). These can be summarised as follows:

- (7) **Establishing Basic Order**
 - a. Frequency
 - b. Pragmatic Neutrality
 - c. Markedness

Textual frequency is probably the most widely used measure of basic word order (cf. Dryer 1997). Though there are several issues with establishing frequency, as discussed in Siewierska (1988: 8-14), Mithun (1992) and Dryer (1997), if two word orders are possible, and both equally neutral, then the more frequent order will be considered basic. For example, in English the order OSV in (8a) is permitted in the context of contrastive focus. However, in addition to being pragmatically marked, it is less frequent in natural discourse than SVO. Consequently, SVO is considered basic (cf. Dryer 2007).

- (8) **Frequency**
 - a. Beans, I like.
 - b. I like beans.

adjectives. The order N-Adj is reserved for the more complex AdjP modifiers (Dryer 2007: 75):

- (10) **Markedness**
- a. the tall woman
 - b. *the woman tall
 - c. the woman taller than John
 - d. *the taller than John woman
- (Dryer 2007: 75)

The simplest adjective, both morphologically and syntactically, is ‘tall’. Thus, the order Adj-N is considered basic, since it is not restricted to the context of relative clauses.

In some cases, these three factors may give conflicting results. For example, Whaley (1997) suggests that in Yagua, frequency and pragmatic neutrality measures would select a morphologically marked order as basic. Similarly, Mithun (1992) argues that there are languages for which the three factors do not select any order as basic. Nonetheless, where a basic order can be established, frequency, pragmatic neutrality and markedness provide the most effective tests (Song 2012).

5.3.3 Analysing the Contexts in which Different Word Orders Occur

A great wealth of literature has emerged concerning the possible uses of different word orders, particularly on the role of context (Nichols 1979, Mithun 1992, Austin 2001). In this section, I briefly illustrate some of the factors that have been shown to affect word order choice, including animacy/definiteness, semantic role and information structure.

Animacy refers to different levels of ‘aliveness’ and sentience among nominal referents. Definiteness relates to the uniqueness and familiarly/givenness of nominal

referents in context (Lyons 1999). Both are seen as scalar notions and have sometimes been thought of as hierarchies (cf. Aissen 2003: 437):²²⁵

- (11) a. **Animacy Hierarchy**
 Human > Animate > Inanimate
- b. **Definiteness Hierarchy**
 Pronoun > Proper Noun > Definite NP > Indefinite Specific NP >
 Indefinite Non-Specific NP

These can affect word order in various ways. For example, in many languages the preferred word order differs depending on whether the object is definite or indefinite. For example, consider the Mayan language K'iche':

- (12) *K'iche'*
- a. **Definite Object (VSO)**
 X-Ø-u-q'aluj le achi **le** **ala.**
 COM-3SG.ABS-3SG.ERG-hug DEF man **DEF** **youth**
 'The man hugged the youth.'
 NOT: 'The youth hugged the man.'
- b. **Indefinite Object (VOS)**
 X-Ø-u-q'aluj **jun** **achi** le ala.
 COM-3SG.ABS-3SG.ERG-hug **one** **man** DEF youth
 'The youth hugged a man.'
 NOT: 'A man hugged the youth.' (England 1991:466-467)

Like a number of Mayan languages, K'iche' is an alternating VSO/VOS language. However, the alternation is not free. Rather, as illustrated in (12), VSO order is found when the object is animate, definite or specific and VOS when the object is indefinite and non-specific (cf. England 1991). Hence, the choice of word order variant may depend on the animacy/definiteness of the verb's arguments.

²²⁵ See also Dik (1978) on the Language Independent Preferred Order of Constituents (LIPOC).

Similarly, word order choices may differ depending on whether arguments are nominal or pronominal. For example, in the Australian language Ngawun pronominal objects occur pre-verbally, whilst nominal objects follow the verb:

- (13) *Ngawun* (Pama-Nyungan)
- a. **Pronominal Object (OV)**
 T^yunu ŋankan^yta ŋana paɬaŋu.
 that.ERG dingo.ERG me bite.FUT
 ‘That dingo might bite me.’
- b. **Nominal Object (VO)**
 Waɬaŋka puwanu ŋapuyu ŋat^yiniŋu.
 that.ERG hit.PST brother me.GEN
 ‘That fellow hit my brother.’ (Breen 1981: 67)

In (13), the animacy of the actor argument also differs: (13a) involves a situation in which the undergoer is more animate than the actor, whilst in (13b) both actor and undergoer are human.²²⁶ Similar word order differences can be seen depending on whether the subject is nominal or pronominal in Fijian, Margany and Maori (Breen 1981, Keenan 1978, Chung 1998). Hence, the animacy/definiteness of the argument can have an effect on the word order used.

Another important factor that can affect word order is the semantic role of nominal arguments. In this thesis, I have been using the generalised notions of ‘actor’ and ‘undergoer’. The actor refers to the most agentive argument and has properties such as volition, control and initiation of the action. The undergoer refers to the most patientive argument and has properties such as undergoing a change-of-state and being affected by the action (see Dowty 1991). Some theories also make reference to a hierarchy of semantic roles (Bresnan et al 2016):

²²⁶ It is possible that the pronouns may also be clitics.

- (14) **Semantic Role Hierarchy**
 Agent > Benefactive/Goal > Recipient/Experiencer > Instrument >
 Theme/patient > Locative

In some languages, including many in the Philippines, word order is determined in part by semantic role information (see SUBSECTION 5.4.1). For example, in Pangasinan in the Northern Philippines, the actor always follows the verb, regardless of whether it is the subject or not:

- (15) *Pangasinan* (Austronesian)
- a. **Actor = subject**
 Nan-luto si Juan na baaw.
 AV-cook SUBJ Juan CORE rice
 ‘Juan cooked the rice.’
- b. **Actor = non-subject**
 Il-luto-an nen Juan si Pedro na sira.
 BV-cook-BV CORE Juan SUBJ Pedro CORE fish
 ‘Juan will cook fish for Pedro.’ (Siewierska 1988: 51-52)

Similarly, in Lakhota, the actor is initial, followed by animate recipients and patients unless the semantic role of the argument is unambiguously determinable from context (cf. Van Valin 1977).

Finally, information structure has been shown to influence word order. Information structure can be understood as a formal mechanism for facilitating effective information exchange or update (Dalrymple & Nikolaeva 2011, Erteschik-Shir 2007). Among the most important information structure roles are: topic, focus and comment (see Erteschik-Shir 2007).²²⁷ The topic is generally understood as an entity that the speaker identifies and about which the proposition is made (Krifka 2007: 30). Thus, the two key defining factors are ‘aboutness’ and ‘accessibility’.

²²⁷ Other information structure roles have also been suggested, including background and completive. See Dalrymple & Nikolaeva (2011) and Mycock & Lowe (2014) for discussion.

c. **Sentence Focus**

Q: What is happening?

A: Bill is eating chocolate.

In summary, both topic and focus represent information structurally prominent roles, whilst the comment is the proposition made about the topic. This can be equivalent to the focus, or contain the focus within it (see Erteschik-Shir 2007).

Information structure has been shown to influence word order in a number of studies (e.g. Harbour, Watkins & Adger 2012 on Kiowa). Many languages have a default order of topic and comment, independently of which grammatical function is topic (cf. Gundel 1988). For example, Latin has flexible word order in terms of grammatical functions but favours a topic-comment order. Thus, the object appears clause-initially in (18), since it is topical, even though it is not promoted to subject via passivisation, as in the English translation:

- (18) *Latin*
Topical Object
Hunc secutus Marcius Rufus quaestor
 he.ACC followed Marcius.NOM Rufus.NOM quaestor.NOM

 navibus xii.
 ships.ABL 12
 ‘He was followed by the quaestor Marcius Rufus with twelve ships.’
 (Siewierska 1988: 65)

Thus, ‘topic’ versus ‘comment’ status can affect word order independently of grammatical functions.

Topic-comment structure can also explain flexible word order in the Hungarian examples in SUBSECTION 5.2.4. Hungarian has flexible word order (see É-Kiss 2002: 2). However, word order is not freely variable but rather reflects different information structures:

- (19) *Hungarian*
- a. **Topic = Subject**
 [János]_{TOP} [fel hívta Marit]_{COMMENT.}
 John up called Mary.ACC
 ‘John called up Mary.’
- b. **Topic = Object**
 [Marit]_{TOP} [fel hívta János]_{COMMENT.}
 Mary.ACC up called John
 ‘Mary was called up by John.’ (É-Kiss 2002: 2)

The grammatical functions of the arguments in (19a) and (19b) do not change. Therefore, (19a) and (19b) could be thought to have SVO and OVS orders respectively. However, both (19a) and (19b) follow a topic-comment structure. Hence, Hungarian is sometimes referred to as ‘discourse configurational’ in the sense that word order is motivated by information structure rather than grammatical functions (É-Kiss 1995).

In addition to topic and comment, focus can also play a role in word order. In many languages, there is a tendency to put new/newsworthy information first (cf. Mithun 1992, B. Blake 1987). In Cayuga, for example, definite, given NPs are often ungrammatical initially, but can occur clause-finally:

- (20) *Cayuga* (Iroquoian)
- a. **Definite Subject**
 *Ne’ John shakoŋqhwe’s Mary.
 DEF John he.likes.her Mary
 For: ‘John likes Mary.’
- b. Mary shakoŋqhwe’s ne’ John.
 Mary he.likes.her DEF John
 ‘John likes Mary.’
- c. **Definite Object**
 *Ne’ Mary shakoŋqhwe’s John.
 DEF Mary he.likes.her John
 For: ‘John likes Mary.’

- d. John shakoŋohwé's ne' Mary.²²⁹
 John he.likes.her DEF Mary
 'John likes Mary.' (Mithun 1992: 27)

Similarly, new information tends to precede old information. This is illustrated in (21), which was produced in answer to the question of whether the speaker likes baked potatoes:

- (21) *Cayuga*
- a. Ne:' kyé:'q thrés i:nq kyé:'q é:ke:k.
 it.is I.guess too far I.guess I.will.eat
 'It's just that I eat them so seldom.'
- b. **Skins = new**
 Ne:' tshq: ne' oa'wistá' the' ní:' t'e:ke:s.
 it.is only DEF peeling not I do.I.eat
 'I just don't eat the skins.'
- c. **Pigs = new**
 Kwiskwís kyé:' hne:' tshq: ka:tí:s ne' oa'wistá'.
 Pig just CONTR just they.eat DEF peeling
 'Only the pigs eat the skin.' (Mithun 1992: 29)

In (21b), the skins are definite, as they are encoded with the definite particle *ne'* and are retrievable from the context of talking about potatoes. Hence, definiteness does not motivate word order variation by itself. Instead, the different word orders in (21b) and (21c) reflect the relative status of the skins as 'new/focus' and 'given/non-focus'. In (21b), the speaker makes a contrast between eating potatoes and eating skins. Hence, the skins represent important focus information. In (21c), in contrast, the skins are no longer new and are realised clause-finally. The focus is the contrast between pigs, who eat the skins, and the speaker, who doesn't. Hence, Mithun (1992: 31-34) suggests that word order in Cayuga may be determined by a principle of

²²⁹ Names may also have a degree of definiteness but certainly cannot co-occur with the definiteness marker in initial position.

‘newsworthiness’. She suggests that information is newsworthy when it conveys new information, introduces a new topic or establishes a new focus/contrast (Mithun 1992).

Newsworthiness plays a role even in languages like Russian, Polish and English, that show a preference for topic-comment order in written text. In these languages, new/prominent information is often ordered before old information in colloquial speech (Siewierska 1988: 74). Hence, word order can be constrained by the relative order of topic and comment, but also according to the position of focus and newsworthy information. In summary, animacy, definiteness, semantic role and information structure can all play a role in word-order variation.²³⁰

5.3.4 Summary

In this section, I argued that an analysis of word order involves identifying both the basic word order and any word-orders variants. I argued that basic word order can be identified using discourse frequency, pragmatic neutrality and formal markedness. I then surveyed a set of factors known to constrain word order, including animacy, definiteness, semantics and information structure. The next section explores word order and word-order variation in a range of Western Austronesian languages in order to compare with Kelabit in SUBSECTION 5.5.

5.4 Word Order in Western Austronesian

As discussed in SUBSECTION 5.1, the two major classes of Western Austronesian languages are said to differ in their basic word order. Philippine-type languages are

²³⁰ Other factors which have been proposed to affect word order include iconicity and human cognition & attention (see Siewierska 1988 and D’Elia 2015 for further discussion). These are not considered further here though future research might explore the role that these factors play in Austronesian and Kelabit.

verb-initial, whilst Indonesian-type languages are verb-medial (Donohue 2007a).²³¹ However, much like with voice systems and clitic types, there is also internal variation within each of the groups, as well as a group of transitional languages that share some word-order characteristics of Philippine-type languages and some characteristics of Indonesian-type languages. This section presents examples of variation in Western Austronesian in terms of word order flexibility, basic word order and the factors that appear to determine the use of different word order variants.

It should be noted that word-order discussions in Austronesian are complicated by the controversy surrounding grammatical functions (see SUBSECTION 1.4.1). In this thesis, I assume that the subject is the argument signalled in the verbal morphology in all Western Austronesian languages (cf. SUBSECTION 2.5.1.1). Hence, in AV the order SVO is equivalent to Actor-Verb-Undergoer, whilst in UV the order SVO is equivalent to Undergoer-Verb-Actor. Other studies assume that S = Actor and O = Undergoer, regardless of voice construction (e.g. Aldridge 2010, Dryer 2013a). As a result, some of the analyses presented in this section have been adapted to fit with the understanding of grammatical functions presented in this thesis.

5.4.1 Philippine-type languages

The vast majority of Philippine-type languages are verb-initial (cf. Donohue 2007). However, these are split between rigid VOS and alternating VSO/VOS languages. The Atayalic (Formosan) languages are said to have fixed VOS order (Holmer 2005). This can be illustrated for Tkdaya Seediq in (22):

²³¹ See Donohue (2007a) for discussion of further word-order differences between Philippine-type and Indonesian-type languages.

- (22) *Tkdaya Seediq* (Formosan)
- a. **Subject = Clause-final**
 Wada biq-un hulama na Ape [ka laqi].
 PST give-UV treat GEN Ape NOM child
 ‘Ape gave the child a treat.’
- b. **Subject = Before Actor**
 *Wada biq-un hulama [ka laki] na Ape.
 PST give-UV treat NOM child GEN Ape
 For: ‘Ape gave the child a treat.’
- c. **Subject = Before Undergoer**
 *Wada biq-un na Ape [ka laki] hulama.
 PST give-UV GEN Ape NOM child treat
 For: ‘Ape gave the child a treat.’ (Aldridge 2010: 171)

In *Tkdaya Seediq*, the subject, *ka laki* ‘the child’, always appears finally. When any other argument follows the subject, the clause is ungrammatical, as shown in (22b) and (22c).

Moreover, the clause-final subject is always definite (Aldridge 2010) and cannot have a focus interpretation:

- (23) *Tkdaya Seediq*
- a. **Clause-final focus**
 *S<m>ebut laqi ka ima?
 <AV>hit child NOM who
 For: ‘Who hits a child?’
- a. **Clause-initial focus**
 Ima s<m>ebut laqi?
 who <AV>hit child
 ‘Who hits a child?’ (Chang 1997: 146)

In order for the subject to be interpreted as focus it must occur pre-verbally in a pragmatically marked construction (see (34b) below). Clause-finally, *wh*-words,

which are inherently focus, are ungrammatical, as shown in (23a). For this reason, Aldridge (2010) argues that the clause-final subject has a topic interpretation.²³²

The verb and the non-subject argument form a single constituent, as they can be co-ordinated:

- (24) *Tkdaya Seediq*
Co-ordination Test for Constituency
 [S-bari=na hulama] ma.
 BV-buy=3SG.GEN treat and
- [s-smalu=na lukus dungan] ka laqi=na
 BV-make.3SG.GEN clothes also NOM child=3SG.GEN
 ‘She buys a treat for and also makes clothes for her child.’
 (Aldridge 2010: 171)

Hence, *Tkdaya Seediq* can be argued to have a rigid VOS order, in which the verb and non-subject argument form a constituent and the subject obligatorily appears clause-finally, where it is interpreted as a topic.

Malagasy is also VOS. Indeed, Malagasy could be described as predicate-initial in that nominal and adjectival predicates also occur initially:

- (25) *Malagasy*
- a. **Verbal Predicate**
 [Mividy ny akoho] i Bao.
 AV.buy the chicken HON Bao
 ‘Bao is buying the chicken.’
- b. **Nominal Predicate**
 [Vorona ratsy feo] ny goaika.
 bird bad voice the crow
 ‘The crow is a bird with an ugly voice.’
- c. **Adjectival Predicate**
 [Faly amin’ ny zanany] Rasoa.
 proud PREP the child.3SG Rasoa
 ‘Rasoa is proud of her children.’

²³² See Aldridge (2010) for additional arguments, such as patterns of clitic doubling.

d. **PP Predicate**

[Any an-tsena] Rakoto.
 PREP ACC-market Rakoto
 ‘Rakoto is at the market.’ (Polinsky & Potsdam, to appear)

Like Seediq, the final subject is always definite and receives a topic interpretation (Pearson 2001: 88).

In both Seediq and Malagasy, VOS order is preferred independently of the voice construction:

(26) *Tkdaya Seediq*

a. **Actor Voice**

Wada m-ari huluma ka Ape.
 PST AV-buy treat NOM Ape
 ‘Ape bought a treat.’

b. **Undergoer Voice**

Wada burig-un na Ape ka patis.
 PST buy-UV GEN Ape NOM book
 ‘Ape bought the book.’ (Aldridge 2006: 4)

(27) *Malagasy*

a. **Actor Voice**

N-an-shuratra ilai taratashi ilai umbiashi.
 PST-AV-write DEF letter DEF soothsayer
 ‘The soothsayer wrote the letter.’

b. **Undergoer Voice**

N-u-shuratra-ana ilai umbiashi ilai taratashi.
 PST-UV-write-UV DEF soothsayer DEF letter
 ‘The soothsayer wrote the letter.’ (Rasolofo 2007: 213)

Hence, Seediq and Malagasy have fixed VOS order in both AV and UV constructions, and could perhaps be analysed as comment-topic languages.²³³

²³³ See Rasolofo (2007) for judgements that suggest any other orders in Malagasy would be ungrammatical.

In contrast, there are other Philippine-type languages that have been described as VSO or alternating VSO/VOS. For example, Chamorro and Tagalog alternate between VSO and VOS orders:²³⁴

- (28) *Chamorro*
- a. **VOS**
 Ha-pula' i patgon-ña i nāna.
 AGR-undress the child-AGR the mother
 'The mother undressed the child.'
- b. **VSO**
 Ha-pula' i nāna i patgon-ña.
 AGR-undress the mother the child-AGR
 'The mother undressed the child.' (Chung 2006: 710)
- (29) *Tagalog*
- a. **VOS**
 S<um>ulat ng=liham si=Juan.
 <PFV.AV>write GEN=letter NOM=Juan
 'Juan wrote a letter.'
- b. **VSO**
 S<um>ulat si=Juan ng=liham.
 <PFV.AV>write NOM=Juan GEN=letter
 'Juan wrote a letter.' (Schachter & Otones 1972: 436)

Hence, Philippine-type languages vary in their degree of word-order flexibility.

Moreover, alternating VSO/VOS languages like Tagalog differ from languages like Seediq in other ways, including the ability for adjuncts to be questioned in initial position:

²³⁴ Alternative analyses of Tagalog also exist. For example, Billings (2005) suggests that Tagalog is like Seediq in having basic VOS order. He treats VSO as relating to second-position clitic phenomena. In his account, the apparent word order flexibility is a product of the optionality of realising certain pronominal and proper noun actors as clitics, depending on their information structure. Other Formosan languages have different word-order patterns. For example, Rukai has alternating VSO/VOS order. Saisiyat has SVO order in AV clauses and word order in Bunun and Amis is determined more by semantic roles than grammatical functions (Elizabeth Zeitoun, p.c.).

- (30) *Tkdaya Seediq*
- a. **Questioning adjuncts**
 M-n-ari inu patis Ape?
 AV-PFV-buy where book Ape
 ‘Where did Ape buy books?’
- b. *Inu m-n-ari patis Ape?
 where AV-PFV-buy book Ape
 For: ‘Where did Ape buy books?’ (Aldridge 2002: 395)
- Tagalog*
- c. Saan b<in>ili ni=Maria ang=libro?
 where <PFV.UV>buy GEN=Maria NOM=book
 ‘Where did Maria buy the book?’ (Aldridge 2006: 1)

In *Tkdaya Seediq*, only subjects can be questioned in initial-position, as in (23a). Adjuncts can only be questioned *in-situ*, as in (30a). In *Tagalog*, adjuncts can also be questioned in initial position. This has led to a variety of different theoretical accounts (see Aldridge 2006 for discussion and Huang et al (1999) for further details on interrogative constructions in Formosan languages).

In flexible Philippine-type languages, word order is not random but rather follows some general tendencies relating to grammatical functions, semantic roles and prosodic factors. For example, word order in *Tagalog* is constrained by the following three principles (cf. Billings 2005, Kroeger 1993):²³⁵

- (31) a. **Early-Actor tendency**
 The highest semantic role, or Actor, tends to appear immediately after the verb
- b. **Late-Subject tendency**
 The subject NP tends to appear last in the clause
- c. **Heavy-shift tendency**
 Heavy NPs tend to appear later

²³⁵ For other languages, such as *Batad Ifugao*, *Pangasinan* and *Cebuano*, the actor always immediately follows the verb, regardless of whether it is the subject or not (cf. Dryer 2007, Siewierska 1988: 51).

These result in different word-order patterns, depending on the voice construction. In non-actor voices, the order VOXS is preferred, since it follows both the ‘Late Subject’ and the ‘Early Actor’ principles. In contrast, VSO violates both principles and is only used in pragmatically marked contexts to contrastively focus the actor:

- (32) *Tagalog*
- a. **VOS**
 S<in>ulat ni=Juan ang=liham.
 <PFV.UV>write GEN=Juan NOM=letter
 ‘Juan wrote the letter.’
 Early Actor = ✓
 Late Subject = ✓
- b. **VSO**
 ?S<in>ulat ang=liham ni=Juan.
 <PFV.UV>write NOM=letter GEN=Juan
 For: ‘Juan wrote the letter.’
 Early Actor = X
 Late Subject = X
- (Kroeger 1993: 111)

Hence, VOS could be considered the basic order in UV contexts, much like in Seediq.

In actor voice, however, the principles contradict each other since the actor is the subject and cannot appear both early and late. This results in two equally preferred orders:

- (33) *Tagalog*
- a. **VOS**
 S<um>ulat ng=liham si=Juan.
 <PFV.AV>write GEN=letter NOM=Juan
 ‘Juan wrote a letter.’
 Early Actor = X
 Late Subject = ✓

b. **VSO**

S<um>ulat	si=Juan	ng=liham.
<PFV.AV>write	NOM=Juan	GEN=letter
‘Juan wrote a letter.’		
Early Actor = ✓		
Late Subject = X		

(Billings 2005: 307, from Schachter & Otanes 1972: 436)

VOS follows the ‘Late Subject’ principle, but violates the ‘Early Actor’ principle. On the other hand, VSO follows the ‘Early Actor’ principle, but violates the ‘Late Subject’ principle. Thus, it is not possible to determine which order is basic in terms of markedness and AV clauses can be said to have greater word-order flexibility (see Billings 2005). Consequently, in Tagalog, as in many other Philippine-type languages, word order is affected by the voice construction.

Finally, even fixed word order languages typically allow SVO as a variant order in pragmatically marked contexts. For example, in Seediq an S_{ka} VO construction can be used when the subject corresponds to focus or newsworthy information:²³⁶

(34) *Seediq Information Structure*

a. **What did Pawan do?**

[Mnimah	sino] _{comment}	ka	[Pawan] _{topic} .
drink	wine	NOM	Pawan
‘Pawan drank wine.’			

b. **Who drank wine?**

[Pawan] _{focus}	ka	[mnimah	sino].
Pawan	NOM	drink	wine
‘Pawan drank wine.’			(Karlsson & Holmer 2011)

Hence, SVO is typically used as an alternative order when the subject has a specific information structure role.

²³⁶ Some might analyse this as a type of cleft (see Karlsson & Holmer 2011). Karlsson et al (2015) present similar findings in Puyuma and Bunun. Seediq *ka* appears similar to the Kelabit particle *teh* (SUBSECTION 2.4.2.14.1).

5.4.2 Indonesian-type languages

In contrast, Indonesian-type languages typically have basic SVO word order (Donohue 2007a).²³⁷ This is an innovation and historical records reveal that older stages of Indonesian-type languages were also verb-initial. For example, Modern Javanese is SVO, but Old Javanese is said to be verb-initial and generally more Philippine-type in its morphosyntax (Blust 2013:467). For many Indonesian-type languages, SVO order is fixed in both AV and UV clauses. This is illustrated for Madurese:

- (35) *Madurese*
- a. **Actor Voice**
Sengko' mokol Alwi.
1SG AV.hit Alwi
'I hit Alwi.'
- b. **Undergoer Voice**
Alwi e-pokol sengko'.
Alwi UV-hit 1SG
'Alwi was hit by me.'
- (Kikusawa 2008: 69)

In AV, the actor is subject and appears pre-verbally. The non-subject undergoer follows the verb. In UV, the undergoer is subject and appears pre-verbally, whilst the non-subject actor follows the verb. Hence, both AV and UV have SVO word order.

In some languages, a variant word order occurs with clitic pronouns in undergoer voice (see SUBSECTION 4.3.2). This can be illustrated for Standard Indonesian, which has SVO order in both AV and UV clauses with NP arguments, but SOV in UV constructions involving 1SG and 2SG clitic actors.²³⁸

²³⁷ There are also verb-initial languages in Indonesia, such as Toba Batak (Cole & Herman 2008).

²³⁸ The construction is sometimes treated as object-fronting.

- (36) *Indonesian*
- a. **Actor Voice**
 Hasan menulis surat.
 Hasan AV.write letter
 ‘Hasan wrote a letter.’
- b. **Undergoer Voice, 3rd person actor**
 Surat di-tulis Hasan.
 letter UV-write Hasan
 ‘The letter was written by Hasan.’
- c. **Undergoer Voice, 1st/2nd person actor**
 Surat saya tulis.
 Letter 1SG UV.write
 ‘The letter was written by me.’

(adapted from Musgrave 2002: 38)

Thus, much like Ngawun in SUBSECTION 5.3.3, word order can be affected by the animacy/definiteness of arguments.²³⁹ This results in different word-order possibilities according to voice construction.

Nonetheless, SVO can be analysed as the basic order on account of textual frequency. For example, Gregor (2013) analysed word order in Jakarta Indonesian and Kupang Malay using the spoken Jakarta-Indonesian and Kupang Malay corpora collected by the MPI-EVA Jakarta Field Station.²⁴⁰ She found that SVO order was used in 88.33% of the clauses in the Jakarta Indonesian corpus, and 91.45% of the Kupang Malay corpus. Hence, both Jakarta Indonesian and Kupang Malay can be said to have basic SVO order (see also Sneddon 2006, Paauw 2008: 178).

Verb-initial orders do occur in naturalistic data, but typically with a marked pragmatic function. For example, Sneddon (1996: 257) suggests that the verb can

²³⁹ Nb. clitic pronouns can also appear pre-verbally in Philippine-type languages when the initial position is filled with something other than the verb (see Billings 2005)

²⁴⁰ The Jakarta Indonesian corpus was collected between 2004-2009 and has 257,662 words and 64,093 utterances. The Kupang Malay corpus was collected between 2008-2011 and has 264,156 words and 53,179 utterances. It should be noted that the number of transitive clauses with two core arguments, including both full NPs and pronouns, was small: Jakarta Indonesian = 96, Kupang Malay = 315.

occur in initial position in Standard Indonesian if foregrounded. Similarly, examples of verb-initial order in the Jakarta Indonesian corpus could be said to have a predicate focus reading:²⁴¹

- (37) *Jakarta Indonesian*
Verb-initial Order
 Beli aja kita di material jadi kalo pasir-nya.
 buy just 1SG LOC material become TOP sand-ASSOC
 ‘I just buy the sand at the material store.’ (Gregor 2013: 16)

Indeed, in (37) the undergoer is separated from the rest of the clause by the topic particle, *kalo*. Hence, this is perhaps interpreted as exhaustive focus. Consequently, in Indonesian-type languages verb-initial orders can be considered pragmatically marked, and SVO basic, whilst in Philippine-type languages verb-initial orders are basic and SVO forms a marked construction.

Finally, Indonesian varieties have different degrees of word-order flexibility. Whilst Standard Indonesian is fairly fixed, Riau Indonesian allows all possible word-order variants:

- (38) *Riau Indonesian*
 a. **SVO**
 Saya pakai kaca mata, Vid.
 1SG use glass eye FAM.David
 ‘I’m wearing my glasses, David.’ (putting them on)
 b. **VSO**
 Beli aku laser, ‘kan.
 buy 1SG laser Q
 ‘I’ll buy a laser, right.’ (contemplating a shopping trip)

²⁴¹ No context is given for (37) in Gregor (2013) so it is difficult to evaluate the precise pragmatics.

- c. **VOS**
 Beli nasi goreng aku.
 buy rice fried 1SG
 'I bought the fried rice.'
 (in answer to the suggestion of the speaker buying cards)
- d. **OVS**
 Honda pakai abang Elly.
 motorcycle use elder.brother Elly
 'Elly's using the motorcycle.' (as an explanation for why the speaker cannot go out and buy food on the motorcycle)
- e. **SOV**
 Si Pai aku usir.
 PERS Pai 1SG send.away
 'Pai sent me away.' (complaining about Pai)
- f. **OSV**
 Abang dia sendiri dia usir.
 elder.brother 3 alone 3 send.away
 'His own brother, he sent away.' (complaining about Pai)
 (Gil 2005: 247-248)

Like Hungarian in SUBSECTION 5.2.4, Gil (2005) analyses Riau Indonesian as having highly flexible word order. He argues that the choice is determined by information structure, in that the linear order reflects the newsworthiness of participants.

Moreover, in languages like Balinese, word-order frequency can differ according to the voice construction. In UV, the verb and the non-subject argument always form a constituent:

- (39) *Balinese*
- a. **Undergoer Voice**
 Siap-e [uber cicing] ke jalan-e.
 Chicken-DEF UV.chase dog to street-DEF
 'The/a dog chased the chicken to the street.'
- b. [Uber cicing] ke jalan-e siap-e.
 UV.chase dog to street-DEF chicken-DEF
 'The/a dog chased the chicken to the street.'

- c. *Uber siap-e cicing ke jalan-e.
 UV.chase chicken-DEF dog to street-DEF
 For: ‘the/a dog chased the chicken to the street.’ (Artawa 1998: 19)

In AV, however, although the non-subject undergoer typically forms a constituent with the verb, it is also possible for the subject to intervene between them:

- (40) *Balinese*
- a. **Actor Voice**
 Iclang [ngae umah].
 1SG AV.build house
 ‘I built a house.’
- b. [Ngae umah] icang.
 AV.build house 1SG
 ‘I built a house.’
- c. Ngae icang umah.
 AV.build 1SG house
 ‘I built a house.’ (Artawa 1998: 19)

Artawa (1998: 20) argues that constructions like (40c) indicate contrastive focus on the verb. Nonetheless, VSO is only possible in AV and not UV. Thus, Indonesian-type languages also differ in their degree of word order flexibility and the effect of voice on word-order patterns.

5.4.3 Transitional Languages

Finally, many languages in Borneo and Sulawesi are transitional between Philippine-type and Indonesian-type languages in terms of their word order. Some preserve Philippine-type verb-initial order, like *Tukang Besi*:

- (42) *Lundayeh*
- a. **Intransitive**
 Ne-langui ui guta abpa' di.
 PFV<INTR.AV>swim 1SG.1 across water PT
 'I swam across the river.'
- b. **Undergoer Voice**
 K<i>kam ipu' Bayoi Mengilo ticu' awan nidi
 <PFV.UV>grasp grandfather Bayoi Mengilo hand wife his
- pa-kabing neh, tinan nidi pa-tinueh neh,
 on-left 3SG.2 mother his on-right 3SG.2
- idi ideh langui.
 then 3PL.1 <INTR.AV>swim
 'Grandfather BM grasped his wife's hand on his left and his mother's
 on his right, then they swam.' (Clayre 2014: 132)

However, Clayre (2014) suggests that there is a tendency for the actor subject to precede the verb in AV clauses.

For other North Sarawak languages, word order is more fixed and SVO is the preferred word order in both AV and UV clauses. However, verb-initial orders are possible, as shown in Berawan and Sa'ban:

- (43) *Berawan* (Rejang-Baram)
- a. **Actor Voice (SVO)**
 Akkoh m-unyih tiu lum kam.
 1SG.1 AV-hide eggs in basket
 'I hide the eggs in the basket.'
- b. **Undergoer Voice (SVO)**
 Tiu nih n-unyih koh lum kam.
 eggs those UV-hide 1SG.2 in basket
 'I hid those eggs in the basket.'
- c. **Undergoer Voice (VOS)**
 Kanen noh buppuun?
 eat.UV.IMPF 2SG.2 porridge
 'Do you eat porridge?' (Clayre 2014: 135)

- (44) *Sa'ban* (Apad Uat)
- a. **Actor Voice (SVO)**
 Éek nnal ieh.
 1SG AV.see 3SG
 'I see him.'
- b. **Undergoer Voice (SVO)**
 Éek inal ieh.
 1SG UV.see 3SG
 'He saw me.' (Clayre 2005: 33)
- c. **Undergoer Voice (VOS)**
 Pi n-net nah Ra' Kueng
 after AV-climb EMPH ant giant

 i-tap Ra' ai hroel ssuek ai.
 UV-bite ant DET egg macaque DET
 'After Giant Ant had climbed up he bit the macaque on his testicles.'
 (Clayre 2014: 138)

In Berawan, it is possible that the choice between SVO and VOS depends on the definiteness of arguments, since SVO order in (43b) is used with a definite undergoer, *tiu nih* 'those eggs', whilst VOS order in (43c) is used with an indefinite, non-referential undergoer, *buppuun* 'porridge'. However, such an account does not extend to *Sa'ban* and further research would be needed to clarify the differences between the word orders in context.

Finally, Kenyah languages have basic SVO order but may mark a sort of UV construction using an SOV variant (see SUBSECTION 3.4.3). This would be similar to the pro=verb constructions described for Indonesian in SUBSECTION 5.4.2 and voice alternations without morphology in SUBSECTION 1.3.2:

- (45) *Kenyah*
- a. **Actor Voice (SVO)**
 Aké metong asu'.
 1SG hit dog
 'I hit a dog.'

b. **Undergoer Voice (SOV?)**

Asu' inih aké metong.
dog this 1SG hit
'I hit this dog.'

(Clayre 2014: 145)

Consequently, Clayre (2014) concludes that the languages of Sarawak are moving towards the fixed word-order patterns typical of Indonesian-type languages. In any case, much like Philippine-type and Indonesian-type languages, transitional languages also differ in their degree of word-order flexibility and the effect of voice.

5.4.4 Historical Change in Word Order

The discussion so far suggests that many Western Austronesian languages are not neatly classifiable as Philippine-type or Indonesian-type on the basis of their word order patterns. This follows from the fact that languages differ in their degree of word order flexibility and also the fact that many languages show different word order tendencies depending on the voice construction. This may, in turn, relate to semantic role, animacy/definiteness of arguments and information structure, as discussed in SUBSECTION 5.3.3. Perhaps a better approach is to consider possible historical changes between verb-initial Philippine-type orders and SVO Indonesian-type orders.

In fact, Western Austronesian languages are thought to have undergone a change in word order via the reanalysis of a topicalisation construction in Philippine-type languages as the basic order of grammatical functions in Indonesian-type languages. This could reflect a general tendency to place subjects first (see SUBSECTION 5.2) or be triggered via contact with predominantly SVO languages in mainland South East Asia, or an unknown substrate in insular South East Asia, as suggested in Donohue (2007a: 357). The summary of word-order patterns seems to fit with this suggestion, in that SVO word order was shown to correlate with

pragmatically marked constructions in Philippine-type languages, whilst it is basic in Indonesian-type languages. Interestingly, an intermediate stage in this change seems to involve the Actor Voice, and therefore word order changes may also be linked to voice and semantic roles. I return to this idea in relation to Kelabit in SUBSECTION 5.5.4.

5.4.5 Summary

In this section, I showed that Western Austronesian languages vary according to their basic word orders. The more conservative, Philippine-type languages tend to be verb-initial, whilst the languages of Indonesia tend to be SVO. This has been argued to represent a historical change driven by the reanalysis of a topicalisation construction as subject position. Hence, word order, much like voice alternations and pronominal systems, is perhaps better analysed as reflecting various historical changes rather than a straight-forward two-way typology.

Accordingly, within Philippine-type languages, Indonesian-type languages and transitional languages, there is variation in the degree of word-order flexibility and the factors that underlie word order choice. Some languages, like Tkdaya Seediq, have fairly rigid word order, whilst others, like Tagalog and Chamorro, allow flexible ordering following the predicate. Even those languages with fairly fixed order may allow variant orders depending on information structure. Finally, word-order choices also differ in a range of languages from Tagalog to Indonesian, depending on which voice construction is being analysed and whether the arguments are nominal or pronominal. Against this background, I now consider word order in Kelabit.

5.5 Word Order in Kelabit

This section outlines some key findings in relation to the word order of core arguments and the verb in Kelabit clauses. SUBSECTION 5.5.1 describes possible word orders in intransitive, AV, UV and bare verb clauses. SUBSECTION 5.5.2 addresses the question of which order is basic using frequency counts of the different voices in different genres. Finally, SUBSECTION 5.5.3 explores the role of information structure in determining word order variants by analysing the use of different word orders in the corpus and elicitation. Ultimately, I conclude that Kelabit word order is transitional in that UV appears more Philippine-type in its word order, whilst AV is more Indonesian-type.

5.5.1 Possible Word Orders

In order to elicit possible word orders, a word game was devised and tested with several consultants. The game involved colour-coded cards to represent the different word classes as follows:²⁴²

²⁴² If repeating the experiment, some of the colour coding could be changed. For example, the particles *teh* and *neh* seem to behave like sentence particles and should perhaps be yellow. It is not clear if the colour coding made any difference to the experiment.

Table 5.6 Word Order Elicitation Game

Colour	Category	Examples
Red	Verb	Intransitive verbs, e.g. <i>tudo</i> ‘sit’ Transitive AV verbs, e.g. <i>kuman</i> ‘eat’ Transitive UV verbs, e.g. <i>kinan/kenen</i> ‘eat’ Transitive IV verbs, e.g. <i>penekul</i> ‘use to spoon up’
Blue	Noun	Proper names, e.g. Peter Pronouns, e.g. FORM 1 <i>uih</i> and FORM 2 <i>kuh</i> Nouns, e.g. <i>la’ih</i> ‘man’, <i>tekul</i> ‘spoon’, <i>buaq kaber</i> ‘pineapple’
Green	Adjuncts	Time adverbs, e.g. <i>ngimalem</i> ‘yesterday’, <i>na’an</i> ‘later’ Place adverbs, e.g. <i>luun asuq</i> ‘on the stool’ Manner adverbs, e.g. <i>dooq-dooq</i> ‘well’ Adjectives, e.g. <i>rayeh</i> ‘big’
Yellow	Auxiliaries, Particles & Quantifiers	Auxiliaries, e.g. <i>laq</i> ‘want’, <i>kereb</i> ‘can’, <i>pengeh</i> ‘already’ Sentence particles, e.g. <i>tebeyq</i> , <i>men</i> , <i>tun</i> , <i>betuq</i> Question particle, e.g. <i>ken</i> Quantifier, e.g. <i>na’am</i> ‘negation’
Orange	Determiners	Indefinite/numeral, e.g. <i>edteh</i> ‘a/one’ Definite/specific/singular, e.g. <i>sinih</i> ‘this one’, <i>sineh</i> ‘that one’ Definite/plural, e.g. <i>(nuk) inih</i> ‘these’, <i>(nuk) ineh</i> ‘those’ Particles, e.g. <i>teh/neh</i> , <i>nih/dih/neh</i> , <i>ih</i> ‘highly identifiable’
White	Question words	<i>wh</i> -words, e.g. <i>ngapeh</i> ‘where’, <i>enun</i> ‘what’, <i>iih</i> ‘who’ Relativisers, e.g. <i>suk</i> ‘specific’, <i>nuk</i> ‘non-specific’

Participants were provided with this selection of cards and asked to create sentences, typically starting with a particular verb. From the sentences that they created, alternative orders were shown by moving the word cards around and judged for their grammaticality. In each instance, consultants were asked to offer a suggestion on the differences in use between potential orders. The possible orders of verb and core arguments were found to differ depending on the nature of the clause, i.e. whether the verb was intransitive or transitive and which voice was used.

5.5.1.1 Intransitive Clauses

Intransitive clauses are minimally composed of a verb and a subject. Typically, the verb is initial. However, it is also possible for the subject to precede the verb:

(46) *Kelabit Intransitives*

- a. **VS**
Tudo uih.
sit 1SG.1
'I sit.'
- b. **SV**
Uih tudo.
1SG.1 sit
'I sit.'

(elicitation, BAR18082014CH_01 00:22:10.740-00:22:13.210)

If the pre-verbal position is filled with an auxiliary, the most neutral order is Aux-S-V. However, it is possible for the subject to follow the verb or precede the auxiliary as seen in (47):

(47) *Auxiliaries*

- a. **Aux S V**
Laq uih tudo.
DESID 1SG.1 sit
'I want to sit.'
- b. **Aux V S**
Laq tudo uih.
DESID sit 1SG.1
'I want to sit.'

(elicitation, BAR18082014CH_01 00:32:17-730-00:32:22.640)

- c. **S Aux V**
Uih laq tudo.
1SG.1 DESID sit
'I want to sit.'

(elicitation, BAR18082014CH_01 00:33:30.450-00:33:32.540)

Thus, both SV and VS are equally possible. When the predicate is in initial position the most neutral order is VS. However, if any other element fills initial position, then SV is preferred.

5.5.1.2 Actor Voice Clauses

AV clauses are transitive and can have both a subject and a non-subject core argument. As illustrated in SUBSECTION 2.5.1, the actor of the AV clause is the subject and the undergoer is a non-subject core argument, which can be optionally omitted. Three orders are possible: SVO, VOS and VSO. This is true for both nominal and pronominal actors:

- (48) *Kelabit Actor Voice*
- a. **SVO**
[Uih] kuman buaq kabēr.
1SG.1 AV.eat fruit pineapple
'I eat pineapple.'
(elicitation, BAR18082014CH_02 00:07:02.570-00:07:04.550)
- b. **VOS**
Kuman buaq kabēr [uih].
AV.eat fruit pineapple 1SG.1
'I eat pineapple.'
(elicitation, BAR18082014CH_02 00:07:39.970-00:07:42.460)
- c. **VSO**
Kuman (neh) [uih] buaq kabēr.
AV.eat (PT) 1SG.1 fruit pineapple
'I eat pineapple.'
(elicitation, BAR18082014CH_02 00:07:43.330-00:07:47.160)
- (49) *Kelabit Actor Voice*
- a. **SVO**
[La'ih sineh] ne-kuman buaq kabēr.
man DEM PFV-AV.eat fruit pineapple
'The man ate pineapple.'
- b. **VOS**
Ne-kuman buaq kabēr [la'ih sineh].
PFV-AV.eat fruit pineapple man DEM
'The man ate pineapple.'

c. **VSO**

Ne-kuman [la'ih sineh] buaq kaber.
 PFV-AV.eat man DEM fruit pineapple
 'The man ate pineapple.' (Florance Apu p.c.)

The order VSO differs in acceptability. Some speakers do not find the order in (48c) acceptable, particularly without the particles *neh* or *teh*.²⁴³ For them, (48c) would mean that the pineapple is eating the speaker. However, a similar structure was used in the prosody experiment, described in CHAPTER 4 and repeated below. This was found to be acceptable by the participants:

(50) **VSO in Prosody Experiment**

Pu'un~pu'un ne-kuman [uih] [edteh buaq kaber
 REDUP~first PFV-AV.eat 1SG.1 a fruit pineapple

 nuk pelaba laam].
 REL very sour
 'First I ate a pineapple that was very sour.'
 (experiment, e.g. BAR18082014CH_03 00:02:22.280-00:02:25.890)

There could be a number of reasons for the different judgements in reference to (48c) and (50). Firstly, there may be dialect or speaker variation. Secondly, it is possible that the acceptability of VSO differs depending on the animacy/definiteness of the undergoer argument. In (48) and (49), the undergoer, *buaq kaber* 'pineapple', could be argued to be non-referential and non-specific. In (50), however, the undergoer is anchored in discourse as it is modified by a relative clause, *nuk pelaba laam* 'that is very sour' (see SUBSECTION 5.3.3). Finally, it is possible that word order in Kelabit is affected by prosodic weight, much like Tagalog (see SUBSECTION 5.4.1). VSO may be more acceptable when the undergoer NP is 'heavy', as in (50). These questions are

²⁴³ Sentence particles often have an information structure function (SUBSECTION 2.4.2.14.1). This supports the conclusion in SUBSECTION 5.5.3.3 that VSO is used for particular information structure purposes in AV.

readdressed in SUBSECTION 5.5.3. In any case, VSO order is considered acceptable by at least some speakers.

In contrast, all OV orders are judged ungrammatical and do not occur in the corpus:

(51) *Kelabit Actor Voice*

a. **OVS**

*Buaq kaber ne-kuman [uih].
 fruit pineapple PFV-AV.eat 1SG.1
 For: 'I ate pineapple.'

(elicitation, BAR18082014CH_02 00:08:01.770-00:08:05.590)

b. **OSV**

*Buaq kaber [uih] ne-kuman.
 fruit pineapple 1SG.1 PFV-AV.eat
 For: 'I ate pineapple.'

c. **SOV**

*[Uih] buaq kaber ne-kuman.
 1SG.1 fruit pineapple PFV-AV.eat
 For: 'I ate pineapple.'

(Florance Apu p.c.)

Thus, in AV there are three possible word orders, all of which are VO in Dryer's (2013b) typology (SUBSECTION 5.2.2). Out of context, VOS is offered as the most natural order. SVO clauses are typically considered less fluent, though offered in elicitation.²⁴⁴ However, if anything other than the verb fills the initial position then, like intransitive clauses, the subject will most likely precede the verb (see SUBSECTION 5.5.3).

²⁴⁴ There may well be speaker differentiation as to how natural SVO order sounds, which could reflect an ongoing change and would remain to be further explored. Anecdotally, it appears as though younger speakers prefer SVO. This could be due to the influence of Malay.

5.5.1.3 Undergoer Voice Clauses

In UV, the undergoer is subject and the actor is a non-subject core argument (SUBSECTION 2.5.1). Word order is flexible, but unlike AV, only two word orders are possible: VOS and SVO:

(52) *Kelabit Undergoer Voice*

a. **VOS**

Kinan kuh [buaq kaber ih].
UV.PFV.eat 1SG.2 fruit pineapple PT
'I ate the pineapple.'

(elicitation, BAR18082014CH_02 00:08:26.400-00:08.29.610)

b. **SVO**

[Buaq kaber ih] kinan kuh.
fruit pineapple PT UV.eat 1SG.2
'I ate the pineapple.'

(elicitation, BAR18082014CH_02 00:05:14.950-00:05.16.880)

Of these, VOS is considered most acceptable out of context and speakers suggest that examples like (52a) would be the most natural way of expressing a past tense proposition. This is almost always the order provided in elicited UV sentences.

All other orders, including VSO, are ungrammatical:

(53) *Kelabit Undergoer Voice*

a. **OVS**

*Uih kinan [buaq kaber].
1SG.1 UV.PFV.eat fruit pineapple
For: 'I ate pineapple.'

(elicitation, BAR18082014CH_02 00:04:46.970-00:04:51.680)

b. **OSV**

*Uih [buaq kaber] kinan.
1SG.1 fruit pineapple UV.PFV.eat
For: 'I ate pineapple.'

(Florance Apu, p.c.)

Hence, AV clauses appear more like intransitive clauses in that there is not necessarily a VP constituent (see SUBSECTION 3.5.1). In intransitive clauses, the verb can also appear alone in initial position. The subject can immediately follow the particle and be followed by an adjunct, as shown in (55):

(55) *Kelabit Intransitives*

a. **V Part S Adj**

[Tudo] teh uih luun asuq.
 sit PT 1SG.1 on stool
 ‘I’m sitting on the hearth-stool.’

(elicitation, BAR18082014CH_01 00:59:24.140-00:59:26.380)

Hence, intransitive and AV transitive clauses share certain word-order similarities. Whilst the actor non-subject of a UV clause is an obligatory argument of the VP, the undergoer non-subject of AV clause can be realised outside the VP, rendering the clause syntactically similar to an intransitive one.

One further interesting difference between Actor Voice and Undergoer Voice concerns their usage with preverbal auxiliaries:

(56) **Preverbal Auxiliaries**

a. [Kuman buaq kaber ngapeh] tebeyq laq Peter?
 AV.eat fruit pineapple where PT DESID Peter
 ‘Where does Peter want to eat pineapple?’

(elicitation, BAR19082014CH_03 01:06:23.236-01:06:27.319)

b. *[Kenen Peter kapeh] tebeyq laq buaq kaber?
 UV.IRR.eat Peter how PT DESID fruit pineapple
 For: ‘how does Peter want to eat pineapple?’

(elicitation, BAR19082014CH_03 01:10:51.978-01:10:59.153)

c. [Laq kuman buaq kaber ngapeh] tebeyq Peter?
 DESID AV.eat fruit pineapple where PT Peter
 ‘Where does Peter want to eat pineapple?’

(elicitation, BAR19082014CH_03 01:06:04.277-01:06:07.122)

- d. [Laq kenen Peter kapeh] tebeyq buaq kabey?
 DESID UV.IRR.eat Peter how PT fruit pineapple
 ‘How does Peter want to eat the pineapple?’
 (elicitation, BAR19082014CH_03 01:11:40.783-01:11:44.525)

Whilst both AV and UV clauses can have a constituent formed of the preverbal auxiliary and its VP complement before the particle, only AV allows an order in which the VP precedes the particle in initial position and the auxiliary + subject follow the particle.

A similar construction is found with intransitive predicates:

- (57) *Kelabit Intransitives*
 a. Tudo laq tuih.
 sit DESID PT=1SG.1
 ‘To sit, I would like.’
 (elicitation, BAR18082014CH_01 00:34:22.840-00:34:23.890)

Hence, AV and intransitive clauses appear to have greater word-order flexibility than UV. The UV clause has only two possible orders: VOS or SVO. VO forms a tight unit and cannot be separated from any pre-verbal auxiliaries.

5.5.1.4 Bare verbs

Finally, the clauses with the most flexible word order are those with bare verbs that are unmarked for voice, including *keliq* ‘know/see’ and *uwan* ‘own’ (SUBSECTION 2.4.2.2.2). Such clauses allow verb-initial and verb-medial orders. Moreover, the order of actor and undergoer is not constrained following the verb and both can appear in initial position, though the restriction against two arguments pre-verbally is maintained. This is illustrated in (58):

- (58) *Kelabit Bare Verbs*
- a. **Verb Actor Undergoer**
 Keliq neh emeyq ih?
 see 3SG.2 goat PT
 ‘Did he see the goat?’
 (text, BAR31072014CH_06 00:03:18.350-00:03:19.520)
- b. **Actor Verb Undergoer**
 Ideh ne-keliq puluq agur.
 3PL.1 PFV-see ten shot
 ‘They found ten bullets.’
 (text, BAR21082014CH_01 00:16:08.640-00:16:11.908)
- c. **Verb Undergoer Actor**
 Na’am keliq enun~enun teh kedideh.
 NEG know REDUP~what PT 3PL.EMPH
 ‘They don’t know anything at all.’
 (text, BAR31072014CH_06 00:10:32.405-00:10:34.775)
- d. **Undergoer Verb Actor**
 Ieh keliq deh selubid lem tanki sineh.
 3SG.1 know 3PL.2 lie in oil.rig DEM
 ‘They found him lying down on that oil rig.’
 (text, BAR21082014CH_01 00:17:31.910-00:17:35.406)

It is possible that unmarked predicates can be interpreted either in analogy with AV or in analogy with UV (see SUBSECTION 2.5.3). Indeed, the FORM 2 pronouns that are normally restricted to UV are used in (58a) and (58d). In any case, word order in bare verb clauses seems to be driven by information structure even more than in other clauses, as discussed in SUBSECTION 5.5.3. The following section explores which of the word-order variants is basic according to textual frequency.

5.5.2 Establishing Basic Order

In the previous section, I established that Kelabit has flexible word order and that word-order patterns are affected by the voice construction and possibly the animacy/definiteness of the arguments. In this section, I address the issue of basic

word order in Kelabit using textual frequency data in narratives and news reports. The narrative corpus is composed of six retellings of the pear story video (Chafe 1980). This involves one speaker watching a five-minute video that shows a series of events surrounding a man picking pears without any sound. The speaker then retells the story to another participant who hasn't seen the original video. There are roughly 33 minutes of pear story recordings and 757 clauses. The news reports corpus was taken from the twice-daily news reports transmitted via the community radio in Bario (SUBSECTION 2.2.4). This includes approximately 30 minutes of recordings and 640 clauses from two different newsreaders.

Following SUBSECTION 5.3.2, the frequency count is restricted to transitive clauses with two overt arguments and intransitive clauses with one overt argument in pragmatically neutral contexts, i.e. declarative, affirmative, independent clauses (Sierwierska 1988:8, SUBSECTION 5.3.2). The following clause-types were excluded:

- (59) **Excluded from Frequency Count**
- a. Negative clauses
 - b. Interrogatives
 - c. Imperatives
 - d. Periphrastic Constructions
 - e. Subordinate Clauses
 - f. Relative Clauses
 - g. Conjunct Clauses
 - h. Clauses with subject doubling

In such clauses, the initial position in the clause is typically filled with something other than the verb, such as a negative, interrogative particle or relative marker. This generally triggers SVO order by default (SUBSECTION 2.5) and could skew results.²⁴⁶

There were a number of cases where the subject was repeated twice, as in (60):

²⁴⁶ Nb. as illustrated in SUBSECTION 5.5.1.3, it is also possible for a negative or preverbal auxiliary to be followed by the verb rather than the subject, though typically SVO order results.

(60) **Subject Doubling**
 Neh [nieh] pakai edtan [tieh] menad buaq
 DEM PT=3SG.1 use ladder PT=3SG.1 AV.climb fruit

sineh neh.

DEM PT

‘Then he used a ladder to climb that fruit tree.’

(pear story, BAR01082014CH_02 00:00:11.770-00:00:15:440)

It is not clear if such clauses represent topicalisation, or are simply products of naturally occurring speech. Consequently, any constructions that require specific word orders or are not clearly analysable as either SVO or verb-initial are excluded from frequency counts.

Clauses with pronominal arguments are included in the frequency count, even though Siewierska (1988) defines basic word order as that found in clauses with full NP participants (SUBSECTION 5.3.2). This is because clauses with two full NP participants are very rare in the corpus. In the pear story corpus, only seven of the 105 examples of transitive clauses contained two nominal arguments, spread across five different stories. In all seven cases, the word order was SVO. However, seven examples are not sufficient to draw any meaningful conclusions. Hence, I consider the frequency of different word orders when pronominal and nominal arguments are classed together.

5.5.2.1 Frequency in Narratives

The pear story corpus contained 156 examples of intransitive clauses with an overt argument in pragmatically neutral contexts, and 105 examples of transitive clauses with two overt arguments in pragmatically neutral contexts. TABLE 5.7 shows the breakdown of word order in intransitive clauses. TABLE 5.8 shows the breakdown of

word order in transitive clauses and TABLE 5.9 shows word order in transitive clauses split by voice construction.

Table 5.7 Word Order in Narrative Intransitive Clauses

	SV	VS	Total
Number	42	114	156
Percentage	27%	73%	100%

Table 5.8 Word Order in Narrative Transitive Clauses

	SVO	VOS	VSO	OVS²⁴⁷	Total
Number	72	30	2	1	105
Percentage	68.5%	28.5%	2%	1%	100%

Table 5.9 Word Order in Narratives by Voice Construction

Voice	Word Order	Number	Percentage
AV	SVO	64/74	86 %
	VSO	2/74	3%
	VOS	8/74	11%
UV	SVO	4/22	18%
	VOS	18/22	82%
Bare verbs	SVO	4/9	44.5%
	VOS	4/9	44.5%
	OVS	1/9	11%

The results in TABLE 5.7 and 5.8 suggest that intransitive predicates in narratives are overwhelmingly verb-initial, whilst transitive predicates are overwhelmingly SVO. This supports the claim in Dryer (1997) that transitive and intransitive clauses can differ in their basic word order.

Furthermore, the results from TABLE 5.9 demonstrate that the frequency of word orders differs depending on the voice of the clause. In AV clauses, which

²⁴⁷ As discussed in SUBSECTION 5.5.1, OV order is only found in clauses with bare predicates and corresponds to the order Undergoer-Verb-Actor. If bare predicates are thought to allow both the mapping of actor to subject, and undergoer to subject, OVS could be thought of as SVO order.

constitute the majority of transitive clauses, SVO word order occurs 86% of the time. Verb-initial orders, including VOS and VSO, occurs only a small percentage. In contrast, in UV clauses, VOS order occurs 82% of the time, and SVO order only a small percentage. Clauses where the main verb is a bare verb, occur equally frequently with verb-initial and SVO order, perhaps supporting the idea that these can be analysed either in analogy with AV or in analogy with UV. Hence, textual frequency in narratives would suggest that Kelabit is much like Lundayeh and West Coast Bajau in that UV clauses have basic verb-initial order, whilst AV clauses have basic SVO order (SUBSECTION 5.4.3).

5.5.2.2 Frequency in News Reports

The news reports corpus contained 55 examples of intransitive clauses and 95 examples of transitive clauses with overt arguments in pragmatically neutral contexts. TABLES 5.10, 5.11 and 5.12 show the breakdown of word order by clause-type and voice construction:

Table 5.10 Word Order in News Report Intransitive Clauses

	SV	VS	Total
Number	49	6	55
Percentage	89%	11%	100%

Table 5.11 Word Order in News Report Transitive Clauses

	SVO	VOS	VSO	OVS	Total
Number	91	1	0	3	95
Percentage	97%	1%	0%	2%	100%

Table 5.12 Word Order in News Reports by Voice Construction

Voice	Word Order	Number	Percentage
AV	SVO	64/64	100%
	VSO	0	0%
	VOS	0	0%
UV	SVO	18/19	95%
	VOS	1/19	5%
Bare verbs	SVO	9/12	75%
	VOS	0	0%
	OVS	3/12	25%

Interestingly, the news reports paint a different picture from the data in SUBSECTION 5.5.2.1. Firstly, both intransitive and transitive clauses occur more frequently in SV order and the number of intransitive clauses is much lower, perhaps because equative clauses and relative clauses are often used in news reports but are excluded from the frequency counts, as discussed in SUBSECTION 5.5.2.

Secondly, and importantly, the percentage of clauses with SVO order is considerably higher in transitive clauses in the news reports corpus. This is largely because the majority of UV clauses in the news report corpus also have SVO order, unlike in the narrative corpus where verb-initial order was more frequent for UV. This suggests that word-order patterns not only vary by clause-type but also by genre.²⁴⁸

In summary, SVO could be argued to be the basic word order in Kelabit as it is the most frequent order overall, but word-order choice is affected by the voice construction and the text genre. In the following section, I explore a potential explanation for these differences and address the implications for Western Austronesian typology.

²⁴⁸ It is possible that the news reports are affected by the fact that they are translated from existing reports in English or Malay (both SVO languages). However, this would not explain the patterns of SVO in narrative contexts, for which there are examples in UV.

5.5.3 Factors Affecting Word Order in Kelabit

In SUBSECTION 5.5.2, I established that word order differs in intransitive, AV and UV clauses. I demonstrated that whilst AV is predominantly SVO, independent of genre, the choice of word order in intransitive and UV clauses differed in narratives and news reports. This leads to the hypothesis that SVO may be used as a pragmatically marked construction in UV. One of the key differences between the genres is that narratives tend to describe action and sequence, whilst news reports adopt more of a topic-comment structure. Hence, if word order reflects ‘newsworthiness’, then the action of the verb is arguably the most important aspect of narratives, whilst introducing new topics is arguably the most important aspect of news reports. This section explores the role of information structure in determining word order choice by analysing examples of the different orders in text and elicitation. This allows us to test the hypothesis that SVO is used for specific information structure purposes in UV, whereas in AV SVO is simply a default.

5.5.3.1 Information Structure and Elicitation

In SUBSECTION 5.3.3, I discussed two aspects of information structure that can affect word order: the division of the clause into topic and comment, and placing ‘newsworthy’ or prominent information first. This section demonstrates that both aspects of information structure play a role in Kelabit. I subsequently extend the analysis to naturally occurring data.

In Kelabit, like Seediq and Malagasy, there appears to be a preference for comment-topic order. Indeed, new information precedes old information following the verb, and focus information can appear in the immediately post-verbal non-subject position. This can be illustrated in the following question-answer pairs:

- (61) **Actor = Given, Undergoer = New**
- Q. Ne-kuman enun teh Peter ngimalem?
 PFV-AV.eat what PT Peter yesterday
 ‘What did Peter eat yesterday?’
- A. [Ne-kuman [buaq kaber]_{focus}]_{comment} [tiah]_{topic} ngimalem.
 PFV-AV.eat fruit pineapple PT=3SG.1 yesterday
 V_{AV} O S
 ‘He ate pineapple yesterday.’
 (elicitation, BAR19082014CH_03 00:12:43.548-00:12:50.769)

- (62) **Actor = New, Undergoer = Given**
- Q. Kenen iih buaq kaber dih?
 UV.IRR.eat who fruit pineapple DEM
 ‘Who will eat the pineapple?’
- A1. [Kenen [ieh]_{focus}]_{comment} [dih]_{topic}.
 UV.IRR.eat 3SG.1 DEM
 V_{UV} O S
 ‘He will eat it.’ (pointing)
 (elicitation, BAR19082014CH_03 00:07:13.834-00:07:23.418)

The questions in (61) and (62) elicit argument focus contexts. In (61), the undergoer is focus and is realised as the non-subject argument of an AV clause, directly following the verb. The actor is topic and follows the undergoer. In (62), the actor is focus and is realised as the non-subject argument of a UV clause, directly following the verb. The undergoer is topic and is realised after the actor. Hence, the immediate post-verbal position can be a focus position.

Like in Tkdaya Seediq, the final subject typically receives a topic interpretation. This is shown in (63) by the fact that question words cannot appear in this position:

- (63) **Final Topic Position**
- a. *Kenen Peter enun?
 UV.IRR.eat Peter what
 For: ‘What does Peter eat?’
 (elicitation, BAR19082014CH_03 00:19:43.260-00:19:45.880)

Although VOS is the preferred order in UV (SUBSECTION 5.5.2), it is not possible when the undergoer is a non-referential *wh*-word. This suggests that the clause-final element must be topical or given information. Hence, we could analyse Kelabit as having a comment-topic order, much like many other Western Austronesian languages.

However, both subject-topic and subject-focus constituents are also allowed in the preverbal position, as illustrated in SUBSECTION 5.5.3.2 and 5.5.3.3. Consequently, I argue that the principle of ‘newsworthiness’ plays a role in Kelabit. Mithun (1992) defined ‘newsworthiness’ as highlighting new information, establishing new topics or expressing contrasts (see SUBSECTION 5.3.3). Hence, it can explain the tendency for new before old, seen in (61), (62) and (63). It can also explain the tendency to place actors before undergoers in contexts where they are equally given or new:

(64) **Actor = given, Undergoer = given**

Q. Ken kinan muh nubaq ih?
 Q UV.PFV.eat 2SG.2 rice PT
 ‘Did you eat the rice?’

A. Mo, kinan kuh nidih.
 yes, UV.PFV.eat 1SG.2 PT=DEM
 V_{UV} O S
 ‘Yes, I ate it.’

(elicitation, BAR19082014CH_03 00:00:27.472-00:00:35.064)

(65) **Actor = new, Undergoer = new**

Q. Enun suk tu'en deh pemudtih nangey?
 what REL UV.IRR.do 3PL.2 behind over.there
 ‘What’s going on behind there?’

A. Ah, neh Peter kuman buaq kaber nangey terun
 Oh, DEM Peter AV.eat fruit pineapple over.there maybe
 S V_{AV} O

‘Peter is eating pineapple there perhaps’

(elicitation, BAR19082014CH_03 00:22:39.140-00:22:45.802)

In (64), both actor and undergoer are given in the context of the question. In (65), both actor and undergoer are unknown as the question elicits sentence focus.²⁴⁹ In both (64) and (65), the actor is ordered before the undergoer, though (64) is VOS UV and (65) is SVO AV. Actors tend to be human, and are often of greater importance in the discourse. Hence, if word order is determined by newsworthiness, this could explain why there is a tendency to place them first (see Mithun 1992).

Finally, ‘newsworthiness’ can explain a tendency to place important topics, focus information and contrastive information first. For example, (66) is an alternative answer to the question in (62), repeated below, with SVO rather than VOS order:

- (66) **Actor = New, Undergoer = Given**
- Q. Kenen iih buaq kaber dih?
 UV.IRR.eat who fruit pineapple DEM
 ‘Who will eat the pineapple?’
- A2. [Dih]_{topic} [kenen [la’ih dih]_{focus}]_{comment}
 it UV.IRR.eat man DEM
 S V_{UV} O
 ‘It will be eaten by the man.’
 (elicitation, BAR19082014CH_03 00:07:13.834-00:07:23.418)

The actor, which constitutes focus information, remains in the immediately post-verbal position. However, the undergoer topic is realised pre-verbally, resulting in SVO order as it established as topic. Similarly, there is a preference for expressing contrastive information in initial position. This is illustrated in (67), which was elicited using a picture stimulus scenario in which a woman was laughing and a child crying:

²⁴⁹ This was elicited in an attempt to find a sentence focus question akin to ‘what’s happening?’ in English. The consultant was asked to imagine that something was happening behind a screen and that the speaker wanted to know what was going on. This question and answer pair were suggested. The use of the final particle *terun*, which translates roughly as ‘maybe’, may reflect an evidential distinction between things that have been seen and things that are assumed to be happening behind a screen. The use of 3PL.2 *deh* marks a generic/indefinite referent – it does not necessarily signify a plural agent.

(67) **Subject = Contrast**

Q. Iih suk riruh?
who REL laugh
'Who is laughing?'

A1. Neh men dedtur sidih riruh~riruh keneh.
DEM PT woman DEM REDUP~laugh PT
S V

'Well obviously the lady is laughing.'

(elicitation, BAR20082014CH_02 00:02:53.000-00:02:58.840)

A2. #Riruh teh dedtur sidih.
laugh PT woman DEM
V S

For: 'The woman is laughing.'

(elicitation, BAR20082014CH_02 00:03:22.010-00:03:26.310)

In contexts where the subject represents contrastive focus, such as (67), verb-initial orders are strange. However, in contexts where the verb is contrasted, verb-initial orders are preferred:

(68) **Verb = Contrast**

Ngi anak sidih riruh ngi
there child DEM laugh there

kadiq [tangey~tangey teh edtur suk ruyung nedih].
but REDUP~cry PT woman REL with 3SG.POSS
V S

'The child is laughing there but the woman who is with him is crying.'

(elicitation, BAR20082014CH_02 00:01:31.870-00:01:40.380)

Hence, Kelabit clauses appear to be ordered via the principle of newsworthiness, much like Cayuga in SUBSECTION 5.3.3, as prominent information is placed in initial position. This could be thought of as a general principle that subsumes other tendencies, such as new before old, actor before undergoer and comment before topic.

In the following sections, I explore the extent to which this holds true of naturalistic texts in Kelabit and whether there is evidence to support the hypothesis

raised above that SVO in UV is driven by information structure, whilst in AV clauses SVO is a default and verb-initial orders are pragmatically marked.

5.5.3.2 Information Structure in UV

On the basis of elicitation, I proposed that the choice between verb-initial and SVO word order in UV clauses may be determined by the principle of newsworthiness. Hence, I predict VOS to be used when the verb and actor are of greater newsworthiness than the undergoer, and SVO to be reserved for marked contexts in which the undergoer constitutes prominent information. This is supported by naturalistic data and can be illustrated in the examples below.

Example (69) contains several UV clauses in SVO and is taken from a news report. The segment begins with the clause in (69a), which introduces the main topic of the news report: a girl who has been arrested for the murder of her mother. In (69b), the topic at the clausal level switches from the girl to the mother. The mother remains the topic in the following equative clause in (69c) and subsequently the report reverts back to talking about the daughter and her boyfriend and provides details of their arrest in (69e):

(69) **SVO in UV**

- a. Edteh anak dedtur beruh nuk beruh~beruh ngeren binala
 one child girl new REL REDUP~new pregnant UV.PFV.say
 deh nuk ne-ngatey tesineh nedih ngi Indonesia.
 3PL.2 REL PFV-AV.kill mother 3SG.POSS in Indonesia
 ‘A girl who had just become pregnant has been said to have killed her
 mother in Indonesia.’
- b. Inih natey deh ngi bawang Bali.
 DEM UV.PFV.kill 3PL.2 in place Bali
 ‘They killed her in Bali.’

- c. Dedtur sinih ieh neh Sheila von Wiese Mack, enem ngepuluq laak.
 woman DEM EQUATIVE Sheila von Wiese Mack, sixty year
 ‘That woman is Sheila von Wiese-Mack, 60 years old.’
- d. Anak nedih ieh neh Heather Mack, 19 laak, ngen kawan
 child 3SG.POSS EQUATIVE Heather Mack, 19 year with friend

 dela'ih nedih ieh neh Tommy Schaefer, 21 laak.
 boy 3SG.POSS EQUATIVE Tommy Schaefer, 21 year
 ‘Her daughter was Heather Mack, 19 years old, and her boyfriend was
 Tommy Schaefer, 21 years old.’
- e. Ideh inep deh edto keteluh igu suk malem.
 3PL.1 UV.PFV.catch 3PL.2 day third week REL last
 ‘They were arrested on Wednesday last week.’
 (news report, BAR21082014CH_01 00:03:09.188-00:03:42.729)

There are three UV clauses with SVO order in this segment, (69a), (69b) and (69e). In the first instance, the undergoer of the clause – the daughter – is established as topic. The second two instances involve a switch in topic. The switch topic is not necessarily new information. For example, the mother in (69b) is encoded with a proximal deictic demonstrative, *inih*, and the daughter and her boyfriend in (69e) are encoded with a pronoun, *ideh*, both of which correspond to high discourse accessibility (see Lambrecht 1994). Nonetheless, they can be considered newsworthy on account of the fact that the topic is switched (see Mithun 1992: 34, SUBSECTION 5.3.3).

Similarly, of the four examples of SVO order in the narrative corpus, two occur in answer to a question where the undergoer represents newsworthy information on account of being new. For example, the question in (70) was used to prompt a pear story retelling:

- (70) **SVO in UV**
- Q. Enun seni'er muh?
 what UV.PFV.see 2SG.2
 'What did you see?'
- A. [Edteh wayang]_{focus} sen'ier kuh na'ah...
 one video UV.PFV.see 1SG.2 before
 'I just saw a video...'
 (pear story, BAR31072014CH_06 00:00:09.640-00:00:15.950)

In (70), the undergoer, *edteh wayang* 'a video', represents the portion of the answer that corresponds to the question word. This is considered a good test for focus status (cf. Lambrecht 1994, Dalrymple & Nikolaeva 2011, SUBSECTION 5.3.3). Hence, SVO in undergoer voice seems to occur in contexts where the undergoer is newsworthy, either representing a topic shift or focus information.

In contrast, VOS order is used where both the actor and the undergoer are given in the discourse. For example, (71) is taken from a pear story. The segment begins after the main participants of the narrative – the man and the fruit – have already been introduced. This is reflected in the fact that the man is encoded through the FORM 1 pronoun *ieh* and the pears are referred to as *buaq nuk ineh* 'those fruits', using a demonstrative to encode the definiteness of the referent. Furthermore, (71c-d) presents a sequence of actions that could be considered foregrounded in the narrative:

- (71) **VOS in UV**
- a. Pakai edtan sineh nieh mey ngalap buaq nuk ineh.
 use ladder DEM PT=3SG.1 go AV.pick fruit REL DEM
 'He used that ladder to go and pick fruit.'
- b. Pengeh neh, eh ni'er uih, neh ieh temurun let dingi.
 after DEM eh AV.see 1SG.1 then 3SG.1 INTR.down from up.there
 'After that, eh I watched, then he climbed down.'
- c. Temurun ieh let dingi keyh.
 INTR.down 3SG.1 from up.there EXCL
 'So he climbs down from up there.'

- d. Senipa neh neh buaq nuk ineh.
 UV.PFV.pack 3SG.2 PT fruit REL DEM
 ‘And put the fruit away [in the basket].’
 [second speaker asks a question at this point]
 (pear story, BAR31072014CH_06 00:01:24.400-00:01:38.737)

The clause in (71d) is VOS. Both actor and undergoer are highly topical. However, the actor could be considered more topical than the undergoer and it is perhaps for this reason that the FORM 2 pronoun *neh* is used (see SUBSECTION 4.7). The most newsworthy piece of information in this segment is the sequence of actions. Hence, the verb is initial, followed by actor and undergoer in order of importance.

A similar analysis extends to the only example of VOS in the news report corpus. In (72c), both actor and undergoer are given. Arguably, the newsworthy information is the act of taking the man to hospital, despite the fact that he had already passed away:

- (72) **VOS in uv**
- a. Pu'un~pu'un duweh lun duweh beken miney muruq [...]
 REDUP~first two people two other PFV.go AV.clean
 ‘First of all, two people went to clean [the oil rig].’
- b. Ideh ne-gagap ne-ni'er edteh la'ih edteh
 3PL PFV-surprised PFV-AV.see one man one

 bekuer idih matey lem tanki sineh.
 coiled.up and dead on oil.rig DEM
 ‘They were shocked to find a man lying coiled up and dead on the oil rig.’
- c. Tu'uh peh kineh nuit deh ieh mey rumaq sakit Miri.
 true PT like.that UV.PFV.take 3PL.2 3SG.1 go hospital Miri
 ‘They took him to hospital anyway.’
- d. Tapi ieh pengeh da'at lem ineh pukul teluh.
 but 3SG.1 already bad in DEM strike three
 ‘But he had already passed by three o'clock.’
 (news report, BAR21082014CH_01 00:17:36.567-00:18:07.593)

Hence, VOS in UV places emphasis on the action, rather than the participants. Consequently, word order in UV clauses can be shown to be affected by information structure and, in particular, the principle of newsworthiness. When the actor and undergoer are both given in discourse and the verb is the most newsworthy element then VOS order is used. However, when the undergoer is particularly newsworthy, in conveying new information, a topic switch or contrastive information, then SVO order is used.

5.5.3.3 Information Structure in AV

SUBSECTION 5.5.3.2 demonstrated that newsworthiness was a strong determiner of SVO word order in UV clauses. In this section, I address whether the same can be said for AV clauses, where SVO is the dominant word order in both narrative and news report genres (SUBSECTION 5.5.2). If this were the case, then we would expect verb-initial orders to be used where both actor and undergoer are given, and SVO to be used only in contexts where the actor subject is particularly prominent or newsworthy. In fact, looking at examples of the different word orders from the corpus, it seems that verb-initial orders are used for particular discourse functions and that SVO is a default order. This can be illustrated with the following examples.

Verb-initial orders tend to serve information-structure purposes in AV clauses. VSO is very rare in the corpus, but occurs in situations where both actor and undergoer are given, much like VOS in UV (see SUBSECTION 5.4.1 and 5.4.3 for similar patterns in other Western Austronesian languages). The only clear example in the texts analysed is (73) from a pear story:

- (73) **VSO in AV**
- a. Neh nieh nipa~nipa lem takub.
 then PT=3SG.1 REDUP~AV.pack in pocket
 ‘Then he put them all into a pocket.’
- b. Edteh takub ngi pema’un batek nedih [...]
 one pocket at front stomach 3SG.POSS
 ‘There was a pocket in front of his belly.’
- c. Neh neh inan neh nipa buaq ih.
 DEM PT EXIST 3SG.2 AV.pack fruit PT
 ‘That was where he put the fruit.’
- d. Ngalap~ngalap tieh buaq.
 REDUP~AV.pick PT=3SG.1 fruit
 ‘So he was picking fruit.’
- e. Am teh munung nedih mawan lem da’un temidteh ih.
 NEG PT face 3SG.POSS be.seen in leaf sometimes
 ‘But you sometimes couldn’t see his face in the leaves.’
 (pear story, BAR31072014CH_06 00:00:53.278-00:01:21.065)

Like example (71), both actor and undergoer are given at the beginning of this segment, and expressed using pronouns and definite NPs. (73a) and (73d) could be considered foregrounded clauses that present key sequences of action in the storyline. The rest of the clauses, and the omitted segment indicated by [...] provide background information. Hence, the function of VSO in (73d) could be to return to the main storyline. The key difference between VSO AV and VOS UV is semantic. Whilst UV clauses like (71) tend to express completed, perfective actions, the AV clause in (73) has a durative, progressive interpretation, partly conditioned by the use of AV and partly by reduplication (see SUBSECTION 2.4.1.4). Hence, VSO in AV contexts may occur in similar situations to VOS in UV, namely where the verb is the most important or newsworthy information in the clause, but an imperfective/progressive interpretation is required.

VOS is used in contexts where the predicate, i.e. the verb and the undergoer, are portrayed as newsworthy. In some instances, VOS order is used to contrastively focus the predicate. This can be seen in (74). The segment describes a scene at the end of the pear story where the man watches as three young boys walk past eating pears. In this version of the pear story, the speaker imagines what the man must be thinking, namely that the three boys have stolen his fruit. However, the speaker concludes that the man must realise this cannot be true, since they are only eating the pears and not carrying the stolen baskets that he is missing:

- (74) **VOS in AV**
- a. Neh nieh bulat ni'er anak nuk ineh.
 DEM PT=3SG.1 wide-eyed AV.see child REL DEM
 'So he looked at those children with wide eyes.'
- b. Kurang-lebih tieh ngelinuh:
 less-more PT=3SG.1 AV.think
 'more or less, he must have been thinking.'
- c. Teyh, ken ideh teh ne-ngalap buaq kudih turun.
 EXCL Q 3PL PT PFV-AV.take fruit 1SG.POSS maybe
 'ah ha, was it them who stole my fruit then!'
- d. Kadiq nidih lit na'am idih.
 reason PT=DEM suddenly NEG present
 'Is that why it's not here all of a sudden.'
- e. Na'am metoq bakul tu'en deh nitin metoq koq.
 NEG PT basket UV.do 3PL.2 AV.carry PT EXCL
 'But they aren't carrying the baskets.'
- f. Kuman buaq ih tupu tideh.
 AV.eat fruit PT only PT=3PL.1
 'They are just eating fruit.'
- g. Adiq am tieh bu'uh ngedeh.
 so NEG PT=3SG.1 angry with.3PL.2
 'So he didn't shout at them.'

(pear story, BAR31072014CH_06 00:10:34.803-00:10:52.376)

The VOS clause in (74f) expresses a contrast, as indicated through the use of the particle *tupu* ‘only’. This is the only way of attaining such a reading, as it is the only voice construction in which verb and undergoer can appear together in initial position (see SUBSECTION 2.4.2.14.1).

VOS can also be used if the predicate (i.e. verb + undergoer) forms new information, or the comment given about an established topic. This is shown in (75):

- (75) **VOS in AV**
- a. Pengeh ineh, am dadan, mirat edteh anak i’it bah.
 after DEM NEG long INTR.appear one child small EXCL
 ‘Not long afterwards, a small boy appeared.’
- b. [Ngimet edteh tupi]_{comment} [tieh]_{topic}.
 AV.wear one hat PT=3SG.1
 ‘He was wearing a hat.’
 (pear story, BAR01082014CH_02 00:00:35.605-00:00:40.335)

In (75b), the verb and undergoer make up a comment relating to the small child, who had been established as a new topic in the previous clause. Hence, VOS tends to be used when the verb and undergoer form the most newsworthy constituent.

That brings us to SVO. By far the greatest number of transitive clauses are SVO AV clauses. In some cases, it could possibly be claimed that this order is used to establish a new topic – i.e. a particularly newsworthy actor. Indeed, this is probably true of all the examples of clauses with two NP arguments, which tend to occur at the beginning of narratives and news reports and appear to serve the function of introducing and establishing important participants. One example is (76):

- (76) **SVO in AV**
- a. [[Lun polis]_{topic} [ne-ngenep telah burur lemulun]_{comment}]_{focus}
 police PFV-AV.catch three body people
- nuk miney ngelai selapang [...] ngi rumaq Menteri
 REL PFV.go AV.shoot gun at house minister
- Datuk Seri Wong Sun Koh.
 Datuk Seri Wong Sun Koh
 ‘The police have caught three people who fired a shotgun at Datuk
 Seri Wong Sun Koh’s house.’
 (news report, BAR21082014CH_01 00:15:09.223-00:15:19.193)

The clauses in (76) begin a news report. Hence, subject, verb and object are all new information in the discourse. One could claim that the report is construed as being about the police, and that the actor subject is therefore particularly prominent and realised as a new topic in initial position, followed by the predicate comment.

However, there are also a number of cases where SVO does not seem to correlate with a newsworthy actor in the sense of Mithun (1992). In (77) for example, SVO is used in a context in which the actor is a continuing topic, and the most newsworthy information is arguably the verb, since it constitutes the answer to the question:

- (77) **SVO in AV**
- Q: Kapeh ieh muit dih remurut, rengaq dih telah bu’an [...]?
 how 3SG.1 AV.take DEM down if DEM three basket
 ‘How did he get them [the fruit] down, if it’s three baskets?’
- A: Kapeh uih mala [...], ieh nutuq buaq mey beneh.
 how 1SG.1 AV.say 3SG.1 AV.drop fruit go low
 ‘How do I say, he dropped the fruit to the ground.’
 (pear story, BAR02082014CH_01 00:01:25.300-00:01:32.280)

In (77), the man is the topic of the question and remains the topic of the answer. The predicate is what answers the question word *kapeh* ‘how’. Hence, we might reasonably

describe this as the most important part of the utterance. Nonetheless, SVO order is used, placing the actor first.

For this reason, I argue that SVO has been reanalysed as the default word order in AV, regardless of whether the actor is newsworthy. This would explain why the word order choice is predominantly SVO in texts of narrative and news report genre and why SVO order does not correlate as strongly with the discourse prominence of the subject as SVO in UV. It may well follow from the tendency to find SVO order when initial position is filled with something other than subject or predicate. An example is given in (78):

- (78) **SVO in AV**
- a. Edteh wayang seni'er kuh na'ah [...] lem ayuq
 one video UV.PFV.see 1SG.2 before in about
 edteh la'ih edteh.
 one man one
 'I just watched a video about a man.'
- b. Neh pu'un~pu'un nieh bukaq keyh.
 DEM REDUP~start PT=3SG.1 open EXCL
 'At the very start when it opens'
- c. Inan buaq lah.
 EXIST fruit EXCL
 'There's some fruit.'
- d. [Neh] la'ih sineh midtet~midtet buaq ih lah.
 then man DEM REDUP~AV.pick fruit PT EXCL
 'And the man is picking the fruit.'
 (pear story, BAR31072014CH_06 00:00:13.671-00:00:42.114)

It is very common in narratives for AV clauses to begin with the particle *neh*, which implies a sequence of action: first this and then that (see SUBSECTION 2.4.2.7). As discussed in SUBSECTION 5.5.1.2, when anything other than the verb appears in initial position, the subject typically follows directly in second position. Hence, the syntactic

context may determine the word order choice. Since these constructions are highly frequent in discourse, SVO could have been reanalysed as the basic order of subject, verb and object, even without the determining context of a filled initial position. This reanalysis may not have taken place in UV clauses, as they the initial position is most commonly filled with the verb rather than auxiliaries or negation, since the UV infix *-in-* also conveys perfective aspect (see SUBSECTION 2.4.1.2.3).

Consequently, information structure can be seen to have the following effects in AV clauses in the corpus. Verb-initial orders portray the verb or predicate as the most newsworthy piece of information and may result in contrastive readings. SVO, in contrast, does not seem to have specific information structure interpretations and can be used when the actor is prominent and newsworthy, but also when the actor is a continuing topic.

5.5.4 Summary

In this section, I addressed possible word orders in Kelabit, the relative frequency of the different orders and the factors that seem to determine word-order choice. I found that Kelabit is a VO language but allows flexible positioning of the subject. I also found that, like Tagalog and Balinese, AV clauses are more flexible than UV clauses, since they also allow VSO order in addition to VOS and SVO. As for basic word order, like West Coast Bajau, the most frequent word order overall was SVO. However, the relative frequencies differed according to voice construction and genre. SVO is most frequent in AV, regardless of context, but word order in UV depends on the genre. VOS is the most frequent order in narratives, but SVO is more frequent in news reports.

Finally, I explored the role of information structure in determining word-order choice in Kelabit. I argued that all clauses are subject to some general tendencies. For

example, new information tends to precede old information, actors tend to precede undergoers and prominent information tends to go first. However, the role of information structure differs by clause-type. In UV clauses, SVO order is pragmatically marked and occurs in contexts where the undergoer is particularly newsworthy, much like in Philippine-type languages. In contrast, in AV clauses, it is verb-initial orders that effect a particular pragmatic interpretation, such as predicate focus, much like Indonesian-type languages. Hence, word-order choices, flexibility and interpretations all differ by voice construction in Kelabit.

5.6 Conclusion

In this chapter, I discussed a second asymmetry between AV and UV in Kelabit, namely word order patterns. AV and UV have different word order preferences and possibilities, which is reflected in their use in naturalistic corpora and elicitation tasks. They also interact with information structure in different ways. In AV, SVO order can be argued to be basic and verb-initial orders are used to contrastively focus the predicate. In UV, however, VOS order is basic and SVO order is used in pragmatically marked contexts where the undergoer is more prominent than the actor. This could explain different frequencies in narrative and news report genres. If we assume that word order is constrained by newsworthiness/discourse prominence, then it is not surprising that VOS UV and SVO AV are the preferred word orders, since these are the main constructions that allow the actor to precede the undergoer and actors are much more likely to be discourse prominent than undergoers (see SUBSECTION 5.5.3.1).

This has two important implications. Firstly, it suggests that the reanalysis of SVO as basic order in Western Austronesian begins in AV and that voice is therefore important in historical word order changes. SVO may be preferred in AV because it is

in keeping with universal tendencies to present topics and actors first, but also keeps the predicate (verb + undergoer) as a constituent (see SUBSECTION 5.2). In UV, SVO order not only places the undergoer before the actor, but also leads to a discontinuous predicate, and is therefore cross-linguistically marked. That AV allows for word-order patterns that are cross-linguistically more common may also explain why AV is used more frequently in Kelabit discourse (SUBSECTION 3.5.3) and in a number of transitional languages (SUBSECTION 3.4.3.3). Since discourse frequency also contributes to analyses of clauses as basic/transitive (SUBSECTION 3.3.3), this may therefore be a trigger for the reanalysis of AV as active rather than antipassive, discussed in CHAPTER 3.

Secondly, word-order typology demonstrates yet again that the traditional two-way typology of Philippine-type and Indonesian-type cannot capture the full extent of variation in Western Austronesian. On the one hand, there is variation within Philippine-type and Indonesian-type languages in terms of possible word orders and word order flexibility. On the other hand, Kelabit and many other languages can be seen as transitional between the two groups in that basic word order appears to differ depending on the voice construction. AV has Indonesian-type properties, whilst UV has Philippine-type properties. Hence, word order, like pronouns and voice, suggest that Western Austronesian languages can be better categorised by looking at different parameters of variation, rather than simply assigning a language to one of two classes. This allows us to explore the interrelationships between word order, information structure and voice that can contribute to a better understanding of the historical changes that have taken place.

Consequently, the study of Kelabit word order, much like Kelabit pronouns and Kelabit voice, supports the idea that Western Austronesian languages vary to a

greater extent than implied by the two-way typology and reinforces the benefits of a parametric approach to synchronic and diachronic variation.

Chapter 6

Conclusion

6.1 Introduction

In this thesis, I have analysed the structure of Kelabit in relation to other Western Austronesian languages in order to address the implications that this has for Western Austronesian typology, as well as ongoing theoretical and historical debates. In CHAPTER 1, I introduced the idea that Western Austronesian languages can be defined typologically as those languages in Taiwan, the Philippines, Malaysia, Indonesia, Madagascar, Borneo, and Sulawesi that possess ‘symmetrical voice’ alternations. In other words, Western Austronesian languages, in contrast to asymmetrical voice languages around the world, possess two or more voices that are syntactically transitive and morphologically marked. Nonetheless, Western Austronesian languages differ along a number of parameters and this has led to a distinction between Philippine-type and Indonesian-type languages that has become prevalent in the literature (cf. Himmelmann 2005a, Arka & Ross 2005).

Subsequently, I outlined two major theoretical debates within Austronesian syntax that relate to the symmetrical voice systems. These are the subject debate and the alignment debate. The subject debate concerns the question of whether subjects and other grammatical functions are relevant notions in Western Austronesian

languages. This is in light of the fact that typical subject properties are split between the argument indicated in the voice morphology (the ASV) and the actor semantic role. The alignment debate relates to the question of whether Western Austronesian languages have accusative alignment, ergative alignment or an alternative system of alignment altogether. It rests largely on the degree to which the voice alternations are viewed as symmetrical and how alignment systems are defined (see CHAPTER 3). I showed that arguments have been proposed for each alignment system in both Philippine-type and Indonesian-type languages. I also introduced the hypothesis that Western Austronesian languages have undergone a change in alignment from ergative in the more conservative Philippine-type languages to accusative in the more innovative Indonesian-type languages.

In the rest of the thesis, I set out to explore these debates through an analysis of the structure of Kelabit, a Western Austronesian language of Northern Sarawak that is spoken in a transitional area between Philippine-type and Indonesian-type languages. Kelabit offers a unique opportunity to understand typological variation in Western Austronesian as it belongs to the Apad Uat subgroup of North Sarawak. This includes languages with Philippine-type characteristics, such as Lun Bawang/Lundayeh, and languages with more innovative characteristics, such as Sa'ban, that have much in common with Indonesian-type languages (Clayre 2005, SUBSECTION 2.2.1).

I addressed the subject debate in CHAPTER 2, where a preliminary grammar sketch of Kelabit was presented. I subsequently addressed the alignment debate in CHAPTERS 3, 4 and 5 via more detailed case studies of voice, pronouns and word order in Kelabit. In each case, the patterns in Kelabit were compared with other Western Austronesian languages, including proto-typically Philippine-type languages, proto-

typically Indonesian-type languages, and a series of languages that appear transitional between the two extremes. This functioned as a vehicle for evaluating the traditional two-way typology of Western Austronesian. It also allowed me to evaluate different arguments in the alignment debate and consider the implications for theories of historical change. In this chapter, I summarise the main findings of the research.

The chapter is structured as follows. SUBSECTION 6.2 summarises the findings in relation to Western Austronesian typology. SUBSECTION 6.3 summarises the findings in relation to wider theoretical and historical debates and SUBSECTION 6.4 details possibilities for future research.

6.2 Western Austronesian Typology

Throughout the thesis, I have emphasised that Western Austronesian languages are typically split into Philippine-type and Indonesian-type systems. The distinction is made on the basis of different structural properties (SUBSECTION 1.3.1). Arka & Ross (2005: 7) summarise the differences between Philippine-type and Indonesian-type as follows:

(1) a. **Philippine-type**

Languages with multiple voice types, marked by verbal morphology and often accompanied by case marking of free nominal arguments. There is always one actor voice, which is either intransitive or lower in transitivity than the other voices, which are conveniently grouped as undergoer voices. [They] allow noun phrases with a variety of semantic roles to become subject: patient, theme, location, instrument, beneficiary etc. (Arka & Ross 2005: 7)

b. **Indonesian-type**

Languages conventionally analysed as having two voices, actor and undergoer, supplemented by applicative suffixes which allow locations, instruments, beneficiaries and noun phrases of other semantic roles to become the undergoer. (Arka & Ross 2005: 7)

This provides a useful model for distinguishing between languages like Tagalog and Indonesian in CHAPTER 1, which share the property of symmetrical voice alternations, but differ in a number of other ways (see TABLE 1.2). For example, Tagalog has case-marking of nominal arguments and verb-initial word order, whilst Indonesian has no case-marking and SVO basic word order.

However, Himmelmann (2002: 8) notes that though the two-way typology ‘provides a useful start for investigating the (internal) typology of Austronesian languages’ it also requires ‘a lot more empirical scrutiny’. In this thesis, I subjected the typology for further empirical scrutiny by exploring differences between Philippine-type and Indonesian-type languages in relation to voice systems, pronoun/clitic systems and word order. What I found was that the typology is inadequate as a means of capturing the full extent of variation in Western Austronesian. This is based on two main findings: firstly, the fact that both Philippine-type and Indonesian-type languages are subject to internal variation along a variety of parameters; and, secondly, that there are a number of languages, including Kelabit, that differ in a non-superficial manner from both Philippine-type and Indonesian-type languages. In the following subsections, I review the typological differences found in relation to each of the three case studies.

6.2.1 Voice and Alignment

In CHAPTER 3, I argued that the grammatical feature ‘voice’ should be understood as a system of alternations in the syntactic, semantic or pragmatic status of arguments that is expressed in the verbal morphology or the morphosyntactic construction as a whole. Consequently, I argued that a comparison of voice systems must take into account the levels of morphology, syntax, semantics and discourse. Furthermore, I suggested that identifying the alignment in a given language involves identifying which voice construction represents the basic transitive clause-type. Each level of structure can provide independent evidence for treating a particular voice as basic, which supports the view that languages can differ in their degree of symmetry (see Riesberg 2014).

The claim introduced in (1) is that the key difference between Philippine-type and Indonesian-type languages is in terms of the number of voice alternations: Philippine-type languages have more than two alternations, whilst Indonesian-type languages have two generic AV and UV constructions. This claim is broadly accurate, as illustrated in CHAPTER 3. However, it implies greater homogeneity than is found in Western Austronesian. In fact, not all Philippine-type languages have the same number of voices and not all two-way voice systems have Indonesian-type features. As to the first point, PAn is reconstructed as having a four-way system of alternations, which is preserved in a number of languages of the Philippines, including Tagalog and Cebuano (see Ross 2002, SUBSECTION 1.3 and 3.2.1.3). However, other languages with typical Philippine-type characteristics, such as case-marking and mood-marking morphology, have three-way voice systems, which differ according to the nature of the third voice. For example, Kavalan has an instrumental voice, but no locative or benefactive voice, whilst Kadazan Dusun has a benefactive voice, but no instrumental voice (SUBSECTION 3.4.1). As to the second point, there are languages like Sa’ban and

Kayan in Northern Sarawak that have two-way voice systems but do not have typical Indonesian properties like applicatives and true passives (SUBSECTION 3.4.3). Hence, the number of voices does not necessarily allow us to make insightful typological predictions.

A somewhat deeper claim is that Philippine-type and Indonesian-type languages differ in terms of their alignment (Aldridge 2011). As discussed above, this translates into the claim that Western Austronesian languages differ in terms of which voice represents the basic transitive clause, schematised in (2):

- | | | |
|-----|--------------------------------|-----------------------------|
| (2) | Philippine-type | Indonesian-type |
| | a. UV is basic/transitive | AV is basic/transitive |
| | b. AV is lower in transitivity | UV is lower in transitivity |
| | c. Alignment = Ergative | Alignment = Accusative |

In order to analyse this claim, I compared the voice systems at the levels of morphology, syntax, semantics and discourse. Again, comparison of the two groups broadly supported the schema in (2). However, it painted a more nuanced picture and sometimes provided conflicting results. As discussed in SUBSECTION 1.3, a defining characteristic of all Western Austronesian languages is that they are morphosyntactically symmetrical. In other words, each voice construction can be analysed as morphosyntactically transitive, with two core arguments. Nonetheless, the discussion in CHAPTER 3 revealed that there is morphosyntactic variation. Both Philippine-type languages, such as Pangutaran Sama, and Indonesian-type languages, like Balinese, have morphological asymmetries. Typically, UV is unmarked, in contrast to AV. Moreover, syntactic patterns may support the analysis of one voice as more basic than the other. For example, in Kapampangan person-marking shows agreement with two core arguments in UV, but only the actor in AV, which might suggest an analysis of UV as basic. Contrastively, in Balinese quantifier-floating can be launched

by both actor and undergoer in AV, but only the undergoer subject in UV, which might support an analysis of AV as basic. Thus, like Riesberg (2014) suggests, Western Austronesian languages appear to differ in the degree of morphosyntactic symmetry between the voices.

The levels of semantics and discourse support an analysis by which Philippine-type and Indonesian-type languages differ in their alignment. Semantic and discourse tests suggest that UV is basic in Philippine-type languages, and that AV is lower in transitivity. Conversely, semantic and discourse tests in relation to Indonesian suggest that AV is basic, and that UV is lower in transitivity. However, much like the number of voices, Western Austronesian languages differ in terms of the semantic and discourse properties associated with AV and UV. In Philippine-type languages, UV always tends to have semantic interpretations and discourse properties that correlate with high transitivity. However, the treatment of AV differs. In languages like Tagalog, there is a strong constraint against definite undergoers, which fits with an analysis of AV as low in transitivity or antipassive-like. In contrast, in Cebuano, the contrast against definite undergoers is less strong (SUBSECTION 3.4.1.2). In Indonesian-type languages, AV tends to have semantic and discourse properties associated with active clauses and there is no constraint against definite undergoers. Instead, the treatment of UV differs. In some languages, UV has semantic and discourse properties associated with high transitivity. However, in other cases, UV has the semantic and discourse properties of passives (SUBSECTION 3.4.2.3). Hence, there is variation in both Philippine-type and Indonesian-type languages in terms of the semantic and discourse properties associated with each voice.

Finally, there are various languages in Sulawesi and Borneo, including Kelabit, whose voice systems are neither proto-typically Philippine-type nor Indonesian-type.

The voice alternations involve a mixture of Philippine-type and Indonesian-type properties, as well as unique morphosyntactic features of their own. Moreover, semantics and discourse properties also differ from other Western Austronesian languages. Taking Kelabit as an example, UV has similar discourse and semantic properties to Philippine-type UV constructions. In contrast, AV has mixed results depending on what level of structure is compared. Like Philippine-type languages, AV tends to have characteristics of low semantic transitivity. However, unlike Philippine-type languages there is no constraint against definite undergoers, and AV clauses have discourse properties associated with basic, active clauses. Hence, Kelabit and other transitional languages appear even more symmetrical than Philippine-type and Indonesian-type languages. This can be summarised as follows:

- (3) **Kelabit**
- a. UV is transitive/basic (but less frequent in discourse)
 - b. AV is transitive/basic (but has some semantic properties of low transitivity)
 - c. Alignment is transitional between ergative and accusative

Consequently, distinguishing between two typological groups – Philippine-type and Indonesian-type – does not capture the full extent of variation within Western Austronesian voice systems. It neither captures the surface level morphosyntactic differences, nor the more fundamental differences in terms of alignment and which voice constitutes the basic transitive clause. Hence, it is better to look at differences in terms of morphology, syntax, semantics and discourse properties, and what implications these have for how proto-typically transitive the different voices are.

6.2.2 Pronouns and Clitic Systems

The second case-study considered variation in terms of pronominal systems in Philippine-type and Indonesian-type languages (see CHAPTER 4). There are two key differences between pronouns in Western Austronesian languages: case-marking and clitic status. Philippine-type languages typically have case-marking for all nominal arguments, including pronouns. As discussed in CHAPTER 4, the main case oppositions are traditionally analysed as nominative and genitive, in keeping with a symmetrical analysis of voice. However, the forms can also be understood as marking an ergative-absolutive case system (SUBSECTION 4.2.1.2). NOM pronouns are used for subject functions, and GEN pronouns for non-subject actors, and sometimes all non-subject core arguments, as in Kimaragang (Kroeger 2005). Both NOM and GEN pronouns are analysed as second-position enclitics (see Billings & Kaufman 2004). Indonesian-type languages, in contrast, do not have case-marking of nominal arguments. However, variant pronouns are used to represent first and second person, non-subject actors in UV contexts (SUBSECTION 4.3.2.1). These are cognate to GEN pronouns in other Austronesian languages, but are not typically analysed as genitive. Instead, they are thought to be clitic pronouns and contrast with the more widely used free-standing pronouns, which are cognate with NOM pronouns elsewhere in Western Austronesian. The non-subject actor clitics in Indonesian are both proclitic and verb-adjacent, in contrast to the conservative Wackernagel enclitics found in the Philippines.

Hence, the two typological groups also differ in terms of whether there is case-marking and what sort of pronominal clitics are used:

- | | | |
|-----|---------------------------|--------------------------|
| (4) | Philippine-type | Indonesian-type |
| a. | case-marking | no case-marking |
| b. | second-position enclitics | verb-adjacent proclitics |

However, analysis of Kelabit reveals that Western Austronesian languages differ in more ways than simply the presence or absence of case-marking and the two clitic subtypes, discussed in (4). In terms of case, Kelabit has two variant pronouns, which are described as FORM 1 and FORM 2 (see SUBSECTION 2.4.2.8). Loosely, these are used to indicate subjects (FORM 1) and non-subject actors (FORM 2), much like NOM and GEN pronouns in Philippine-type languages. However, a Philippine-type case-based analysis cannot be applied to Kelabit for the following reasons. Firstly, FORM 1 is also used for non-subject functions, including as an alternative to FORM 2. Secondly, FORM 2 pronouns are used to express actor subjects for certain intransitive and experiential transitive predicates (SUBSECTION 4.2.2). Hence, the difference between the pronouns is better understood as differential marking rather than as case in the strictest sense.

In terms of clitic status, prosodic and syntactic analyses also reveal differences between Kelabit FORM 1 and FORM 2 pronouns and both Philippine-type and Indonesian-type systems. FORM 2 pronouns are similar to Indonesian-type clitics in that they are exclusively verb-adjacent clitics. However, FORM 1 pronouns differ somewhat from Indonesian-type free-standing pronouns. Firstly, they can also be prosodic clitics when they occur in the immediately post-verbal position. Secondly, they have some syntactic characteristics of second-position enclitics in that they can appear immediately following a negative or pre-verbal auxiliary (SUBSECTION 4.3.3). Moreover, both FORM 1 and FORM 2 pronouns are enclitic to their syntactic hosts, rather than proclitic. Nonetheless, they attach prosodically to the following prosodic word. Hence, there is a mismatch between syntax and prosody. It remains to be seen if this also applies to other Western Austronesian languages. The Kelabit patterns can be summarised as follows:

- (5) **Kelabit**
- a. no case-marking (but differential marking with FORM 1 and FORM 2)
 - b. FORM 2 is a verb-adjacent clitic, that is prosodically proclitic and syntactically enclitic
 - c. FORM 1 is sometimes clitic and sometimes free-standing

A similar system is also found in the Kulawi language of Sulawesi (see SUBSECTION 4.3.2). Thus, a two-way typology also obscures differences between Western Austronesian clitic systems. SUBSECTION 4.3.2 illustrates that languages differ not only in terms of clitic position, but also in terms of whether NOM and GEN pronouns differ in their clitic status. Moreover, the pronouns may have different analyses (proclitic vs enclitic) depending on whether they are studied from a syntactic or a prosodic point of view. Hence, it may be better to consider variation along a number of parameters, rather than attempting to fit the variety of different clitic systems into two overarching umbrellas.

6.2.3 Word Order

Finally, CHAPTER 5 considered variation in Western Austronesian word order. In the literature, it has been suggested that a key difference between Philippine-type and Indonesian-type languages is word order (Donohue 2007a). This can be schematised in (6):

- | | | |
|-----|------------------------|------------------------|
| (6) | Philippine-type | Indonesian-type |
| a. | verb-initial order | SVO order |

However, I proposed in CHAPTER 5 that any analysis of word order should consider not only basic word order but also alternative possible orders and the factors that affect word order choice. In particular, I explored the role of animacy/definiteness, semantic roles and information structure in word order variation across Western Austronesian.

Yet again, Western Austronesian languages differ in their word order patterns to a greater extent than is predicted by the two-way typology. Firstly, both Philippine-type and Indonesian-type languages differ in their word order flexibility. Some Philippine-type languages, like Seediq, are fixed VOS languages. They only allow variant orders, like SVO, in highly marked pragmatic contexts – and even then it might be possible to analyse such structures as biclausal (SUBSECTION 5.4.1). Other Philippine-type languages, such as Tagalog, are alternating VOS/VSO languages. Similarly, some Indonesian-type languages, like Madurese, are fixed SVO languages. However, other Indonesian-type languages allow other word order variants, depending on information structural context, as in Riau Indonesian, or the animacy/definiteness of arguments, as in the bare UV construction in Standard Indonesian (SUBSECTION 5.4.2). Secondly, languages differ in the extent to which word order patterns are affected by voice construction. In fixed word order languages, like Seediq and Madurese, basic word order is the same in both AV and UV. However, in flexible word order languages, like Tagalog and Balinese, both possible word orders and word order preferences may differ according to the voice construction. Finally, even within verb-initial languages, there are different word-order correlations. In particular, some verb-initial languages allow *wh*-first questioning of adjuncts, whilst others do not (SUBSECTION 5.4.1). This has led to a number of different theoretical analyses. Hence, a two-way typology does not account for all the word order differences in Western Austronesian languages, which may correspond to deeper structural differences beyond the surface level word order variations (see Aldridge 2006).

Furthermore, Kelabit word order patterns support an analysis of Kelabit as intermediate in the transition from verb-initial to SVO language. In keeping with the analysis of AV as innovative in CHAPTER 3, AV clauses tend to be SVO order. That is,

SVO is the most frequent word order both in clauses where the subject follows an element such as a negative or pre-verbal auxiliary, and in clauses where the subject is clause-initial. Verb-initial orders also occur, but typically in contexts where the VP is particularly important or newsworthy. In contrast, UV clauses, which preserve many of the Philippine-type semantic and discourse properties, tend to have basic VOS order. Moreover, they are used in contexts where both actor and undergoer are given and typically correspond to foregrounded clauses. SVO order occurs as a variant, but mainly in contexts where the undergoer is particularly newsworthy. These patterns can be summarised as follows:

- (7) **Kelabit**
- a. AV has basic SVO word order
 - b. UV has basic verb-initial order

Hence, Kelabit has mixed properties of Philippine-type and Indonesian-type languages. AV has the word order patterns of an Indonesian-type languages but UV has the word order patterns of a Philippine-type language. Moreover, similar patterns are identified in a number of languages in Borneo, including West Coast Bajau and Lundayeh (SUBSECTION 5.4.3). Consequently, word order appears to be another area of Western Austronesian syntax in which a two-way typology does not do justice to the variation found.

6.2.4 Summary

In summary, the preliminary survey of voice systems, pronominal systems and word order in Kelabit and other Western Austronesian languages demonstrates that a two-way classification of languages into Philippine-type and Indonesian-type does not accurately reflect the level of variation within Western Austronesian. A better

approach seems to be to consider the semantic and discourse properties of each of the voice constructions, as well as the different morphological forms and syntactic word orders associated with each voice. In each case, there are a series of parameters that need to be considered, some of which are summarised in (8):

- (8) *Parameters of Variation in Western Austronesian*
- a. **Voice Systems**
 - Number of Voices
 - Degree of Morphosyntactic Symmetry
 - Semantic & Discourse Characteristics of UV (ergative to passive)
 - Semantic & Discourse Characteristics of AV (antipassive to active)

 - b. **Pronominal Systems**
 - Case-marking (canonical to differential marking to absent)
 - Clitic subtype (Wackernagel to verb-adjacent to affix)
 - Clitic position (enclitic or proclitic)
 - Syntax-prosody mismatch (yes or no)
 - Differences and similarities between NOM and GEN pronouns

 - c. **Word Order**
 - Word Order Flexibility (fixed to flexible)
 - Basic Word Order (verb-initial to SVO)
 - Differences and similarities between AV and UV
 - Role of Information-Structure in different configurations

There may well be additional parameters of variation and additional possibilities within existing parameters, as discussed in SUBSECTION 6.4. A better understanding of the synchronic variation will enable more detailed and more accurate accounts of diachronic change. Nonetheless, the parameters listed in (8) have several implications for historical models of syntactic change and wider theoretical debates within Austronesian syntax, and I address these in SUBSECTION 6.3.

6.3 Wider Theoretical and Historical Debates

In SUBSECTION 6.2, I summarised the findings of the three case-studies in this thesis, all of which support an analysis of Kelabit as intermediate between the more conservative Philippine-type languages and the more innovative Indonesian-type languages. In this section, I discuss the implications that the study of an intermediate-type language like Kelabit has for the subject and alignment debates (SUBSECTION 1.4).

6.3.1 The Subject Debate

As shown in SUBSECTION 1.4.1.1, in Tagalog typical subject properties are split between the ASV and the actor. This led Schachter (1976) to conclude that subject was not a viable category in the languages of the Philippines. However, largely similar patterns are also found in Indonesian (SUBSECTION 1.4.1.2) and Kelabit (SUBSECTION 2.5.1.1). Hence, split subject properties appear to be a characteristic of all Western Austronesian languages, just like symmetrical voice alternations (cf. Blust 2013). The split properties are summarised for Tagalog, Indonesian and Kelabit in (9):

- (9) a. *Tagalog*
- | | |
|--------------------------------|------------------------|
| ASV = Subject | Actor = Subject |
| Obligatory Argument | Reflexivisation |
| Launching Floating Quantifiers | Control |
| Relativisation | Imperative Addressee |
| Agreement Marking | |
- b. *Indonesian*
- | | |
|----------------------|------------------------|
| ASV = Subject | Actor = Subject |
| Relativisation | Reflexivisation |
| Control | |
| Raising | |

c. <i>Kelabit</i>	
ASV = Subject	Actor = Subject
Relativisation	Reflexivisation
Control	
Co-ordination	
Pre-verbal position	
Post particle position	

There are some striking similarities between Tagalog, Indonesian and Kelabit. In all three languages, reflexives are bound by the actor, even in non-actor voices. This – and the fact that actors are cross-linguistically more likely to be mapped to subject – are the main support for treating the actor as subject in Western Austronesian. However, reflexives also tend to be bound by the actor in syntactically ergative languages, like Inuit, which have been analysed as having undergoer subjects, following the inverse approach in Manning (1996). Moreover, practically all other tests suggest that the argument signalled in the verbal morphology (ASV) is subject.²⁵⁰ Consequently, the study of Kelabit provides additional support for the proposal in SUBSECTION 1.4.1.3 that reflexivisation is a property of the highest argument at argument structure and that it is ‘reference related’ tests, such as relativisation, that identify the highest grammatical function or ‘subject’ (cf. Manning 1996).²⁵¹ Since these tests inevitably identify the ASV as subject, this entails treating the voice alternations as alternations in the mapping of semantic arguments to grammatical functions. Hence, a theoretical conclusion supported by this thesis is that Western Austronesian voice alternations – whether Philippine-type, Indonesian-type or

²⁵⁰ See Riesberg (2014) and Kroeger (1993) for discussion of control in Tagalog.

²⁵¹ Equally, it could provide support for Falk’s (2006) model in which ‘subjects’ are subdivided into the highest grammatical function and a sentence-level pivot (SUBSECTION 1.4.1.3). Evaluating the proposals in Manning (1996) versus Falk (2006) depends on whether it is important to maintain that grammatical functions and semantic roles align. Since Manning (1996) defines ‘subject’ purely in terms of ‘pivot-like’ properties and deals with semantic properties at the level of argument structure, it does not require stipulating grammatical functions that automatically align with semantic roles. I therefore find this a simpler approach to Western Austronesian.

Kelabit-type – are not as dissimilar from canonical voice systems like active/passive and ergative/antipassive, in that they also involve a remapping of arguments to functions, but one that is symmetrical rather than asymmetrical (see SUBSECTION 3.2.1).

6.3.2 The Alignment Debate

The second major debate introduced in CHAPTER 1 is the nature of alignment in Western Austronesian languages. There have been three main hypotheses proposed:

- (10) **Western Austronesian Alignment**
 - a. The Accusative Hypothesis
 - b. The Ergative Hypothesis
 - c. The Philippine-type Alignment/Symmetrical Hypothesis

The accusative hypothesis rests on UV being treated as a passive construction. The ergative hypothesis rests on AV being treated as an antipassive. Finally, the symmetrical hypothesis rests of both AV and UV being treated as equally transitive. In SUBSECTION 1.4.2, a range of morphosyntactic arguments were put forward for treating both AV and UV as having two core arguments in the languages of the Philippines and Indonesia. In SUBSECTION 2.5.1, it was shown that similar arguments can be made for Kelabit. Hence, Western Austronesian languages cannot be either ergative or accusative in the traditional sense. That is, since both passives and antipassives involve syntactic detransitivisation, and Western Austronesian symmetrical voice alternations do not, it follows that Western Austronesian AV and UV constructions cannot be either passives or antipassives and that alignment cannot be proto-typically ergative or accusative.

In CHAPTER 3, I returned to the alignment debate in light of a broader definition of concepts such as active, passive, ergative and antipassive (SUBSECTION 3.2.4).

Using examples from a diverse range of languages, I demonstrated that voice alternations not only affect the mapping of arguments to functions but also have different semantic entailments and different discourse properties. For this reason, I argued that an analysis of voice and alignment should take semantics and discourse factors into account, as well as the traditional morphosyntactic properties discussed in CHAPTER 1. Identifying alignment involves identifying which of a set of possible transitive clauses is the most proto-typically transitive (see SUBSECTION 3.3). This may differ depending on whether clauses are compared at the levels of morphosyntax, semantics, discourse or indeed any other level. The proto-typical characteristics of active/transitive clauses in CHAPTER 3, as well as non-default alternations, are summarised for each level of structure in TABLE 6.1:

Table 6.1 Default and Non-Default Clauses

	Transitive Proto-type	Non-default Alternations
Morphology	morphologically unmarked	morphologically marked
Syntax	two (or more) core arguments	only one core argument
Semantics	higher semantic transitivity	lower semantic transitivity
Discourse	higher frequency	lower frequency
	two topical arguments: actor more topical than undergoer	undergoer more topical than actor or only one topical argument

This framework allows us to explore alignment in more detail in Western Austronesian, and allows for the possibility that languages may have ergative or accusative tendencies as well as morphosyntactically symmetrical voice alternations.

When Austronesian languages are compared in this manner, as discussed in SUBSECTION 6.2.1, some interesting aspects of variation arise. A number of languages in the Philippines appear to have ergative alignment, as proposed by Aldridge (2004,

2011, 2012), in terms of semantic and discourse asymmetries between AV and UV. Similarly, languages like Indonesian, can be argued to have accusative alignment in terms of semantic and particularly discourse asymmetries between the two voices. What is particularly striking, is that Kelabit, and other transitional languages discussed in this thesis, appear to constitute various intermediate stages between these two extremes. In these languages, AV and UV both have a mixture of proto-typically transitive and non-default semantic and discourse properties. This seems to support the proposal of Aldridge (2011) that Western Austronesian languages are undergoing a shift in alignment from ergative to accusative, via a reanalysis of AV as active rather than antipassive (see SUBSECTION 1.4.2.3). However, the shift occurs at the levels of semantics and discourse, rather than in the morphosyntax, since Western Austronesian languages are morphosyntactically ‘symmetrical’, even if the voices sometimes differ in the degree of ‘coreness’ of their arguments, as discussed in CHAPTER 3.

This has two important implications for Western Austronesian and general linguistic theory. Firstly, it suggests that alignment can be symmetrical in some ways and asymmetrical in others. Secondly, it suggests that largescale structural changes can occur at discourse or semantic levels without necessarily producing changes in the morphosyntax. Both of these facts, would need to be adequately accounted for in any theoretical model of Western Austronesian voice.

6.3.3 Summary

In this section, I summarised the findings of this thesis in relation to two wider syntactic debates within Western Austronesian. Firstly, I demonstrated that Kelabit, much like other Western Austronesian languages, can be analysed as having a subject grammatical function, so long as reflexivisation is taken as a property governed by the

highest semantic role, i.e. the actor, and not the highest grammatical function, i.e. the subject. This suggests that it is ‘pivot’ functions that are shared by grammatical subjects cross-linguistically and further motivates distinguishing between grammatical functions and semantic roles in syntactic analyses.

Secondly, I demonstrated that Kelabit, much like other Western Austronesian languages, can be analysed as morphosyntactically symmetrical. This implies that it is neither a proto-typically ergative language, nor a proto-typically accusative language. Nonetheless, I argued that alignment could also be analysed using semantic and discourse tests and that Kelabit has mixed alignment properties at these levels. UV is semantically high in transitivity, but has the non-default property of lower discourse frequency. In contrast, AV is high frequency, but tends to have properties of low semantic transitivity. This is taken to represent an intermediate stage in a process of alignment shift. Specifically, it marks a point at which AV has been reanalysed from a discourse/semantic antipassive to a discourse/semantic active clause, but UV has not been reanalysed as a discourse/semantic passive. Thus, Kelabit supports the analysis of grammatical functions in Manning (1996), the analysis of symmetrical alternations in Kroeger (1993), Foley (2008) and Riesberg (2014) and also the analysis of alignment shift in Aldridge (2012).

6.4 Future Research

The findings summarised in SUBSECTION 6.2 and 6.3 suggest a number of avenues for future research.²⁵² Firstly, if Western Austronesian languages are not all neatly classifiable as either Philippine-type or Indonesian-type, the question arises of what exactly the extent of the variation in Western Austronesian languages is. This is

²⁵² There are also a number of questions relating to the grammar of Kelabit, as outlined in CHAPTER 2.

particularly relevant in the context of Sarawak and Borneo more generally, where a great number of languages are endangered and very few languages have substantial documentation and description (see SUBSECTION 2.2.2). The parametric approach to variation employed in this thesis could easily be extended to languages like Lundayeh and Sa'ban to gain a better understanding of the extent of variation within closely-related Western Austronesian subgroups. It could also be employed in a wider typological survey of Western Austronesian languages in order to reveal where proto-typical Philippine-type languages start to acquire the transitional features associated with Kelabit, and what the relationship is between Indonesian-type languages and languages like Kelabit. This would also allow us to assess whether there are any implicational relationships between the various phenomena discussed in this thesis, and whether we might find a language with a Philippine-type pronominal system but Indonesian-type voice or whether structural changes in Austronesian are ordered in a particular manner.

Secondly, having established that variation exists in Western Austronesian languages, it would be interesting to understand what determines this variation synchronically. In particular, in Kelabit and many other Western Austronesian languages, it was shown that the choice of voice construction, pronoun form and word order variant is not entirely constrained by syntax. What then motivates the syntactic choices that speakers make? In CHAPTER 5, information structural concepts were briefly introduced in relation to word order choices. In fact, information structure has been argued to play a role in voice and differential case marking as well, both in Western Austronesian languages and other languages around the world (see Valle 2011, Santiago 2015). Consequently, gaining a better understanding of information structure in Western Austronesian languages may well shed light on the motivations

for using different structures that involve particular configurations of verbal morphology, nominal morphology and word order. Moreover, as indicated in the survey of word-order patterns in CHAPTER 5, languages may also differ in the nature of the role that information structure plays. Hence, information structure may also reveal itself to be a parameter of variation in Western Austronesian languages.

Finally, I suggested throughout the thesis that the different structural properties that characterise Philippine-type and Indonesian-type languages may reflect a series of diachronic changes. In particular, I discussed the question of alignment shift from ergative to accusative; the grammaticalization of second-position clitics as affixes and the reanalysis of the clause-initial position as the default position of subjects, rather than a position that is information-structurally marked. A better understanding of variation within Western Austronesian would not only help to present a more accurate typological picture, as discussed above, but also to build a more informed view of how the various proposed structural changes take place and identify additional intermediate stages, in addition to those represented by Kelabit.

Hence, future research could look to extend the methodologies developed in this thesis to a greater number of Western Austronesian languages in order to build up a fine-grained picture of variation within the family. It could look for potential explanations for patterns of variation found within languages, such as the use of multiple voices or flexible patterns of word order. Finally, it could look for potential historical explanations for patterns of variation between different languages, particularly in relation to the functions of different voices, and the correlations in terms of pronouns and word order. This would have far reaching implications not only for the study of Western Austronesian, but for theoretical models of alignment shift and other structural changes, as well as general models of syntax and information structure.

6.5 Conclusion

Arka & Ross (2005: 1) consider Western Austronesian voice alternations to present ‘a significant challenge to historical, descriptive, and typological linguistics, as well as to theoretical syntax’. In particular, the symmetrical nature of the alternations has led to debate surrounding grammatical functions and alignment. In this thesis, I aimed to contribute to ongoing debates by exploring voice and related morphosyntactic phenomena in the Kelabit language of Northern Sarawak. This functioned as a unique opportunity to explore grammatical relations and alignment in a language that is spoken in a transitional area between the more conservative Philippine-type and more innovative Indonesian-type languages. As such, it also functioned as a means of exploring whether the traditional two-way typology was sufficient to capture the full extent of variation in Western Austronesian languages.

Throughout the thesis, I have demonstrated that Kelabit and the languages of Borneo pose a problem for the two-way typology as they show a mixture of Philippine-type and Indonesian-type properties, as well as unique features of their own. This applies in relation to voice alternations but also in relation to clitic and word-order typology. Hence, a better model of syntactic variation is needed to account for Western Austronesian languages. In this thesis, I proposed that a parametric approach to variation not only reveals important structural differences between Western Austronesian languages, but is potentially enlightening in terms of possible historical variation and paths of diachronic change. It is hoped that extending this sort of approach to other languages in Sarawak and beyond may allow us to gain a better understanding of Western Austronesian voice and syntactic typology, and in turn improve our models of historical change and general linguistics.

Bibliography

- Abbi, Anvita. 2001. *A Manual of Linguistic Field Work and Indian Language Structures*. Lincom Handbooks in Linguistics 17. Munich: Lincom Europa.
- Adelaar, Alexander. 1992. *Proto-Malayic: The reconstruction of its phonology and parts of its morphology and lexicon*. Revised version of 1985 PhD Dissertation. Canberra: Pacific Linguistics.
- Adelaar, Alexander. 1995. 'Borneo as a Crossroads for Comparative Austronesian Linguistics'. In Peter Bellwood, James J. Fox & Darrell Tyron (eds.) *The Austronesians: Historical and Comparative Perspectives*, 75-95. Canberra: Dept. of Anthropology, Australian National University.
- Adelaar, Alexander. 2005. 'The Austronesian languages of Asia and Madagascar: a historical perspective'. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.) *The Austronesian Languages of Asia and Madagascar*, 350-376. London: Routledge.
- Adelaar, Alexander. 2009. 'The various origins of the passive prefix *di-*'. In Alexander Adelaar & Andrew Pawley (eds.) *Austronesian historical linguistics and culture history: A festschrift for Bob Blust*, 121–133. Canberra: Pacific Linguistics.

- Adelaar, Alexander. 2013. 'Voice variation in Austronesian languages of Indonesia: Introduction'. In Alexander Adelaar (ed.) *Voice variation in Austronesian languages of Indonesia*, 54: 1-3. Jakarta: NUSA.
- Ahmady, Nur. 2009. 'Subject and preverbal nominal in Sasak Ngeno-Ngene Pancor'. *Linguistika* 16 (30): 1-15.
- Aikhenvald, Alexandra Y. 2007. 'Typological Distinctions in Word Formation'. In Timothy Shopen (ed.) *Language Typology and Syntactic Description, vol III: Grammatical Categories and the Lexicon, 2nd edition*, 1-65. Cambridge: Cambridge University Press.
- Aissen, Judith. 1999. 'Agent Focus and Inverse in Tzotzil'. *Language* 75: 451-485.
- Aissen, Judith. 2003. 'Differential Object Marking: Iconicity vs. Economy'. *Natural Language and Linguistic Theory* 21: 435-483.
- Aldridge, Edith. 2002. 'Wh-movement in Seediq and Tagalog'. *MIT Working Papers in Linguistics* 44: 1-28.
- Aldridge, Edith. 2004. Ergativity and Word Order in Austronesian Languages. PhD Dissertation, Cornell University.
- Aldridge, Edith. 2006. 'The Heterogeneity of VOS and Extraction in Austronesian Languages'. Unpublished Manuscript, Northwestern University.
- Aldridge, Edith. 2008. 'Phase-based Account of Extraction in Indonesian'. *Lingua* 118 (10): 1440-1469.
- Aldridge, Edith. 2010. 'Directionality in Word Order Change in Austronesian Languages'. In A. Breitbarth, C. Lucas, S. Watts & D. Willis (eds.) *Continuity and Change in Grammar*, 169-180. Amsterdam: John Benjamins.

- Aldridge, Edith. 2011. 'Antipassive in Austronesian Alignment Change'. In D. Jonas, J. Whitman & A. Garrett (eds.) *Grammatical Change: Origins, Nature, Outcomes*, 331-345. Oxford: Oxford University Press.
- Aldridge, Edith. 2012. 'Antipassive and Ergativity in Tagalog'. *Lingua* 122: 192-203.
- Aldridge, Edith. 2016. 'Ergativity from subjunctive in Austronesian languages'. *Language and Linguistics* 17 (1): 27-62.
- Amster, Matthew H. 1995. *Kelabit/English, English/Kelabit glossary: A concise guide to the Kelabit language*. Kuching: Rurum Kelabit Sarawak.
- Amster, Matthew H. 1998. Community, Ethnicity, and Modes of Association among the Kelabit of Sarawak, East Malaysia. PhD Dissertation, Brandeis University, Waltham MA.
- Amster, Matthew. 1999. 'Tradition, Ethnicity and Change: Kelabit Practices of Name Changing'. *Sarawak Museum Journal* 54 (75): 183-200.
- Amster, Matthew. 2003. 'Gender Complementarity and Death among the Kelabit'. In William. D. Wilder (ed.) *Journeys of the Soul. Anthropological Studies of Death, Burial and Reburial*, 251-307. Kuching: Borneo Research Council.
- Amster, Matthew. 2006. 'Narrating the Border: Perspectives from the Kelabit Highlands of Borneo'. Alexander Horstmann & Reed L. Wadley (eds.) *Centering the Margin: Agency and Narrative in Southeast Asian Borderlands*, 209-230. New York: Berghahn Books.
- Andersen, Suree & T. David Andersen. 2005. 'Semantic analysis of the Moronene verbal prefix *moN-*'. In I. Wayan Arka & Malcolm D. Ross (eds.) *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, 243-278. Canberra: Pacific Linguistics.
- Anderson, Stephen R. 1992. *A-morphous Morphology*. Cambridge: CUP.

- Antworth, Evan L. 1979. *A grammatical sketch of Botolan Sambal*. Manila: Linguistic Society of the Philippines.
- Appell, George N. 1969. 'Inventory of urgent anthropological research for Borneo: I. *Borneo Research Bulletin* 1 (2): 10-12.
- Arka, I. Wayan. 1998. *From Morphosyntax to Pragmatics in Balinese: A Lexical-Functional Approach*. PhD Dissertation, Department of Linguistics, University of Sydney.
- Arka, I Wayan. 2000. 'Voice and Being Core: Evidence from (Eastern) Indonesian Languages'. Invited paper presented at the AFLA7 Conference, Amsterdam, May 11-13.
- Arka, I. Wayan. 2002. 'Voice systems in the Austronesian languages of Nusantara: Typology, symmetry and Undergoer orientation'. Paper presented at the *10th National Symposium of the Indonesia Linguistics Society*, Bali, Indonesia.
- Arka, I. Wayan. 2003. *Balinese Morphosyntax: A Lexical-Functional Approach*. Canberra: Pacific Linguistics.
- Arka, I. Wayan. 2005. 'The core-oblique distinction and core index in some Austronesian languages of Indonesia'. Keynote Paper presented at the International Association of Linguistic Typology (ALT) VI Conference, Padang, Indonesia. Accessed 2016-02-09
http://users.ox.ac.uk/~cpgl0015/Arka_Core_oblique_distinction_and_core_index_in_some_Austronesian_languages.pdf
- Arka, I. Wayan. 2009. 'On the zero (voice) prefix and bare verbs in Austronesian languages of Nusa Tenggara, Indonesia'. In Bethwyn Evans (ed.) *Discovering*

- history through language: papers in honour of Malcolm Ross*, 247-270.
Canberra, Australia: Pacific Linguistics.
- Arka, I Wayan & Christopher Manning. 1998. *Voice and Grammatical Relations in Indonesian: A New Perspective*. Stanford: CSLI Publications.
- Arka, I. Wayan & Christopher D. Manning. 2008. 'Voice and Grammatical Relations in Indonesian: a New Perspective'. In Peter K. Austin & Simon Musgrave (eds.) *Voice and Grammatical Relations in Austronesian Languages*, 90-127. Stanford: CSLI Publications.
- Arka, I. Wayan & Malcolm D. Ross (eds.). 2005. *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*. Canberra: Pacific Linguistics.
- Arka, I. Wayan & Fay Wouk. 2014. 'Voice-related constructions in the Austronesian languages of Flores'. In I. Wayan Arka & N Indrawati (eds.) *Argument realisations and related constructions in Austronesian languages, Papers from 12-ICAL, Vol. 2*, 313-333. Canberra: Asia-Pacific Linguistics.
- Arnold, Jennifer. 1997. 'The Inverse System in Mapudungan and other Languages'. *Revista de Linguística Teórica y Aplicada* 34.
- Artawa, I Ketut. 1998. *Ergativity and Balinese syntax* (Parts 1, 2 and 3). Jakarta: NUSA.
- Artawa, I Ketut. 2013. 'The Basic Verb Construction in Balinese'. In Alexander Adelaar (ed.) *Voice variation in Austronesian languages of Indonesia*, 5-27. Jakarta: NUSA.
- Asmah Haji Omar. 1983. *The Malay Peoples of Malaysia and Their Languages*. Kuala Lumpur: Art Printing Works.
- Asmah Haji Omar. 1993. *Language and Society in Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

- Asmah Haji Omar. 2004. *The encyclopaedia of Malaysia: languages and literatures*. Singapore: Archipelago Press.
- Austin, Peter K. 2001. 'Word order in a free word order language: the case of Jiwarli'. Jane Simpson, David Nash, Mary Laughren, Peter Austin & Barry Alpher (eds.) *Forty years on: Ken Hale and Australian languages*, 205-323. Canberra: Pacific Linguistics.
- Austin, Peter K. 2004. 'Clitics in Sasak, eastern Indonesia'. Unpublished Ms, Linguistics Association of Great Britain Annual Conference, Sheffield, UK.
- Austin, Peter K. 2011. 'Tense, aspect, mood and evidentiality in Sasak, eastern Indonesia'. Paper presented at *Chronos 10*, Aston University.
- Austin, Peter K. 2013. 'Too many nasal verbs: dialect variation in the voice system of Sasak'. In Alexander Adelaar (ed.) *Voice variation in Austronesian languages of Indonesia*, 29-46. Jakarta: NUSA.
- Austin, Peter K. & Julia Sallabank, eds. 2011. *Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.
- Ayres, Glenn. 1977. ¿Es el Antipasivo siempre una voz? Unpublished Manuscript.
- Ayres, Glenn. 1983. 'The Antipassive Voice in Ixil'. *International Journal of American Linguistics* 49: 20-45.
- Bakker, Dik. 1998. 'Flexibility and consistency in word order patterns in the languages of Europe'. In A. Siewierska (ed.) *Constituent order in the languages of Europe Empirical Approaches to Language Typology*, 384-419. Berlin: Mouton de Gruyter.
- Bala, Poline. 2002. *Changing Borders and Identities in the Kelabit Highlands: anthropological reflections on growing up near an international border*. Koto Samarahan: Unit Penerbitan Universiti Malaysia.

- Barber, E.J.W. 1975. 'Voice: Beyond the Passive'. *Berkeley Linguistics Society* 1: 16-24.
- Barnes, Robert. 1986. 'Review of Carol Rubenstein, *The Honey Tree Song: Poems and Chants of Sarawak Dayaks*'. *JASO* 17 (1): 73.
- Beech, Mervyn W. H. 1982. *Bala Luk Do'. Buri' Lun Bawang luk titen buri' Lun Dayeh te*. Singapore: Richard Clay (SE Asia) Pte., Ltd.
- Behagel, O. 1909/10. 'Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern'. *Indogermanische Forschungen* 25: 110-42.
- Bell, S.J. 1983. 'Advancements and ascensions in Cebuano'. In D. M. Perlmutter (ed.) *Studies in Relational Grammar* 1, 143-218. Chicago: University of Chicago Press.
- Bellwood, Peter. 1985. *Prehistory of the Indo-Malaysian Archipelago*. London: Academic Press.
- Beckman, M. E., & J. Edwards. 1990. 'Lengthenings and shortenings and the nature of prosodic constituency'. In J. Kingston & M. E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and the physics of speech*, 152-178. Cambridge: Cambridge University Press.
- Berendsen, E. 1986. *The Phonology of Cliticization*. Dordrecht: Foris.
- van den Berg, René. 1989. *A Grammar of the Muna Language*. Dordrecht/Providence: Foris.
- van den Berg, René. 1995. 'Focus in Sulawesi: A Preliminary Overview'. Paper presented at the Workshop on Voice in Austronesian, Leiden, 1st September, 1995.

- van den Berg, René. 1996. 'The demise of focus and the spread of conjugated verbs in Sulawesi'. In Hein Steinhauer (ed.) *Papers in Austronesian Linguistics* 3, 89-114. Canberra: Pacific Linguistics.
- Berinstein, Ava. 1985. *Evidence for Multiattachment in K'ekchi Mayan*. New York: Garland.
- Bibi Aminah Abdul Ghani. 2006. 'Maintenance and Shift in a Language of Sarawak: A study of Bintulu'. In Chong Shin, Karim Harun & Yabit Alas (eds.) *Reflections in Borneo Rivers*, 111-126. Pontianak: STAIN Pontianak Press.
- Billings, Loren. 2005. 'Ordering clitics and postverbal R-expressions in Tagalog: a unified analysis?' In Andrew Carnie, Sheila Dooley & Heidi Harley (eds.) *Verb First: on the syntax of verb-initial languages*, 303-339. Amsterdam: John Benjamins.
- Billings, Loren & Daniel Kaufman. 2004. 'Towards a typology of Austronesian pronominal clisis'. *ZAS Papers in Linguistics* 34: 15-29. (Proceedings of AFLA 11).
- Billings, Loren & Abigail Konopasky. 2002. 'The role of morphology in ordering verb-adjacent clitics: From syntax to prosody in Bulgarian and Tagalog'. *Linguistics in Potsdam* 19: 1-26.
- Bird, Steven & Gary Simons. 2003. 'Seven dimensions of portability for language documentation and description'. *Language* 79: 557-582.
- Blake, Barry J. 1987. *Australian Aboriginal Grammar*. London: Croom Helm.
- Blake, Barry J. 1988. 'Tagalog and the Manila - Mt Isa axis'. *La Trobe Working Papers in Linguistics* 1: 77-90.
- Blake, Barry J. 1990. *Relational Grammar*. London: Routledge.

- Blake, F. R. 1925. *A grammar of the Tagalog language*. New Haven: American Oriental Society.
- Blevins, Juliette. 2006. 'Phonetically-based sound patterns: Typological tendencies or phonological universals?' *Proceedings of Lab Phon10*. Berlin: Mouton de Gruyter.
- Bloomfield, Leonard. 1917. *Tagalog texts with grammatical analysis*. Chicago: University of Illinois Studies in Language and Literature.
- Blust, Robert. 1974a. The Proto-North Sarawak vowel deletion hypothesis. PhD Dissertation, University of Hawai'i at Manoa.
- Blust, Robert. 1977. 'Sketches of the morphology and phonology of Bornean languages, 1: Uma Juman (Kayan)'. In S. A. Wurm (ed.) *Papers in Borneo and Western Austronesian Linguistics 2*, 7-122. Canberra: Pacific Linguistics.
- Blust, Robert. 1993. 'Kelabit-English vocabulary'. *Sarawak Museum Journal* 44 (65): 141-226.
- Blust, Robert. 1997. 'Ablaut in Northwest Borneo'. *Diachronica* 14 (1): 1-30.
- Blust, Robert. 2001. 'Language, Dialect and Riotous Sound Change: The case of Sa'ban'. In Graham W. Thurgood (ed.) *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*, 249-359. Tempe: Arizona State University.
- Blust, Robert. 2003. 'Three notes on early Austronesian morphology'. *Oceanic Linguistics* 42 (2): 438-478.
- Blust, Robert. 2004. 'Austronesian Nasal Substitution: A Survey'. *Oceanic Linguistics* 43 (1): 73-148.
- Blust, Robert. 2006. 'The Origin of the Kelabit Voiced Aspirates: A Historical Hypothesis Revisited'. *Oceanic Linguistics* 45 (2): 311-338.

- Blust, Robert. 2013. *The Austronesian Languages*, revised edition. Canberra: Pacific Linguistics.
- Bögel, Tina. 2010. 'Pashto (Endo-)Clitics in a Parallel Architecture'. In Miriam Butt & Tracy Holloway King (eds.) *Proceedings of the LFG10 Conference*, 85–105. Stanford: CSLI Publications.
- Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy Holloway & Maxwell III, John T. 2010. 'Second Position and the Prosody-Syntax Interface'. In Miriam Butt & Tracy Holloway King (eds.) *Proceedings of the LFG10 Conference*, 106–126. Stanford: CSLI Publications.
- Du Bois, John W. 1987. 'The Discourse Basis of Ergativity'. *Language* 63 (4): 805–855.
- Du Bois, John W. 2003. 'Argument Structure: Grammar in Use'. In John W. Du Bois, Lorraine E. Kumpf & William J. Ashby (eds.) *Preferred Argument Structure: Grammar as Architecture for Function*, 11–60. Amsterdam: Benjamins.
- Bolang, Alexander & Tom Harrisson. 1949. 'Murut and related vocabularies with special reference to North Borneo terminology'. *Sarawak Museum Journal* 5: 116–124.
- Booij, G. 1996. 'Cliticization as Prosodic Integration: the case of Dutch'. *The Linguistic Review* 13: 219–242.
- Boutin, Michael E. 1988. 'Problems in Analyzing Focus in the Languages of Sabah'. In Charles Peck (ed.) *Borneo Language Studies I: Sabah Syntax Papers*, 53–80. Dallas: Summer Institute of Linguistics.
- Boutin, Michael E. 1996. 'A tale of two passives in Bonggi'. Paper presented at the *Fourth Borneo Research Council Conference*, Brunei, June 1996.

- Boutin, Michael E. 2010. 'Kemaloh Lundayeh-English dictionary and bibliographic list of materials relating to the Lundayeh-Lun Bawang-Kelabit and related groups of Sarawak, Sabah, Brunei and East Kalimantan. [Borneo Research Council Reference Series 1.] by Ricky Ganang; Jay Crain; Vicki Pearson-Rounds'. *Bijdragen tot de Taal-, Land- en Volkenkunde* 166 (1): 121-124.
- Bowerman, Melissa & Eric Pederson. 1992. 'Topological relations picture series'. In Stephen C. Levinson (ed.) *Space Stimuli Kit 1.2*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Bowern, Claire. 2008. *Linguistic Fieldwork: A Practical Guide*. New York: Palgrave MacMillan.
- Breen, John Gayan. 1981. 'Margany and Gunya'. In R. W. Dixon & Barry J. Blake (eds.) *Handbook of Australian Languages 2*, 274-393. Amsterdam: John Benjamins.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-Functional Syntax*, 2nd edition. Oxford: Wiley-Blackwell.
- Brickell, Tim. 2014. 'Tondano Pronominal Clitics: Philippine-type or Indonesian-type'. Seminar presented to the Centre for Research on Linguistic Diversity (CRLD), La Trobe University.
- Bulan, Lucy & David Lian Labang. 2008. 'Community efforts in developing and financing an MLE project: The Kelabit Language of Sarawak'. Paper presented at *the 2nd International Conference on Language Development, Language Revitalisation, and Multilingual Education in Ethnolinguistic Communities*, Bangkok, Thailand.
- Byarushengo, Ernest, A. Duranti & Larry M. Hyman (eds.). 1977. *Haya Grammatical Structures*. Los Angeles: University of Southern California.

- Campbell, Lyle. 2001. 'What's wrong with grammaticalization?' *Language Sciences* 23: 113-161.
- Cardinaletti, Anna & Michal Starke. 1995. 'Deficient pronouns: A view from Germanic'. *GenGenP* 3: 22-49.
- Cardinaletti, Anna & Michal Starke. 1999. 'The typology of structural deficiency: A case study of three classes of pronouns'. In Henk van Riemsdijk (ed.) *Clitics in the languages of Europe*, 145-233. Berlin: Mouton.
- Ceña, R.M. 1977. 'Patient primary in Tagalog'. Paper presented at the LSA Winter Meeting, Chicago.
- Cense, A. A. & E. Uhlenbeck. 1958. *Critical Survey of Studies in the Languages of Borneo*. 's-Gravenhage: M. Nijhoff.
- Chafe, Wallace L. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Chafe, Wallace. 1987. 'Cognitive Constraints on Information Flow'. In Russell Tomlin (ed.) *Coherence and Grounding in Discourse*, 21-51. Amsterdam: John Benjamins.
- Chang, Anna H. 2006. *A Reference Grammar of Paiwan*. PhD Dissertation, Australian National University, Canberra.
- Chang, Yung-li. 1997. *Voice, Case and Agreement in Seediq and Kavalan*. PhD Dissertation, Hsinchu: National Tsing Hua University.
- Charles, Mathew. 1974. 'Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages'. *Oceanic Linguistics* 13: 457-509.
- Chelliah, Shobhana L. & Willem J. de Reuse. 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht: Springer.

- Cho, T. 2004. 'Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English'. *Journal of Phonetics* 23: 141–176.
- Chung, Sandra. 1976. 'On the Subject of Two Passives in Indonesian'. In Charles Li (ed.) *Subject and Topic*, 57-98. New York: Academic Press.
- Chung, Sandra. 1998. *The design of agreement: Evidence from Chamorro*. Chicago: University of Chicago Press.
- Chung, Sandra. 2003. 'The Syntax and Prosody of Weak Pronouns in Chamorro'. *Linguistic Inquiry* 34: 547-599.
- Chung, Sandra. 2006. 'Properties of VOS languages'. In M. Everaert & H. van Riemsdijk (eds.) *The Blackwell Companion to Syntax*, 685–720. Malden: Blackwell.
- Chung, Sandra. 2008. 'Indonesian Clause Structure from an Austronesian Perspective'. *Lingua* 118 (10): 1554-1582.
- Clayre, Beatrice M. 1972. 'A preliminary comparative study of the Lun Bawang (Murut) and Sa'ban languages of Sarawak'. *Sarawak Museum Journal* 20: 40-41, 45-47.
- Clayre, Beatrice. 1991. 'Focus in Lun Dayeh'. *Sarawak Museum Journal* 63: 413-434.
- Clayre, Beatrice. 1994. 'Sa'ban: a case of language change'. In Peter W. Martin (ed) *Shifting Patterns of Language Use in Borneo*, 209-226. Williamsburg VA: Borneo Research Council.
- Clayre, Beatrice. 2002. 'Observations on Lundayeh auxiliaries and the case of aru''. In Alexander Adelaar & Robert Blust (eds.) *Between Worlds: linguistic Papers in memory of David John Prentice*, 49-64. Canberra: Pacific Linguistics.
- Clayre, Beatrice. 2005. 'Kelabitic languages and the fate of 'focus': evidence from the Kerayan'. In I Wayan Arka & Malcolm Ross (eds.) *The many faces of*

- Austronesian voice systems: some new empirical studies*, 17-57. Canberra: Pacific Linguistics.
- Clayre, Beatrice. 2014. 'A preliminary typology of the languages of Middle Borneo'. In Peter Sercombe, Michael Boutin & Adrian Clynes (eds.) *Advances in research on cultural and linguistic practices in Borneo*, 123-151. Phillips, Maine USA: Borneo Research Council.
- Clemens, Lauren Eby & Maria Polinsky. to appear. 'Verb-initial Word Orders (Primarily in Austronesian and Mayan Languages)'. In M. Everaert & H. van Riemsdijk (eds.) *The Blackwell Companion to Syntax*. 2nd edition. Hoboken: Wiley-Blackwell.
- Clynes, Adrian. 2012. 'Dominant Language Transfer in Minority Language Documentation Projects: Some Examples from Brunei'. *Language Documentation and Conservation* 6: 253-267.
- Cobbinah, Alex & Friederike Lüpke. 2012. 'Not cut to fit - zero-coded passives in African languages'. In Matthias Brenzinger & Anna-Maria Fehn (eds.) *Proceedings of the 6th World Congress in African Linguistics*, 152-165. Cologne: Rüdiger Köppe.
- Cole, Peter & Gabriella Hermon. 1998. 'The Typology of *wh*-Movement: *wh*-Questions in Malay'. *Syntax* 1: 221-258.
- Cole, Peter & Gabriella Hermon. 2005. 'Subject and non-subject Relativization in Indonesian'. *Journal of East Asian Linguistics* 14: 59-88.
- Cole, Peter & Gabriella Hermon. 2008. 'VP Raising in a VOS Language'. *Syntax* 11 (2): 144-197.
- Cole, Peter, Gabriella Hermon & Yanti. 2008. 'Voice in Malay/Indonesian'. *Lingua* 118 (10): 1500-1553.

- Coluzzi, Paolo. 2010. 'Endangered Languages in Borneo: A Survey among the Iban and Murut (Lun Bawang) in Temburong, Brunei'. *Oceanic Linguistics* 49: 119-143.
- Comrie, Bernard. 1980. 'Inverse verb forms in Siberia: Evidence from Chukchee, Koryak, and Kamchadal'. *Folia Linguistica Historica* 1: 61-74.
- Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.
- Comrie, Bernard. 1988. 'Passive and Voice'. In M. Shibatani (ed.) *Passive and Voice*, 9-23. Amsterdam: Benjamins.
- Comrie, Bernard, Madzhid Khalilov & Zaira Khalilova. 2015. 'Valency and Valency Classes in Bezhta'. In Bernard Comrie & Andrej Malchukov (eds.) *Valency Classes in the World's Languages*, 541-570. Berlin: Mouton de Gruyter.
- Cooreman, Ann. 1987. *Transitivity and Discourse Continuity in Chamorro Narratives*. Berlin: Mouton de Gruyter.
- Cooreman, Ann. 1994. 'A functional typology of antipassive'. In Barbara Fox & Paul J. Hopper (eds.) *Voice: Form and Function*, 49-88. Amsterdam and Philadelphia: John Benjamins.
- Cooreman, A., B. Fox & T. Givón. 1984. 'The discourse definition of ergativity'. *Studies in Language* 8: 1-34.
- de Crespigny, Claude Champion. 1896. *Memoirs of Sir Claude Champion de Crespigny*. London: Lawrence and Bullen.
- Croft, William. 1994. 'Voice: Beyond Control and Affectedness'. In Barbara Fox & Paul J. Hopper (eds.) *Voice: Form and Function*, 89-118. Amsterdam: John Benjamins.

- Crouch, Sophie E. 2009. Voice and Verb Morphology in Minangkabau, a Language of West Sumatra, Indonesian. MA Thesis, University of Western Australia.
- Cruttenden, Alan. 1997. *Intonation*. 2nd edition. New York: Cambridge University Press.
- Cullip, Peter F. 2000. 'Language Use and Attitudes of the Remun of Sarawak'. *The Sarawak Museum Journal* LV (76): 1-44.
- Cumming, Susanna. 1991. *Functional change: The case of Malay constituent order*. Berlin & New York: Mouton de Gruyter.
- Cumming, Susanna. 1995. 'Multifunctionality and the realization problem in modelling discourse production.' In Abraham Werner, Talmy Givón & Sandra A. Thompson (eds.) *Discourse Grammar and Typology: Papers in honor of John W. M. Verhaar*, 247-274. Amsterdam: John Benjamins.
- Dahlstrom, Amy. 1991. *Plains Cree Morphosyntax*. New York: Garland.
- Dahlstrom, Amy. nd. 'Inverse Verbs and Other Syntactic Issues'. Unpublished Manuscript, University of Chicago.
- Dalrymple, Mary & Suriel Mofu. 2012. Plural Semantics, reduplication and numeral modification in Indonesian. *Journal of Semantics* 29 (2): 229-260.
- Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.
- David, Maya K. & Caesar Dealwis. 2006. 'Language Shift among the Telugas of Sarawak'. Paper presented at the 8th Biennial Conference of the Borneo Research Council, Kuching, Malaysia.
- Davies, William. 1991. 'Against an Ergative Analysis of Eastern Javanese'. In Yongkyoon No & Mark Libucha (eds.) *Proceedings of the 7th Annual Meeting*

- of the Eastern States Conference on Linguistics (ESCOL '90), 79-89.
Columbus, OH: The Ohio State University.
- Davies, William. 2005. 'The Richness of Madurese Voice'. In I. Wayan Arka & Malcolm Ross (eds.) *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, 193-216. Canberra: Pacific Linguistics.
- Davies, William. 2010. *A Grammar of Madurese*. Berlin: Mouton de Gruyter.
- Dealwis, Caeser. 2008. Language choice among the Dayak Bidayuh Undergraduates.
PhD Dissertation, Faculty of Language and Linguistics, University of Malaya, Kuala Lumpur.
- Diesing, Molly, Filipović-Đurđević, Dušica & Zec, Draga. 2009. 'Clitic Placement in Serbian: Corpus and Experimental Evidence'. In Sam Featherston & Susanne Winkler (eds.) *The Fruits of Empirical Linguistics*, 61–75. Berlin: Mouton de Gruyter.
- Dik, Simon C. 1978. *Functional Grammar*. Amsterdam: North-Holland.
- Dimmendaal, Gerrit J. 2001. 'Places and people: field sites and informants'. In Paul Newman & Martha Ratliff (eds.) *Linguistic fieldwork*, 54-75. Cambridge: Cambridge University Press.
- Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. 2010. *Basic Linguistic Theory*, 2 volumes. Oxford: Oxford University Press.
- Donohue, Mark. 1999. *A Grammar of Tukang Besi*. Berlin: Mouton de Gruyter.
- Donohue, Mark. 2002. 'Voice in Tukang Besi and the Austronesian voice system'. In Fay Wouk & Malcolm Ross (eds.) *The History and Typology of Western Austronesian Voice Systems*, 81-99. Canberra: Pacific Linguistics.

- Donohue, Mark. 2007a. 'Word order in Austronesian from north to south and west to east'. *Linguistic Typology* 11 (2): 349-391.
- Donohue, Mark. 2007b. 'Variation in voice in Indonesian/Malay: Historical and synchronic perspectives'. In Yoshiko Matsumoto, David Oshima, Orrin Robinson & Peter Sells (eds.) *Diversity in Language: Perspectives and Implications*, 71-129. Stanford: CSLI Publications.
- Douglas, R. S. 1911. 'A comparative vocabulary of the Kayan, Kenyan and Kelabit dialects'. *Sarawak Museum Journal* 1 (1): 75-119.
- Dowty, D. R. 1991. 'Thematic proto roles and argument selection'. *Language* 67 (3): 547-619.
- Drossard, Werner. 1984. *Das Tagalog als Repräsentant des Aktivischen Sprachbaus*. Tübingen: Gunter Narr Verlag.
- Dryer, Matthew S. 1983. 'Coos word order'. Paper delivered at the *Western Conference on Language*, Eugene, Oregon.
- Dryer, Matthew S. 1989. 'Large Linguistic Areas and Language Sampling'. *Studies in Language* 13: 257-92.
- Dryer, Matthew. 1990. 'What determines antipassive in Dyirbal?' In Y. No & M. Libucha, (eds.) *Proceedings of the Seventh Eastern States Conference on Linguistics*, 90-101. The Ohio State University.
- Dryer, Matthew S. 1991. 'SVO languages and the OV/VO typology'. *Journal of Linguistics* 27: 443-482.
- Dryer, Matthew S. 1992. 'The Greenbergian Word Order Correlations'. *Language* 68: 81-138.
- Dryer, Matthew S. 1997. 'On the 6-way word order typology'. *Studies in Language* 21: 69-103.

- Dryer, Matthew S. 2007. 'Word Order'. In Timothy Shopen (ed.) *Language Typology and Syntactic Description*, 2nd Edition, 61-131. Cambridge: Cambridge University Press.
- Dryer, Matthew S. 2013a. 'Order of Subject, Object and Verb'. In Matthew S. Dryer & Martin Haspelmath (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
(Available online at <http://wals.info/chapter/81>, Accessed on 2016-02-16.)
- Dryer, Matthew S. 2013b. 'Against the six-way order typology, again'. *Studies in Language* 37: 267-301.
- Dryer, Matthew S. 2013c. 'Order of Object and Verb'. In Matthew S. Dryer & Martin Haspelmath (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
(Available online at <http://wals.info/chapter/83>, Accessed on 2016-02-16.)
- Duncan, Lachlan. 2003. 'The syntactic structure of Tz'utujil Maya'. In Miriam Butt & Tracy Holloway King (eds.) *Proceedings of the LFG03 Conference*, 164-183. Stanford: CSLI Publications.
- Dyen, Isidore. 1965. *A Lexicostatistical Classification of the Austronesian Languages*. Bloomington: Indiana University Publications in Anthropology and Linguistics.
- Edwards, L. & P. Stephens. 1971. *Short Histories of the Lawas and Kanowit Districts*. Sarawak: Borneo Literature Bureau.
- Eghenter, C. & J. Langub. 2008. 'Past meets future: a trans-border forum for a sustainable future for the highlands of Borneo'. *Borneo Research Bulletin* 39: 286-94.

- Eisenbeiss, S. & W.B. McGregor. 1999. 'The circle of dirt'. Ms. Nijmegen: Max-Planck-Institute for Psycholinguistics.
- É-Kiss, Katalin. 1981. 'Structural Relations in Hungarian: a "free" word order language'. *Linguistic Inquiry* 12: 185-213.
- É-Kiss, Katalin. 1995. *Discourse Configurational Languages*. Oxford: Oxford University Press.
- É-Kiss, Katalin. 1998. 'Identificational Focus versus Information Focus'. *Language* 74 (2): 245-273.
- É-Kiss, Katalin. 2002. *The Syntax of Hungarian*. Cambridge: Cambridge University Press.
- Ewing, Michael C. 2005. 'Colloquial Indonesian'. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.) *The Austronesian Languages of Asia and Madagascar*, 227-258. London: Routledge.
- D'Elia, Sam. 2015. 'The Effect of the Perception of Distance on Complement Order Preference in Two Types of Double Complement Constructions: Evidence from a Picture-Sentence Matching Task'. Paper presented at the 2nd *Linguistics Research Student Conference*, SOAS, University of London, 15th May 2015.
- England, Nora. 1991. 'Changes in basic word order in Mayan languages'. *International Journal of American Linguistics* 57, 446-486.
- Erteschik Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford: Oxford University Press.
- Esser, S. J. 1964. *De Uma-taal (West midden-Celebes). Spraakkundige schets en teksten*. s'-Gravenhage: Nijhoff.
- Everett, Daniel. 1996. *Why there are no clitics. An Alternative Perspective on Pronominal Allomorphy*. Dallas: Summer Institute of Linguistics.

- Falk, Yehuda. 2006. *Subjects and Universal Grammar: An Explanatory Theory*.
Cambridge: Cambridge University Press.
- Féry, Caroline & Manfred Krifka. 2008. 'Information structure notional distinctions, ways of expression'. In Piet van Sterkenburg (ed.) *Unity and diversity of languages*, 123-136. Amsterdam: John Benjamins.
- Foley, William A. 2008. 'The Place of Philippine Languages in a Typology of Voice Systems'. In Peter K. Austin & Simon Musgrave (eds.) *Voice and Grammatical Relations in Austronesian Languages*, 22-44. Stanford: CSLI Publications.
- Foley, W. A. & R. D. Van Valin. 1984. *Functional Syntax and Universal Grammar*.
New York: Cambridge University Press.
- Fox, Anthony. 2000. *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. Oxford: Oxford University Press.
- Fry, Brandon J. & Michael David Hamilton. 2014. 'Long-distance Agreement in Algonquian: Accounting for syntactic variation'. Paper presented at the 50th Annual Meeting of the Chicago Linguistics Society, University of Chicago, 11th April 2014.
- Galih, Balang. 1965. Kapah Ayo' Tana' Inih Pangah Penudut Guma Nepeled. How the World was Made by Guma Nepeled: A Kelabit-Murut Story. *The Sarawak Gazette*, May 31, 152.
- Ganang, Ricky, Jay Crain & Vicky Pearson-Rounds. 2008. *Kemaloh Lundayeh - English dictionary; and bibliographic list of materials relating to the Lundayeh - Lun Bawang - Kelabit and related groups of Sarawak, Sabah, Brunei and East Kalimantan*. Williamsburg, VA: Borneo Research Council.

- Garman, M.A., P. D. Griffiths & R. J. Wales. 1970. 'Prepositions and noun particles in children's speech'. *Sarawak Museum Journal* 18: 353-376.
- Gault, Jo Ann M. 1999. *An ergative description of Sama Bangingi*. Manila: Linguistic Society of the Philippines.
- van Gelderen, Elly. 2011. *The Linguistic Cycle: Language Change and the Language Faculty*. Oxford: OUP.
- Gerdts, D. B. 1988. 'Antipassives and causatives in Ilokano: evidence for an ergative analysis'. In Richard McGinn (ed.) *Studies in Austronesian Linguistics*, 295-321. Athens, OH: Ohio University Press.
- Gerlach, Birgit & Janet Grijzenhout. 2001. 'Clitics from different perspectives'. In Birgit Gerlach & Janet Grijzenhout (eds.) *Clitics in Phonology, Morphology & Syntax*, 1-29. Amsterdam: John Benjamins.
- Gil, David. 2005. 'Word order without syntactic categories: How Riau Indonesian does it'. In Andrew Carnie, Sheila Dooley & Heidi Harley (eds.) *Verb First: on the syntax of verb-initial languages*, 243-263. Amsterdam: John Benjamins.
- Gil, David. 2008. 'The Acquisition of Voice Morphology in Jakarta Indonesian'. In Natalia Gagarina & Insa Gülzow (eds.) *The Acquisition of Verbs and their Grammar: The Effect of Particular Languages*, 201-228. Dordrecht: Springer.
- Gil, David. 2014. 'Those Little k forms in Malayic Languages'. Paper presented at the Workshop on Applicatives in 'Indonesian-type' Languages, ISMIL 18, 16th June, Procida, Naples, Italy.
- Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.
- Givón, Talmy. 1981. 'Typology and functional domains'. *Studies in Language* 5 (2): 163-193.

- Givón, Talmy. 1983. *Topic Continuity in Discourse: Quantitative Cross-Language Studies*. Amsterdam: John Benjamins.
- Givón, Talmy. 1994. *Voice and Inversion*. Amsterdam: John Benjamins.
- Golden, M. & M. Milojević Sheppard. 2000. 'Slovene Pronominal Clitics'. In F. Beukema & M. den Dikken (eds.) *The Clitic Phenomena in European Languages*, 191-207. Amsterdam: John Benjamins.
- Goudswaard, Nelleke. 2005. The Begak (Ida'an) language of Sabah. PhD Dissertation, Free University of Amsterdam.
- Greenberg, Joseph H. 1963. 'Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements'. In J. H. Greenberg (ed.) *Universals of Language*, 58-90. Cambridge, MA: MIT Press.
- Greenhill, S. J., Blust, R. & Gray, R. D. 2008. 'The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics'. *Evolutionary Bioinformatics* 4: 271-283.
- Gregor, Tina. 2013. Word Order in Two Malayic Languages: A Corpus Study. MA Thesis, University of Leipzig.
- Grenoble, Lenore. 2007. 'The importance and challenges of documenting pragmatics'. *Language Documentation and Description* 4: 145-162.
- Grice, Martine & Stefan Baumann. 2007. 'An Introduction to Intonation – Functions and Models'. In Trouvain, Jürgen & Ulrike Gut (eds.) *Non-Native Prosody. Phonetic Description and Teaching Practice*, 25-51. Berlin: De Gruyter.
- Grinevald, C. & M. Peake. 2012. 'Ergativity and voice in Mayan languages: a functional-typological approach'. In K. Haude & G. Authier (eds.) *Ergativity, Valency and Voice*, 15-50. Berlin, Mouton de Gruyter.

- Gudai, Darmansyah H. 1985. A Grammar of Maanyan: a Language of Central Kalimantan. PhD Dissertation, Australian National University, Canberra.
- Guilfoyle, E., H. Hung & L. Travis. 1992. 'Spec of IP and Spec of VP: Two subjects in Austronesian languages'. *Natural Language & Linguistic Theory* 10: 375-414.
- Guili, Bilhamin. 1978. 'Ya bauw aw ubal (The turtle and monkey)'. *Studies in Philippine Linguistics* 2 (2): 69-76.
- Gundel, Jeanette K. 1988. 'Universals of topic-comment structure'. In M. Hammond, E. Moravczik & J. Wirth (eds.) *Studies in Syntactic Typology*, 209-239, Amsterdam: John Benjamins.
- De Guzman, V. P. 1988. 'Ergative analysis for Philippine languages: an analysis'. In R. McGinn (ed.) *Studies in Austronesian Linguistics*. Athens, OH: Ohio University Press.
- Hagège, Claude. 1990. 'Do the Classical Morphological Types have Clear-cut Limits?' In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.) *Contemporary Morphology*, 297-308. Berlin: Mouton de Gruyter.
- Hamlaoui, Fatima & Emmanuel-Moselly Makasso. 2013. 'Object left-dislocation, topicalization and the syntax-phonology mapping of intonation phrases in Bàsàa'. Paper presented at *Bantu 5*, Paris.
- Hale, Kenneth. 1983. 'Warlpiri and the Grammar of Non-configurational languages'. *Natural Language and Linguistic Theory* 1: 5-74.
- Harbour D, L. J. Watkins L J, & D. Adger. 2012. 'Information Structure, Discourse Structure and Noun Phrase Position in Kiowa'. *International Journal or American Linguistics* 78 (1): 97-126.

- Harris, CA. & Harris, RW. 2011. 'Information and Communication Technologies for Cultural Transmission among Indigenous Peoples'. *The Electronic Journal of Information Systems in Developing Countries*, vol. 45.
- Harrisson, Tom. 1954. 'Outside Influences on the Culture of the Kelabits of North Central Borneo'. *Sarawak Museum Journal* 6 (4): 104-125.
- Harrisson, Tom. 1959a. *World Within: A Borneo Story*. London: The Cresset Press.
- Harrisson, Tom. 1959b. 'The Kelabits and Muruts'. In Tom Harrisson (ed.) *The Peoples of Sarawak*, 57-72. Kuching: Government Printing Press.
- Harrisson, Tom. 1960. 'A Kelabit Diary'. *Sarawak Gazette* Vol. LXXXVI.
- Harrisson, Tom. 1961. 'Douglas' 1911 Vocabulary in Practise'. *Sarawak Museum Journal* 17-18 (new): 125-126.
- Haspelmath, Martin. 1993. 'More on the typology of inchoative/causative verb alternations'. In Bernard Comrie & Maria Polinsky (eds.) *Causatives and Transitivity*, 87-120. Amsterdam/Philadelphia: John Benjamins.
- Haspelmath, Martin. 2004. 'On directionality in language change with particular reference to grammaticalization'. In Olga Fischer, Muriel Norde & Harry Perridon (eds.) *Up and down the cline: The nature of grammaticalization*, 17-44. Amsterdam: Benjamins.
- Hayes, Bruce. 1989. 'The prosodic hierarchy in meter'. In Paul Kiparsky & Gilbert Youmans (eds.) *Rhythm and Meter*, 201-260. Orlando: Academic Press.
- Healey, A. 1958. 'Notes on Yogad'. *Oceania Linguistic Monographs* 3: 77-82.
- Heine, Bernd & Tania Kuteva. 2007. *The Genesis of Grammar: A Reconstruction*. Oxford: Oxford University Press.
- Hemmings, Charlotte. 2012. Causatives and Applicatives: The case for Polysemy in Javanese. Unpublished MA Thesis, SOAS, University of London.

- Hemmings, Charlotte. 2013. 'Causatives and Applicatives: The case for Polysemy in Javanese'. *SOAS Working Papers in Linguistics 16*: 167-194.
- Hemmings, Charlotte. 2015. 'Kelabit Voice: Philippine-type, Indonesian-type or something a bit different?' *Transactions of the Philological Society* 113 (3): 383-405.
- Himmelman, Nikolaus P. 1991. *The Philippine Challenge to Universal Grammar*. [University of Köln Working Papers 15]. Cologne: Institute for Linguistics.
- Himmelman, Nikolaus P. 1996. 'Person marking and grammatical relations in Sulawesi'. In Hein Steinhauer (ed.) *Papers in Austronesian Linguistics, Vol. 3*, 115-136. Canberra: Pacific Linguistics.
- Himmelman, Nikolaus P. 1998. 'Documentary and Descriptive Linguistics'. *Linguistics* 36: 161-195.
- Himmelman, Nikolaus P. 1999. 'The lack of zero anaphora and incipient person marking in Tagalog'. *Oceanic Linguistics* 38: 231-269.
- Himmelman, Nikolaus P. 2002. 'Voice in western Austronesian: an update'. In Fay Wouk & Malcolm D. Ross (eds.) *The History and Typology of Western Austronesian Voice Systems*, 7-16. Canberra: Pacific Linguistics.
- Himmelman, Nikolaus P. 2005a. 'The Austronesian Languages of Asia and Madagascar: Typological Characteristics'. In Alexander Adelaar & Nikolaus P. Himmelman (eds.) *The Austronesian Languages of Asia and Madagascar*, 110-181. London: Routledge.
- Himmelman, Nikolaus P. 2005b. 'Tagalog'. In A. Adelaar & N. P. Himmelman (eds.) *The Austronesian Languages of Asia and Madagascar*, 350-376. London: Routledge.

- Himmelman, Nikolaus P. 2006a. 'Language Documentation: what is it and what is it good for?' In Jost Gippert, Nikolaus P. Himmelman & Ulrike Mosel (eds.) *Essentials of Language Documentation*, 1-30. Berlin: Mouton de Gruyter.
- Himmelman, Nikolaus P. 2006b. 'Prosody in language documentation'. In Jost Gippert, Nikolaus P. Himmelman & Ulrike Mosel (eds.) *Essentials of language Documentation*, 163-181. Berlin: Mouton de Gruyter.
- Himmelman, Nikolaus P. 2008. 'Lexical Categories and Voice in Tagalog'. In Peter K. Austin & Simon Musgrave (eds.) *Voice and Grammatical Relations in Austronesian Languages*, 247-93. Stanford: CSLI.
- Himmelman, Nikolaus P. & Robert D. Ladd. 2008. 'Prosodic Description: An Introduction for Fieldworkers'. *Language Documentation & Conservation* 2 (2): 244–274.
- Hirst, Daniel & Albert Di Cristo. 1998. *Intonation systems: a survey of twenty languages*. Cambridge: Cambridge University Press.
- Holmer, Arthur. 2005. 'Seediq: Antisymmetry and final particles in a Formosan VOS language'. In Andrew Carnie, Heidi Harley & Sheila Ann Dooley (eds.) *Verb first: On the syntax of verb-initial languages*, 175–201. Amsterdam: Benjamins.
- Holmer, Arthur & Loren Billings. 2014. 'Clitic pronouns in Seediq'. In I. Wayan Arka & N Indrawati (eds.) *Argument realisations and related constructions in Austronesian languages, Papers from 12-ICAL, Vol. 2*, 111-139. Canberra: Asia-Pacific Linguistics.
- Hopper, Paul J. 1979. 'Some observations on the typology of focus and aspect in narrative language'. *Studies in Language* 3(1). 37–64.

- Hopper, Paul J. 1983. 'Ergative, passive, and active in Malay narrative'. In F. Klein-Andreu (ed.) *Discourse perspectives on Syntax*, 67–88. New York: Academic Press.
- Hopper, Paul J. & Sandra A. Thompspon. 1980. 'Transitivity in Grammar and Discourse'. *Language* 56: 251-299.
- Hose, C. & W. McDougall. 1912. *The Pagan Tribes of Borneo*. London: Macmillan.
- Huang, M. Lillian, Elizabeth Zeitoun, Marie M. Yeh, Anna H. Chang & Joy J. Wu. 1999. 'Interrogative constructions in some Formosan languages'. In Yun-mei Yin, Yi-Li Yang & Huei-Chen Chang (eds.) *Chinese languages and linguistics V: Interactions in language*, 639–680. Taipei: Academia Sinica.
- Hudson, Alfred B. 1978. 'Linguistic relations among Bornean peoples with special reference to Sarawak: An interim report'. *Studies in Third World Societies* 3: 1-44.
- Hudson, Alfred B. 1994. 'Linguistic relations among Bornean peoples with special reference to Sarawak: An interim report'. Reworked in Peter W. Martin (ed.) *Shifting Patterns of Language Use in Borneo*, 1-40. Williamsburg VA: Borneo Research Council.
- Hurlbut, Hope M. 1988. *Verb Morphology in Eastern Kadazan*. Canberra: Pacific Linguistics.
- Hymes, Dell. 1971. *Foundations in sociolinguistics. The ethnography of communication*. Philadelphia: The University of Pennsylvania Press.
- Iivonen, Antti. 1998. 'Intonation in Finnish'. In Daniel Hirst & Albert Di Cristo (eds.) *Intonation systems: a survey of twenty languages*, 314-330. Cambridge: Cambridge University Press.

- Janowski, Monica. 1988. 'The motivating forces behind changes in the wet rice agricultural system in the Kelabit Highlands'. *Sarawak Gazette* 114 (1504): 9-21.
- Janowski, Monica. 1991. *Rice, Work and Community among the Kelabit of Sarawak, East Malaysia*. PhD Dissertation, London School of Economics, University of London.
- Janowski, Monica. 2003. *The forest, source of life: The Kelabit of Sarawak*. Kuching, Sarawak: Sarawak Museum.
- Janowski, Monica. 2012. 'Imagining the force(s) of life and the cosmos in the Kelabit Highlands'. In Monica Janowski & Tim Ingold (eds.) *Imagining Landscapes: Past Present and Future*, 143-163. Ashgate. (Anthropological Studies of Creativity and Perception).
- Johnson, D.E. 1977. 'On Relational Constraints in Grammars'. In Cole, P. & Sadock J (eds.) *Syntax and Semantics 8: Grammatical Relations*, 151-178. New York: Academic Press.
- Johnston, Raymond L. 1989. *Grammar and Basic Vocabulary in Oceanic Austronesian Languages: A Standard Elicitation Schedule*. Ukarumpa: Summer Institute of Linguistics.
- Jones, William. 1911. 'Algonquian (Fox)'. In Franz Boas (ed.) *Handbook of North American Languages*, 735-834. Bureau of American Ethnology Bulletin 40. Washington: G.P.O.
- Kaisse, E. 1985. *Connected Speech. The Interaction of Syntax and Phonology*. New York: Academic Press.
- Kalmar, I. 1979. 'The Antipassive and Grammatical Relations in Eskimo'. In F. Plank (ed.) *Ergativity*, 117-143. New York: Academic Press.

- Karlsson, Anastasia & Arthur Holmer. 2011. 'Interaction between word order, information structure and intonation in Puyuma'. In Mitsuaki Endo (ed.) *Workshop on Austroasiatic and Austronesian Linguistics*, 28-38. Tokyo: Linguistic Circle for the Study of Eastern Eurasian Languages.
- Karlsson, Anastasia, Arthur Holmer, Jan-Olof Svantesson & David House. 2015. 'Correlation between word order and intonation in three verb-initial Formosan languages'. Paper presented at the *Workshop on Information Structure in Endangered Languages*, SOAS, University of London.
- Kaswanti Purwo, Bambang. 1988. 'Voice in Indonesian: A discourse study'. In Masayoshi Shibatani (ed.) *Passive and voice*, 195-241. Amsterdam: John Benjamins.
- Katagiri, Masumi. 2005. 'Voice, ergativity and transitivity in Tagalog and other Philippine languages: a typological perspective'. In I Wayan Arka & Malcolm Ross (eds.) *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, 153-174. Canberra: Pacific Linguistics.
- Kaufman, Daniel. 2005. 'Aspects of Pragmatic Focus in Tagalog'. In I Wayan Arka & Malcolm Ross (eds.) *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, 175-196. Canberra: Pacific Linguistics.
- Kaufman, Daniel. 2009. 'Austronesian Nominalism and its Consequences: A Tagalog case study'. *Theoretical Linguistics* 35 (1): 1-49.
- Kaufman, Daniel. to appear. 'Lexical Category and Alignment in Austronesian'. In Lisa Travis, Jessica Coon & Diane Massam (eds.) *Handbook of Ergativity*. Oxford: OUP.
- Keenan, Edward L. 1976. 'Towards a Universal Definition of 'Subject''. In Charles N. Li (ed.) *Subject and Topic*, 303-333. New York: Academic Press.

- Keenan, Edward L. 1978. 'The syntax of subject-final languages'. In Winifred P. Lehmann (ed.) *Syntactic typology: studies in the phenomenology of language*, 267–328. Austin: University of Texas Press
- Keenan, Edward L. 1985. 'Passive in the world's languages'. In Timothy Shopen (ed.) *Language Typology and Syntactic Description, Vol. 1*, 243–281. Cambridge: Cambridge University Press.
- Keenan, Edward L. & Bernard Comrie. 1979. 'Data on the Noun Phrase Accessibility Hierarchy'. *Language* 55 (2): 333-351.
- Keenan, Edward L. & Matthew S. Dryer. 2006. 'Passive in the world's languages'. In Timothy Shopen (ed.) *Clause Structure, Language Typology and Syntactic Description, Vol. 1*, 325–361. Second Edition. Cambridge: Cambridge University Press.
- Kemmer, Suzanne. 1994. 'Middle Voice, Transitivity, and the Elaboration of Events'. In Barbara Fox & Paul J. Hopper (eds.), *Voice: Form and Function*, 179-230. Amsterdam: John Benjamins.
- Kern, Hendrik. 1889. 'Taalkundige gegevens ter bepaling van het stamland der Maleisch-Polynesische volken'. In *Verlagen en Wededeelingen der Koninklijke Akademie van Wetenschappen afdeeling letterkunde, Vol. 3, No. 6*, 270-287. Amsterdam: Johannes Müller.
- Klaiman, Miriam H. 1991. *Grammatical Voice*. Cambridge: Cambridge University Press.
- Klavans, Judith L. 1983. 'The morphology of cliticization'. *CLS* 15: 103-21.
- Klavans, Judith L. 1985. 'The Independence of syntax and phonology in cliticization'. *Language* 61: 95-120.

- Kikusawa, Ritsuko. 2008. 'Historical Changes in Pronoun Positions in Extra-Formosan Languages'. *Studies in Philippine Languages and Cultures* 17: 66-87.
- Kimenyi, Alexandre. 1980. *A relational grammar of Kinyarwanda*. Berkeley, CA: University of California Press.
- King, Victor, T. 1993. *The Peoples of Borneo*. Oxford: Blackwell.
- Kinyalolo, K. 1991. Syntactic Dependencies and the Spec-head Hypothesis in KiLega. PhD Dissertation, UCLA.
- Kitillä, Seppo. 2011. 'Transitivity Typology'. In Jae. J. Song (ed.) *The Oxford Handbook of Linguistic Typology*, 368-398. Oxford: Oxford University Press.
- Kosta, Peter & Anton Zimmerling. 2013. 'Slavic Clitic Systems in a Typological Perspective'. In L. Schürcks, U. Exteberria, A. Giannakidou & P. Kosta (eds.) *Nominal Constructions in Slavic and beyond*, 441-490. Berlin: Walter de Gruyter.
- Krifka, Manfred. 2007. 'Basic notions of information structure'. In Caroline Fery & Manfred Krifka (eds.) *Interdisciplinary Studies on Information Structure* 6, 13-56. Potsdam: Universitätsverlag.
- Kroeger, Paul R. 1988. 'Verbal focus in Kimaragang'. In Hein Steinhauer (ed.) *Papers in Western Austronesian Linguistics, Vol. 3*, 217-240. Canberra: Pacific Linguistics.
- Kroeger, Paul R. 1990. 'Stative Aspect and Unaccusativity in Kimaragang Dusun'. *Oceanic Linguistics* 29 (2): 110-131.
- Kroeger, Paul R. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford: CSLI Publications.

- Kroeger, Paul R. 1998a. 'Language classification in Sarawak: a status report'.
Sarawak Museum Journal 53 (74): 137-73.
- Kroeger, Paul. 1998b. 'Nouns and Verbs in Tagalog: A Reply to Foley'. Paper presented at Workshop on Voice and Grammatical Functions in Austronesian Languages, LFG98, Brisbane.
- Kroeger, Paul. 1998c. 'Clitics and clause structure in Tagalog'. In Bautista (ed.),
Pagtanaw: Essays on language in honor of Teodoro A. Llamzon, 53–72.
Manila: Linguistic Society of the Philippines.
- Kroeger, Paul R. 2004. *Analysing Syntax*. Cambridge: Cambridge University Press.
- Kroeger, Paul. 2005. 'Kimaragang'. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.) *The Austronesian Languages of Asia and Madagascar*, 397–428. London: Routledge.
- Kroeger, Paul R. 2007. 'McKaughan's analysis of Philippine voice'. In L. Billings & N. Goudswaard (eds.) *Piakandatu ami Dr. Howard P. McKaughan*, 41-46.
Manila: Linguistic Society of the Philippines and SIL Philippines.
- Kroeger, Paul R. 2014. 'Passive agents in Malay: The binding properties and discourse functions of agentive =*nya*'. In Siaw-Fong Chung & Hiroki Nomoto (eds.)
Current Trends in Malay Linguistics, 5–29. Jakarta: NUSA.
- Kulikov, Leonid. 2011. 'Voice typology'. In Jae. J. Song (ed.) *The Oxford Handbook of Linguistic Typology*, 368-398. Oxford: Oxford University Press.
- Kuo, Jonathan Cheng-Chuen. 2015. Argument Alternation and Argument Structure in Symmetrical Voice Languages: A Case Study of Transfer Verbs in Amis, Puyuma and Seediq. PhD Dissertation, University of Hawai'i at Mānoa.
- Kuryłowicz, Jerzy. 1964. *The Inflectional Categories of Indo-European*. Heidelberg: Winter.

- Labang, David Lian. 2012. 'Karuh Tauh – Bahasa Kelabit: Simple Orthography'. In Rurum Kelabit Sarawak Education Unit, 11th Annual Education Excellence Award, Miri, Sarawak, 28th April 2012.
- Labo Pur, Samuel. 1961. *Kamus Lun Dayeh*. Kuching: Borneo Literature Bureau.
- Ladefoged, Peter. 2003. *Phonetic Data Analysis*. Malden, MA: Blackwell.
- Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Langub, Jayl. 1987. 'Ethnic Self-Labeling of the Murut or *Lun Bawang* of Sarawak'. *Sarawak Gazette*. Vol. CXIII, 1502: 30-35.
- Lazdina, Tereza Budina. 1966. *Teach Yourself Latvian*. London: The English Universities Press.
- LeBar, Frank M. 1972. 'Kelabitic Murut'. In Frank M. LeBar (ed.) *Ethnic groups of insular Southeast Asia, vol 1: Indonesia, Andaman Islands, and Madagascar*, 158-163. New Haven: Human Relations Area Files Press.
- Lee, Boon Thong & Tengku Shamsul Bahrain. 1993. 'The Bario Exodus: A Conception of Sarawak Urbanisation'. *Borneo Review* 4 (2): 273-89.
- Lee, Celeste & Loren Billings. 2005. 'Wackernagel and Verb-Adjacent Clisis in Central Philippines'. *UCLA Working Papers in Linguistics* 12: 241-254.
- Lees, Shirley. 1959. 'Lun Daye phonemics'. *Sarawak Museum Journal* 9/13-14: 56-62.
- Lehmann, Winifred P. 1973. 'A Structural Principle of Language and Its Implications'. *Language* 49: 47-66.

- LeSourd, Philip. 1976. 'Verb Agreement in Fox'. In Jorge Hankamer & Judith Aissen (eds.) *Harvard Studies in Syntax and Semantics* 2, 445-528. Cambridge, MA: Harvard University.
- van der Leeuw, Frank. 1997. *Clitics: Prosodic Studies*. The Hague: Holland Academic Graphics.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2016. *Ethnologue: Languages of the World, Nineteenth Edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Li, P. J. & S. Tsuchida. 2006. *Kavalan dictionary*. Taipei: Institute of Linguistics, Academia Sinica.
- Liao, H. 2004. Transitivity and Ergativity in Formosan and Philippine Languages. PhD Dissertation, University of Hawai'i at Manoa.
- Lichtenberk, Frantisek. 1985. 'Multiple Uses of Reciprocal Constructions'. *Australian Journal of Linguistics* 5: 19-41.
- Lichtenberk, Frantisek. 2000. 'Inclusory Pronominals'. *Oceanic Linguistics* 39 (1): 1-32.
- Llamzon, T. 1968. Modern Tagalog. PhD Dissertation, Georgetown University.
- Lobel, Jason. 2013. Philippine and North Bornean Languages: Issues in Description, Subgrouping and Reconstruction. PhD Disstertation, University of Hawai'i at Manoa.
- Lødrup, Helge. 2000. 'Underspecification in Lexical Mapping Theory: The Case of Norwegian Existentials and Resultatives'. In Miriam Butt & Tracy Holloway King (eds.) *Argument Realization*, 55-83. Stanford: CSLI Publications.

- Lowe, John J. 2011. 'Ṛgvedic clitics and prosodic movement'. In Miriam Butt & Tracy Holloway King (eds.) *Proceedings of the LFG11 Conference*, 360-380. Stanford: CSLI Publications.
- Lüpke, Friederike. 2009. 'Data collection methods for field-based language documentation'. *Language Documentation and Description* 6: 53-100.
- Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- Macaulay, Monica. 2009. 'On Prominence Hierarchies: Evidence from Algonquian'. *Linguistic Typology* 13: 357-389.
- Mallinson, Graham & Barry Blake. 1981. *Language Typology: Cross-linguistic Studies in Syntax*. Amsterdam: North Holland.
- Manley, Timothy M. 1972. 'Outline of Sre Structure'. *Oceanic Linguistics Special Publications* 12: 1, 3-100, 102-239.
- Manning, Christopher D. 1996. *Ergativity, Argument Structure and Grammatical Relations*. Stanford: CSLI Publications.
- Marten, Lutz & Jenneke van der Wal. 2014. 'A typology of Bantu subject inversion'. *Linguistic Variation* 14 (2): 318-368.
- Martens, M. 1988a. 'Notes on Uma Verbs'. In Hein Steinhauer (ed.) *Papers in Western Austronesian Linguistics, Vol. 4*, 167-237. Canberra: Pacific Linguistics.
- Martens, M. 1988b. 'Focus and Discourse in Uma'. In Hein Steinhauer (ed.) *Papers in Western Austronesian Linguistics, Vol. 4*, 239-256. Canberra: Pacific Linguistics.

- Martens, M. 1988c. 'Focus or Ergativity? Pronouns Sets in Uma'. In Hein Steinhauer (ed.) *Papers in Western Austronesian Linguistics, Vol. 4*, 263-277. Canberra: Pacific Linguistics.
- Martin, Peter W. 1995. 'Wither the Indigenous Languages of Brunei Darussalam?' *Oceanic Linguistics* 34 (1): 27-43.
- Martin, Peter W. 1996. 'A Comparative Ethnolinguistic Survey of the Murut (Lun Bawang) with Special Reference to Brunei'. In Peter W. Martin, Conrad Ožóg & Gloria Poedjosoedarmo (eds.) *Language Use and Language Change in Brunei Darussalam*, 268-279. Athens, OH: Ohio University Press.
- Martin, Peter W. & Eileen Yen. 1994. 'Language use among the Kelabit living in urban centres'. In Peter W. Martin (ed) *Shifting Patterns of Language Use in Borneo*, 147-163. Williamsburg VA: Borneo Research Council.
- Marušič, F. 2008. 'Slovenian Clitics have no Unique Syntactic Position'. In A. Antonenko, J. Bailyn & C. Bethin (eds.) *Formal Approaches to Slavic Linguistics* 16, 266-281. Ann Arbor: Michigan Slavic Publications.
- Mashman, Valerie. 2014. 'A song about feasting and the loss of beads'. Paper presented at the 12th *International Borneo Research Council Conference*, Universiti Malaysia Sabah, Kota Kinabalu, 5-7 August, 2014.
- Matti, David F. 1994. 'Mamasa pronoun sets'. In René van den Berg (ed.) *Studies in Sulawesi Linguistics, Vol. 3*, 65-89. Jakarta: NUSA.
- McKaughan, Howard. 1958. *The inflection and syntax of Maranao verbs*. Manila: The Institute of National Language – Philippines.
- McLellan, James. 2014. 'Strategies for revitalizing endangered Borneo languages: A comparison between Negara Brunei Darussalam and Sarawak, Malaysia'. *Southeast Asia: A Multidisciplinary Journal* 14: 14–22.

- Mead, David E. 1998. Proto-Bungku-Tolaki: Reconstruction of its phonology and aspects of its morphosyntax. PhD Dissertation, Rice University.
- Miller, Mark. 2007. A Grammar of West Coast Bajau. PhD Dissertation, University of Texas at Arlington.
- Miller, Mark. 2014. 'A comparative look at the major voice oppositions in Sama-Bajau languages and Indonesian/Malay'. In I. Wayan Arka & N Indrawati (eds.) *Argument realisations and related constructions in Austronesian languages, Papers from 12-ICAL, Vol. 2*, 303-312. Canberra: Asia-Pacific Linguistics.
- Mistica, Meladel, Avery Andrews, I. Wayan Arka & Timothy Baldwin. 2009. 'Double-Double, Morphology and Trouble: Looking into Reduplication in Indonesian'. In Luiz Augusto Pizzato & Rolf Schwitter (eds.) *Proceedings of the 2009 Australian Language Technology Association Workshop*, 44-52. Sydney, Australia: Australian Language Technology Association.
- Mithun, Marianne. 1992. 'Is basic word order universal?' In Doris L Payne (ed.) *Pragmatics of Word Order Flexibility*, 15-62. Amsterdam: Benjamins.
- Mithun, Marianne. 1994. 'The implications of ergativity for a Philippine voice system'. In Barbara Fox & Paul J. Hopper (eds.) *Voice: Form and Function*, 247-277. Amsterdam: John Benjamins.
- Mithun, Marianne. 2000. 'Valency-changing derivation in Central Alaskan Yup'ik'. In R. Dixon & A. Aikhenvald (eds.) *Changing Valency: Case Studies in Transitivity*, 84-114. Cambridge: Cambridge University Press.
- Mithun, Marianne. 2001. 'Who shapes the record: the speaker and the linguist'. In Paul Newman & Martha Ratliff (eds.) *Linguistic fieldwork*, 34-54. Cambridge: Cambridge University Press.

- Mkude, Daniel J. 1974. A study of Kiluguru syntax with special reference to the transformational history of sentences with permuted subject and object. PhD Dissertation, SOAS, University of London.
- Morimoto, Yukiko. 2000. Discourse Configurationality in Bantu Morphosyntax. PhD Dissertation, Stanford University, Stanford CA.
- Morimoto, Yukiko. 2003. 'Markedness hierarchies and optimality in Bantu'. In R. Laury, G. McMenamin, S. Okamoto, V. Samiian & K. Subbarao (eds.) *Perspectives in Linguistics: Papers in Honor of P. J. Mistry*, 261–280. New Delhi: Indian Institute of Language Studies.
- Morimoto, Yukiko. 2006. 'Agreement properties and word order in comparative Bantu'. In Laura Downing, Sabine Zerbian & Lutz Marten (eds.) *Papers in Bantu grammar and description*, 161-187. Berlin: Zentrum für Allgemeine Sprachwissenschaft.
- Mosel, Ulrike. 2006. 'Sketch grammar'. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.) *Essentials of Language Documentation*, 301-309. Berlin: Mouton de Gruyter.
- Musgrave, Simon. 2002. Non-subject arguments in Indonesian. PhD Dissertation, Melbourne University.
- Mycock, Louise & John Lowe. 2014. 'Representing Information Structure'. Paper presented at the *14th South of England LFG Meeting*, London 10th May 2014.
- Næss, Åshild. 2014. 'Symmetrical voice in Äiwoo and its implications for the history of Oceanic'. Paper presented at the *7th Austronesian and Papuan Languages and Linguistics Conference*, SOAS, University of London.
- Nagaya, Naonori. 2009a. 'The middle voice in Tagalog'. *Journal of the Southeast Asian Linguistics Society* 1: 159-188.

- Nagaya, Naonori. 2009b. 'Subject and topic in Lamaholot, Eastern Flores'. Presented at the 11th *International Conference on Austronesian Linguistics*, Aussois, France, June 22-26, 2009.
- Nagaya, Naonori. 2010. 'Voice and grammatical relations in Lamaholot of eastern Indonesia'. In A. Shiohara (ed.) *Proceedings of the Workshop on Indonesian-type Voice System*, 3-23. Tokyo: ILCAA, Tokyo University of Foreign Studies.
- Ndayiragije, Juvénal. 1999. 'Checking economy'. *Linguistic Inquiry* 30: 399-444.
- Needham, Stephanie & Ida Toivonen. 2011. 'Derived Arguments'. In Miriam Butt and Tracey Holloway King (eds.) *Proceedings of the LFG11 Conference*, 401-421. Stanford: CSLI Publications.
- Nespor, Marina. 1993. 'The phonology of clitic groups'. In L. Hellan & H. van Riemsdijk (eds.) *Clitic Doubling and Clitic Groups* [Eurotyp Working Papers 5], 179-192. Dordrecht: Foris.
- Nespor, Marina & Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Nevis, Joel, Brian D. Joseph, Dieter Wanner & Arnold M. Zwicky. 1994. *Clitics: a comprehensive bibliography, 1982-1991*. Amsterdam: John Benjamins.
- Newmeyer, Frederick J. 2004. 'Basic word order in formal and functional linguistics and the typological status of 'canonical' sentence types'. In Dominique Willems, Bart Defrancq, Timothy Coleman & Dirk Noël (eds.) *Contrastive analysis in language: Identifying linguistic units of comparison*, 69-88. Basingstoke: Palgrave Macmillan.
- Newmeyer, Frederick J. 2005. *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press.
- Nichols, Johanna 1979. 'Syntax and pragmatics in Manchu-Tungus languages'. In Paul R. Clyne, William F. Hanks & Carol L. Hofbauer (eds.) *The elements: a*

- parasession on linguistic units and levels*, 420-28. Chicago: Chicago Linguistic Society.
- Nolasco, R. M. 2005. 'What ergativity in Philippine languages really means?' Paper presented at Taiwan-Japan Joint Workshop on Austronesian Languages, National Taiwan University, Taipei, Taiwan, June 23-24.
- Nomoto, Hiroki & Kartini Abd. Wahab. 2014. 'Person Restriction on Passive Agents in Malay: Information Structure and Syntax'. In Siaw-Fong Chung & Hiroki Nomoto (eds.) *Current Trends in Malay Linguistics*, 31–50. Jakarta: NUSA.
- Norde, Muriel. 2006. 'Demarcating degrammaticalization: The Swedish s-genitive revisited'. *Nordic Journal of Linguistics* 29: 201–328.
- Norman, William M. 1978. 'Advancement Rules and Syntactic Change: The loss of Instrumental Voice in Mayan'. In Jeri J. Jaeger et al (eds.) *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistic Society*, 458-76. Berkeley, CA: Berkeley Linguistics Society.
- Norwood, Clodagh. 2002. 'Voice and Valency Alternations in Karo Batak'. In Fay Wouk & Malcolm Ross (eds.) *The History and Typology of Western Austronesian Voice Systems*, 181-207. Canberra: Pacific Linguistics.
- Ochs, Elinor. 1979. 'Planned and unplanned discourse'. In Talmy Givón (ed.) *Discourse and Syntax*, 51-80. San Diego: Academic Press.
- Paauw, S. H. 2008. *The Malay Contact Varieties of Eastern Indonesia: A Typological Comparison*. PhD Dissertation, State University of New York, New York.
- Padan, James Noah. 1971. *Lun Dayeh (Murut) phrase book*. Kuching: Borneo Literature Bureau.
- Pallesen, A. Kemp. 1985. *Culture contact and language convergence*. Manila: Linguistic Society of the Philippines.

- Pastika, Wayan. 1999. Voice Selection in Balinese Narrative Discourse. PhD Dissertation, Australian National University, Canberra.
- Payne, Doris. 1990. *The pragmatics of word order: typological dimensions of verb initial languages*. Berlin: Mouton de Gruyter.
- Payne, Doris. 1994. 'The Tupí-Guaraní inverse'. In Barbara Fox and Paul J. Hopper (eds.) *Voice: Form and Function*, 313-340. Amsterdam: John Benjamins.
- Payne, Thomas. 1982. 'Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog'. *Studies in Language* 6 (1): 75-106.
- Payne, Thomas. 1983. 'Yagua Subject Clitics: Syntactic and Phonological Misalignment and Another Possible Source of Ergativity'. In J. Richardson et al (eds.) *Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax*, 173-184. Chicago: Chicago Linguistic Society.
- Pearson, Matthew. 2001. The Clause Structure of Malagasy: A Minimalist Approach. Ph. D. thesis, University of California, Los Angeles.
- Pearson, Matthew. 2005. 'The Malagasy subject/topic as an A'-element'. *Natural Language and Linguistic Theory* 23: 381-457.
- Peperkamp, S. 1996. 'On the prosodic representation of clitics'. In U. Kleinherz (ed.) *Interfaces in Phonology, Studia Grammatica* 41, 102-127. Berlin: Akademie Verlag.
- Perlmutter, D. 1978. 'Impersonal Passives and the Unaccusative Hypothesis'. In Jeri J. Jaeger et al (eds.) *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society*, 157-189. Berkeley: Berkeley Linguistics Society.
- Perlmutter, D. & R. Rhodes. 1988. 'Thematic-syntactic alignments in Ojibwa: evidence for subject-object reversal'. Paper presented at the LSA.

- Pierrehumbert, Janet. 1980. *The Phonology and Phonetics of English Intonation*. PhD Dissertation, MIT, Cambridge, MA.
- Pinkerton, Sandra. 1978. 'Word Order and the Antipassive in K'ekchi'. In Nora England (ed.) *Papers in Mayan Linguistics*, 157-168. Columbia: University of Missouri Press.
- Polinsky, Maria. 2012. 'Headedness, again'. *UCLA Working Papers in Linguistics* 17: 348-359.
- Polinsky, Maria. 2013. 'Antipassive Constructions'. In Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
(Available online at <http://wals.info/chapter/108>, Accessed on 2016-02-09)
- Polinsky, Maria. to appear. 'Antipassive'. In Jessica Coon, Diane Massam & Lisa Travis (eds.) *The Oxford Handbook of Ergativity*. Oxford: Oxford University Press.
- Polinsky, Maria & Eric Potsdam. to appear. 'Austronesian Syntax'. In Bill Palmer (ed.) *Oceania*. Berlin: Mouton de Gruyter.
- Pollard, F. H. 1933. 'The Muruts of Sarawak'. *Sarawak Museum Journal* 4: 139-155.
- Potsdam, Eric. 2009. 'Austronesian verb-initial languages and wh-question strategies'. *Natural Language and Linguistic Theory* 27: 737-771.
- Potsdam, Eric & Maria Polinsky. 2012. 'Questions and Word Order in Polynesian'. In Claire Moyse-Faurie & Joachim Sabel (eds.) *Topics in Oceanic Morphosyntax*, 83-109. Berlin: Mouton de Gruyter.
- Prentice, D. J. 1970. 'The linguistic situation in northern Borneo'. In Stephen A. Wurm & D. C. Laycock (eds.) *Pacific Linguistics Studies in Honor of Arthur Capell*, 369-408. Canberra: Pacific Linguistics.

- Prentice, D. J. 1971. *The Murut Languages of Sabah*. Canberra: Pacific Linguistics.
- Prentice, D. J. 1995. 'Timugon Murut'. In D. T. Tyron (ed.) *Comparative Austronesian Dictionary*, 381-391. Berlin: Mouton de Gruyter.
- Quakenbush, J. Stephen. 1992. 'Word order and discourse type: An Austronesian example'. In Doris L Payne (ed.) *Pragmatics of Word Order Flexibility*, 279-303. Amsterdam: Benjamins.
- Quick, Phil. 2005. 'Topic Continuity, Voice and Word Order in Pendau'. In I. Wayan Arka & Malcolm D. Ross (eds.) *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, 221-242. Canberra: Pacific Linguistics.
- Rackowski, Andrea. 2002. The Structure of Tagalog: Specificity, voice and the distribution of arguments. Doctoral Dissertation, MIT, Cambridge, MA.
- Rackowski, Andrea & Norvin Richards. 2005. 'Phase edge extraction: A Tagalog case study'. *Linguistic Inquiry* 36: 565-599.
- Rafferty, E. 1982. *Discourse Structures of the Chinese Indonesian of Malang*. Jakarta: NUSA.
- Rahmajanti, Ruli. 1995. 'Sistem kebahasaan suku Kenyah Leppo' Ke di Apau Peng Kecamatan Long Pujungan. Culture and Conservation in East Kalimantan'. WWF Kayan-Mentarang Conservation Project, Jakarta.
- Rasolofo, A. 2007. 'Word order and transitive clauses in Malagasy'. In Iwasaki Shoichi, Andrew Simpson, Karen Adams & Paul Sidwell (eds.) *SEALS XIII Papers from the 13th Annual Meeting of the Southeast Asian Linguistics Society 2003*, 213-222. Canberra: Pacific Linguistics.
- Ray, Sidney H. 1913. 'The Languages of Borneo'. *Sarawak Museum Journal* 1 (4): 1-196.

- Reid, Lawrence A. 1966. *An Ivatan Syntax*. Honolulu: Pacific and Asian Linguistics Institute, University of Hawai'i at Manoa.
- Reid, Lawrence A. 1970. *Central Bontoc: Sentence, Paragraph and Discourse*. University of Oklahoma: Summer Institute of Linguistics.
- Reid, Lawrence A. & Hsiu-chuan Liao. 2004. 'A Brief Syntactic Typology of Philippine Languages'. *Language and Linguistics* 5 (2): 433-490.
- Rethinasamy, Soubakeavathi. 2014. 'Language use among the Kelabit in Bario, Sarawak'. Unpublished ms, UNIMAS, Sarawak, Malaysia.
- Rethinasamy, Soubakeavathi, Chuah Kee Man, Ameer Joan & Salbia Hassan. 2013a. 'Kelabit Language Use and Attitude: A Study in the Kelabit Highlands of Bario'. Poster presented at the UNIMAS R&D Expo, Kota Samarahan, Sarawak.
- Rethinasamy, Souba, Norazila Abd Aziz, Fitri Suraya Mohamad, Mohd Hafizan Hashim, & Dayang Sariah Abang Suhai. 2013b. 'Developing literacy and knowledge preservation skills among remote rural children'. In A. Pandian, C. C. L. Liew, D. A. L. Tan, M. Jayagowri, B. C. Lee, & C. H. Toh (eds.) *New literacies: Reconstructing education and language*, 94-107. Newcastle Upon Tyne: Cambridge Scholars Publishing.
- Rhodes, Richard A. 1994. 'Agency, Inversion and Thematic Alignment in Ojibwe'. In Susanna Gahl et al (eds.) *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*, 431-446. Berkeley: Berkeley Linguistics Society.
- Riesberg, Sonja. 2014. *Symmetrical Voice and Linking in Western Austronesian Languages*. Berlin: Walter de Gruyter.
- Ross, Malcolm. 2002. 'The History and Transitivity of Western Austronesian Voice and Voice-marking'. In Fay Wouk & Malcolm Ross (eds.) *The History and*

- Typology of Western Austronesian Voice Systems*, 17-62. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2006. 'Reconstructing the case-marking and personal pronoun systems of Proto Austronesian'. In Henry Yungli Chang, Lillian M. Huang & Dah-an Ho (eds.) *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, 521-564. Taipei: Institute of Linguistics, Academia Sinica.
- Ross, Malcolm. 2009. 'Proto Austronesian Verbal Morphology: A Reappraisal'. In Alexander Adelaar & Andrew Pawley (eds.) *Austronesian Historical Linguistics and Culture History: a Festschrift for Robert Blust*, 295-236. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2015. Ross, Malcolm. 2015. 'Revising the reconstruction of early Austronesian personal pronouns'. In Malcolm D. Ross & I Wayan Arka (eds.) *Language change in Austronesian languages, Papers from 12-ICAL, Vol. 3*, 109–121. Canberra: Asia-Pacific Linguistics.
- Roth, Henry Ling. 1896. *The Natives of Sarawak and British North Borneo*. 2 vols. London: Truslove and Hanson. Rep. Malaysia: University of Malaya Press.
- Round, Erich R. 2012. 'Durational Correlates of Prosodic Structure in Kayardild Vowels'. In Felicity Cox, Susan Lin, Kelly Miles, Sallyanne Palethroe, Jason Shaw & Ivan Yuen (eds.) *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, 221-224. Sydney, Australia: Macquarie University.
- Rubenstein, Carol. 1973. *Poems of the Indigenous Peoples of Sarawak: Some Songs and Chants*. Kuching: Sarawak Museum.

- Rubino, Carl. 2005. 'Iloko'. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.) *The Austronesian Languages of Asia and Madagascar*, 326-349. London: Routledge.
- Ruhlen, Merritt. 1975. 'A Guide to the Languages of the World'. Language Universals Project, Stanford University.
- Rutter, Owen. 1929. *The Pagans of North Borneo*. London: Hutchinson & Co. Ltd.
- Saging, Robert Lian R. 1967/77. An Ethno-history of the Kelabit Tribe of Sarawak. A brief look at the Kelabit tribe before World War II and after. Graduation Exercise submitted in partial fulfilment of the requirements for the Degree of Bachelor of Arts, Hons, Jabatan Sejarah, University of Malaya, Kuala Lumpur.
- Saging, Robert Lian R. & Lucy Bulan. 1989. 'Kelabit Ethnography: A Brief Report'. *Sarawak Museum Journal* 40/61: 89-118.
- Santiago, Paul J. 2015. 'Information Structure and Voice Alternation in Kalanguya'. *Journal of the Southeast Asian Linguistics Society (JSEALS)* 8: 1-23.
- Schachter, Paul. 1976. 'The subject in Philippine languages: Topic, actor, actor-topic, or none of the above'. In Charles N. Li (ed.) *Subject and Topic*. New York: Academic Press.
- Schachter, Paul. 1995. 'The Subject in Tagalog: Still none of the above'. *UCLA Occasional Papers in Linguistics, Vol. 15*. Los Angeles: UCLA Department of Linguistics.
- Schachter, Paul & Otones, Fé T. 1972. *Tagalog Reference Grammar*. Berkeley: University of California Press.
- Schmidt, P. S. V. D. 1926. *Die Sprachfamilien und Sprachenkreise der Erde*. Heidelberg: Winter.

- Schneeberger, W. F. 1945. 'The Kerayan-Kalabit Highland of Central Northeast Borneo'. *Geographical Review* 35: 544-562.
- Schneeberger, W. F. 1979. *Contributions to the Ethnology of Central Northeast Borneo*. Bern: The University of Berne Institute of Ethnology.
- Schultze-Berndt, Eva. 2006. 'Linguistic annotation'. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.) *Essentials of Language Documentation*, 213-251. Berlin: Mouton de Gruyter.
- Sedeng, I N. 2000. *Kalimat Kompleks dan Relasi Gramatikal Bahasa Sikka*. MA Thesis, Universitas Udayana, Denpasar.
- Seifart, Frank. 2008. 'On the representativeness of language documentation'. *Language Documentation and Description* 5: 60-76.
- Selkirk, Elisabeth. 1980. 'Prosodic domains in phonology. Sanskrit revisited'. In M. Aronoff & M. L. Kean (eds.) *Juncture. A collection of original papers*, 107-129. Saratoga: Anma Libri.
- Selkirk, Elisabeth. 1984. *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Selkirk, Elisabeth. 1995. 'The prosodic structure of function words'. In J. Beckman et al (eds.) *Papers in Optimality Theory. University of Massachusetts Occasional Papers* 18, 439-469. Amherst, MA: GSLA, University of Massachusetts.
- Shibatani, Masayoshi. 1985. 'Passive and related constructions: a prototype analysis'. *Language* 61 (4): 821-848.
- Shibatani, Masayoshi. 1988. 'Voice in Philippine Languages'. In Masayoshi Shibatani (ed.) *Passive and Voice*, 85-142. Amsterdam: John Benjamins.

- Shibatani, Masayoshi. 2005. 'The attrition of the Austronesian focus system'. In *Proceedings of the Taiwan-Japan Joint Workshop on Austronesian Languages*, 1-18. Taipei: National Taiwan University.
- Shibatani, Masayoshi. 2006. 'On the conceptual framework for voice phenomena'. *Linguistics* 44 (2): 217-269.
- Shibatani, Masayoshi. 2008. 'Relativization in Sasak and Sumbawa, Eastern Indonesia'. *Language and Linguistics* 9 (4): 865-916.
- Shibatani, Masayoshi. 2009. 'Focus constructions without focus morphology in the AN languages of Nusa Tenggara'. Presented at the *11th International Conference on Austronesian Linguistics*, Aussois, France, June 22-26, 2009.
- Shiohara, Asako. 2012. 'Applicatives in Standard Indonesian'. In Wataru Nakamura & Ritsuko Kikusawa (eds.) *Objectivization and Subjectivization: A Typology of Voice Systems*, 59-76. Osaka: National Museum of Ethnology.
- Siewierska, Anna. 1984. *The passive: a comparative linguistic analysis*. London: Croom Helm.
- Siewierska, Anna. 1988. *Word Order Rules*. London: Croom Helm.
- Siewierska, Anna. 1998. 'Variation in major constituent order; a global and a European perspective'. In Anna Siewierska (ed.) *Constituent Order in the Languages of Europe, Empirical Approaches to language typology*, 475-551. Berlin: Mouton de Gruyter.
- Silitonga, Mangasa. 1973. *Some Rules Reordering Constituents and their Constraints in Batak*. PhD Dissertation, University of Illinois.
- Smith-Stark, Thomas. 1978. 'The Mayan Antipassive: Some Facts and Fictions'. In Nora England (ed.) *Papers in Mayan Linguistics*, 169-187. Columbia: University of Missouri Press.

- Sneddon, James N. 1975. *Tondano Phonology and Grammar*. Canberra: Pacific Linguistics B-38.
- Sneddon, James N. 1996. *Indonesian: A Comprehensive Grammar*. London and New York: Routledge.
- Sneddon, James N. 2006. *Colloquial Jakartan Indonesian*. Canberra: Australian National University.
- Soh, Hooi Ling & Hiroki Nomoto. 2008. 'Aspectual Effects of the Prefix *meN-* in Malay'. Paper presented at *Twelfth International Symposium on Malay/Indonesian Linguistics*, 26th-27th June, Leiden, The Netherlands.
- Soh, Hooi Ling & Hiroki Nomoto. 2009. 'Progressive aspect, the verbal prefix *meN-*, and stative sentences in Malay'. *Oceanic Linguistics* 48 (1): 148-171.
- Soh, Hooi Ling & Hiroki Nomoto. 2015. 'Degree achievements, telicity and the verbal prefix *meN-* in Malay'. *Journal of Linguistics* 51 (1): 147-183.
- Song, Jae Jung. 2011. 'Word Order Typology'. In Jae Jung Song (ed.) *The Oxford Handbook of Linguistic Typology*, 258-279. Oxford: Oxford University Press.
- Song, Jae Jung. 2012. *Word Order*. Cambridge: Cambridge University Press.
- Soriente, Antonia. 2013. 'Undergoer Voice in Borneo: Penan, Punan, Kenyah and Kayan languages'. In Alexander Adelaar (ed.) *Voice variation in Austronesian languages of Indonesia*, 175-203. Jakarta: NUSA.
- Soriente, Antonia. 2014. 'Quotative Verbs and the Behaviour of the Multifunctional Prefix *Ke-* in Punan Tubu' and Kenyah'. Paper presented at the *12th International Borneo Research Council Conference*, Universiti Malaysia Sabah, Kota Kinabalu, 5-7 August, 2014.
- Southwell, C. Hudson. 1949. 'The Structure of the Murut Language'. *Sarawak Museum Journal* 5: 104-115.

- Spencer, Andrew. 1991. *Morphological Theory*. Oxford: Blackwell.
- Spreng, Bettina. 2010. 'On the conditions for antipassives'. *Language and Linguistics Compass* 4 (7): 556–575.
- Starosta, Stanley. 1986. 'Focus as recentralisation'. In P. Geraghty, L. Carrington, & S.A. Wurm (eds.) *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*, 73-95. Canberra: Pacific Linguistics.
- Starosta, Stanley. 2009a [1997]. 'Formosan clause structure: transitivity, ergativity, and case-marking'. In Elizabeth Zeitoun (ed.) *Formosan Linguistics: Stanley Starosta's Contributions. Vol. 2*, 657-680. Taipei: Institute of Linguistics, Academica Sinica.
- Starosta, Stanley. 2009b [1999]. 'Transitivity, ergativity, and the best analysis of Atayal case marking. In Elizabeth Zeitoun (ed.) *Formosan Linguistics: Stanley Starosta's Contributions. Vol. 2*, 681-704. Taipei: Institute of Linguistics, Academica Sinica.
- Starosta, Stanley. 2009c [2004]. 'Transitivity and Ergativity in Austronesian languages'. In Elizabeth Zeitoun (ed.) *Formosan Linguistics: Stanley Starosta's Contributions. Vol. 2*, 705-730. Taipei: Institute of Linguistics, Academica Sinica.
- Starosta, Stanley. 2009d [2002]. 'Austronesian 'Focus' as derivation: Evidence from nominalization'. In Elizabeth Zeitoun (ed.) *Formosan Linguistics: Stanley Starosta's Contributions. Vol. 2*, 553-606. Taipei: Institute of Linguistics, Academica Sinica.
- Starosta, Stanley. 2009e [1988]. 'A grammatical typology of Formosan languages'. In Elizabeth Zeitoun (ed.) *Formosan Linguistics: Stanley Starosta's*

- Contributions. Vol. 2*, 267-296. Taipei: Institute of Linguistics, Academia Sinica.
- Starosta, Stanley, Andrew K. Pawley & Lawrence Reid. 1982. 'The evolution of focus in Austronesian'. In Stephen A. Wurm & Lois Carrington (eds.) *Papers from the Third International Conference on Austronesian Linguistics, Vol. 2 tracking the travellers*, 145-170. Canberra: Pacific Linguistics.
- Stiebels, Barbara. 2006. 'Agent focus in Mayan languages'. *Natural Language and Linguistic Theory* 24: 501-570.
- Svelmoe, Gordon & Thelma Svelmoe. 1974. *Notes on Mansaka Grammar*. Huntington Beach, CA: Summer Institute of Linguistics.
- Swadesh, Morris. 1952. 'Lexicostatistic Dating of Prehistoric Ethnic Contacts'. *Proceedings of the American Philosophical Society* 96: 452-463.
- Szabolcsi, Anna. 1981. 'The Semantics of Topic-Focus Articulation'. In Jan Groenendijk et al. (eds.) *Formal Methods in the Study of Language*. Amsterdam: Mathematisch Centrum.
- Talla, Yahya. 1979. *The Kelabit of the Kelabit Highlands, Sarawak*. Pulau Pinang: Social Anthropology Section, School of Comparative Sciences, Universiti Sains Malaysia.
- Tallerman, Maggie. 2005. *Understanding Syntax*, 2nd Edition. London: Hodder Arnold.
- Tay, Mary. 1971. 'Report on Kelabit Survey: 27 October 1971'. Unpublished Manuscript.
- Teng, Stacy Fang-ching. 2005. 'Grammatical relations in Puyuma'. In I. Wayan Arka & Malcolm D. Ross (eds.) *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, 137-151. Canberra: Pacific Linguistics.

- Timberlake, Alan. 1977. 'Reanalysis and Actualization in Syntactic Change'. In Charles N. Li (ed.) *Mechanisms of Syntactic Change*, 141-177. Austin: Texas University Press.
- Ting, Su-Hie. 2003. 'Impact of Language Planning on Language Attitudes: A Case Study in Sarawak'. *Journal of Multilingual and Multicultural Development* 24 (3): 195-210.
- Ting, Su-Hie & Yvonne Campbell. 2007. *Emergence of Bahasa Melayu as a language for family communication: Case study of a Bidayuh family in Sarawak*. Kota Samarahan: Universiti Malaysia Sarawak.
- Tjia, Jonny. 2007. A Grammar of Mualang: an Ibanic language of Western Kalimantan, Indonesia. PhD Dissertation, Leiden University.
- Tomlin, R. 1986. *Basic Word Order: Functional Principles*. London: Croom Helm.
- Tsukida, Naomi. 2005. 'Seediq'. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.) *The Austronesian Languages of Asia and Madagascar*, 191-325. London: Routledge.
- Uriagereka, J. 1995. 'Aspects of the syntax of clitic placement in Western Romance'. *Linguistic Inquiry* 26: 79-123.
- Van Valin, Robert D. Jr. 1977. Aspects of Lakhota Syntax. PhD Dissertation, University of California, Berkeley.
- Valle, Daniel. 2011. 'Differential Subject Marking triggered by Information Structure'. Paper presented at the 5th Conference on Indigenous Languages of Latin America, University of Texas at Austin, 6-8 October 2011.
- Vennemann, T. 1974. 'Analogy in Generative Grammar: The Origin of Word Order'. In L. Heilmann (ed.) *Proceedings of the Eleventh International Congress of Linguists*, 79-83. Bologna: Il Mulino.

- Verhaar, John W. M. 1988. 'Syntactic ergativity in contemporary Indonesian'. In Richard McGinn (ed.) *Studies in Austronesian Linguistics*, 347–384. Athens, OH: Ohio University Press.
- Vincent, Nigel. 2013. 'Conative'. *Linguistic Typology* 17: 269-289.
- van de Visser, Mario. 2006. *The marked status of ergativity*. Utrecht: LOT Publications.
- Wackernagel, Jacob. 1892. 'Über ein Gesetz der indogermanischen Wortstellung'. *Indogermanische Forschungen* 1: 333–436.
- Walters, Dennis. 1994. 'Discourse-Based Evidence for an Ergative Analysis of Cebuano'. *UTA Working Papers in Linguistics* 1: 127-140.
- Walton, Charles. 1983. Sama verbal semantics: classification, derivation and inflection. MA Thesis, University of Philadelphia.
- Weber, Tobias. 2011. 'Argument marking in Western Austronesian: a cross-linguistic assessment'. Paper presented at *21st Annual Meeting of the Southeast Asian Linguistic Society*, Kasetsart University, Bangkok, May 11-13.
- Wechsler, Stephen & I Wayan Arka. 1998. 'Syntactic Ergativity in Balinese: An Argument Structure Based Theory'. *Natural Language and Linguistic Theory* 16 (2): 387-441.
- Whaley, Lindsay J. 1997. *Introduction to typology: The unity and diversity of language*. Thousand Oaks, CA: Sage.
- Whistler, K. W. 1985. 'Focus, perspective, and inverse person marking in Nootkan'. In Johanna Nichols & Anthony C. Woodbury (eds.) *Grammar inside and outside the clause: Some approaches to theory from the field*, 227-265. Cambridge: Cambridge University Press.

- White, L. & Turk, A. E. 2010. 'English words on the Procrustean bed: Polysyllabic shortening reconsidered'. *Journal of Phonetics* 38: 459-471.
- Whiteley, Wilfred H. 1972. 'Case complexes in Swahili'. *Studies in African Linguistics* 3: 1-45
- Whiteley, Wilfred H. & J. D. Mganga. 1969. 'Focus and entailment: Further problems of transitivity in Swahili'. *African Language Review* 8: 108-25.
- Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. 'Segmental durations in the vicinity of prosodic phrase boundaries'. *Journal of Acoustical Society of America* 91 (3): 1707-17.
- Wolfart, H. Christoph. 1973. *Plains Cree: A Grammatical Study*. Philadelphia: American Philosophical Society.
- Wolfenden, E. 1961. *A re-statement of Tagalog grammar*. Manila: SIL and Institute of National Language.
- Wolff, John U. 1967. *Beginning Cebuano: Part II*. New Haven: Yale University Press.
- Wolff, John U. 1996. 'The development of the Passive Verb with Pronominal Prefix in Western Austronesian Languages'. In Bernd Nothofer (ed.) *Reconstruction, Classification, Description. Papers in Honor of Isidore Dyen on the Occasion of his 80th Birthday*, 15-40. Hamburg: Abera Verlag Meyer & Co. KG.
- Woodbury, Anthony C. 2003. 'Defining Documentary Linguistics'. *Language Documentation and Description* 1: 35-51.
- Woodbury, Anthony C. 2011. 'Language Documentation'. In Peter K. Austin & Julia Sallabank (eds.) *The Cambridge Handbook of Endangered Languages*, 159-186. Cambridge: Cambridge University Press.
- Wouk, Fay. 1989. *The Impact of Discourse on Grammar: Verb Morphology in Spoken*

- Jakarta Indonesian. PhD Dissertation, University of California, Los Angeles.
- Wouk, Fay. 1996. 'Voice in Indonesian Discourse and its Implications for Theories of the Development of Ergativity'. *Studies in Language* 20: 361-410.
- Wouk, Fay. 1999. 'Sasak is Different: A Discourse Perspective on Voice'. *Oceanic Linguistics* 38 (1): 91-114.
- Wouk, Fay. 2004. 'Nasal Transitive Prefixes in Spoken Jakarta Indonesian'. In Bambang Kaswanti Purwo (ed.) *Studies in Malay Dialects, Part IV*, 21-28. Jakarta: NUSA.
- Wouk, Fay. 2010. 'Voice in the Languages of Eastern Indonesia: A Discourse Perspective'. Paper presented at the *5th International Symposium on Austronesian Languages & Literatures*. Denpasar, Indonesia, July 2010.
- Xu, Y. 2011. 'Speech prosody: A methodological review'. *Journal of Speech Sciences* 1: 85-115.
- Xu, Y. 2013. 'ProsodyPro - A Tool for Large-scale Systematic Prosody Analysis'. In Brigitte Bigi & Daniel Hirst (eds.) *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), 7-10. Aix-en-Provence: Laboratoire Parole et Langage (LPL).
- Zavala, Roberto. 1997. 'Functional Analysis of Akatek Voice Constructions'. *International Journal of American Linguistics* 63: 439-474.
- Zeitoun, Elizabeth and Chen-huei Wu. 2006. 'Reduplication in Formosan languages'. In Henry Y. Chang, Lillian M. Huang and Dah-an Ho (eds.) *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th Birthday*, 97-142. Taipei: Academia Sinica.

- Zeitoun, Elizabeth & Stacy Fang-ching Teng. 2016. 'Reassessing the position of Kanakanavu and Saaroa among the Formosan Languages'. *Oceanic Linguistics* 55 (1): 163-198.
- Zerbian, Sabine. 2006. 'Inversion structures in Northern Sotho'. *Southern African Linguistics and Applied Language Studies* 24 (3): 361-376.
- Zobel, Eric. 2002. 'The position of Chamorro and Palauan in the Austronesian family tree: evidence from verb morphosyntax'. In Fay Wouk & Malcolm Ross (eds) *The history and typology of Austronesian voice systems*, 405-34. Canberra: Pacific Linguistics.
- Zwicky, Arnold M. 1977. *On Clitics*. Bloomington: Indiana University Linguistics Club.
- Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. 'Cliticization vs Inflection: English n't'. *Language* 59 (3): 502-513.

Appendix 1

Documentation and Description of Kelabit

A1.1 Introduction

This appendix briefly summarises available resources on the Kelabit language. SUBSECTION A1.2 describes previous research into the language and SUBSECTION A1.3 describes the documentary corpus collected during my PhD research.

A1.2 Previous Documentation

There are very few references on any of the Apad Uat languages or dialects and no full length descriptive studies (cf. Martin 1996: 271). The earliest resources are word lists collected by travellers, missionaries and government officers during the period of Brooke rule, including Rutter (1929), de Crespigny (1896), Roth (1896), Douglas (1911) and Ray (1913).²⁵³ Martin (1996: 273) argues that Ray's (1913) comparative list of lexical items is the most important reference of the time as it includes roughly 200 words in seven different dialects, including Tring and Kelabit. Several works also provide anecdotal reference to the mutual intelligibility of Apad Uat varieties and

²⁵³ Sadly, most lists do not specify which dialect of Kelabit they are taken from and are subject to spelling inconsistencies (see Blust 1993).

basic, non-technical descriptions of the sound systems involved (Hose & McDougall 1912, Douglas 1911, Pollard 1933 and Bolang & Harrisson 1949).²⁵⁴

The first major attempt to describe the structure of an Apad Uat language is Southwell (1949). He developed a provisional orthography, based on the Pa Kemaloh dialect of Lundayeh, and published notes on verbal morphology. Pa Kemaloh has since become the standard variety of Lundayeh and appears in several published works, including Labo Pur's (1961) dictionary and Padan's (1971) phrase book. This served as the basis for translations with the Borneo Evangelical Mission, such as the Lun Bawang Bible, *Bala Luk Do'* (1982). Similarly, Lees (1959) developed a phonemic inventory and practical orthography for Lundayeh and Tay (1971) produced comparative linguistic notes on Lun Bawang and Kelabit.

The main descriptive works on Kelabit are Asmah (1983) and Blust (1974a, 1993, 2006 and details in 2013). Both authors present preliminary descriptions of the phonology and morphosyntax of the Bario dialect of Kelabit. Clayre (1972; 1991, 1994, 2002, 2005, 2014) presents a number of studies of Lundayeh, Lun Bawang and Sa'ban, looking particularly at phonology, morphology and the voice systems. Her work also presents an account of language change, based on over twenty years of field experience (Clayre 1994). Finally, Garman, Griffiths & Wales (1970) present a study of language acquisition among the Lun Bawang. There are also Kelabit language materials in ethnolinguistic works, such as Saging & Bulan (1989), Rubenstein

²⁵⁴ For example, Pollard (1933) states that Murut (Lundayeh) is a 'guttural' language; Bolang & Harrisson (1949) describe Sa'ban as being spoken in a 'not un-Chinese sort of singsong' and Harrisson (1961: 126) suggests that 'Kelabit has one of the more refined pronunciation patterns and sound rhythms in Borneo'.

(1973)²⁵⁵ and Bala (2002) and two short Kelabit-English dictionaries (Amster 1995, Blust 1993).

More recently, a major dictionary *Kemaloh Lun Dayeh-English* was produced (Ganang, Crain & Pearson-Rounds 2008). This does not contain a grammar sketch but does include some notes on pronunciation and an extensive list of bibliographic references (reviewed in Boutin 2010). Furthermore, the Summer Institute of Linguistics (SIL) is in the process of producing primers and children's storybooks in Kelabit and many other languages of Sarawak and there are ongoing attempts to produce a larger Kelabit dictionary by community members. Finally, there are a few online resources, including the Kelabit Portal (<http://kelabitportal.com/>) created by the Institute of Social Informatics and Technological Innovations (ISITI) UNIMAS, Sarawak.

Thus, there are resources available. However, many of these do not state where and how the data was collected, do not explain how the orthography and pronunciation are connected and/or do not provide sufficient data to conduct a complete analysis of the syntax or morphology (see Blust 1993). Hence, previous documentation and description of Kelabit and the Apad Uat languages is limited.

A1.3 The Kelabit Corpus

In order to contribute to ongoing efforts to document and describe minority languages in Borneo and Sarawak, I collected a documentary corpus of audio and video recordings of Kelabit during my PhD research. These were collected over a period of roughly six and a half months during 2013 and 2014 (see SUBSECTION 2.1) and provide

²⁵⁵ Sadly, the primary data for Rubenstein (1973) has since been lost and although a word-by-word gloss was apparently carried out, this was never published making the texts less useful to a linguist (cf. Barnes 1986).

much of the data used in this thesis. In order to be as representative, complete and comprehensive as possible, a diverse range of texts were collected (see Seifart 2008 for discussion of representativeness in language documentation). These were made accessible to the widest possible audience by including transcription into the working orthography (TABLE 2.3) and translation into English, using ELAN. Finally, metadata concerning each recording session and each speaker was recorded in an Excel file. Standard formats for annotation, grammatical description and metadata were followed wherever applicable (cf. Himmelmann 1998, 2006a, Bird & Simons 2003, Mosel 2006, Schultze-Bernd 2006, Ladefoged 2003, Bowerman 2008, Woodbury 2003, 2011, among others).

The current corpus includes audio and video recordings of varying lengths, as well as written materials. They are subdivided into elicitation, experiment and text. This division reflects different methods used in the text collection process. Elicitation involves collecting materials with explicit instructions or directions, and is therefore the least naturalistic data source (see SUBSECTION A1.3.1). In contrast, texts can be considered the most naturalistic data source, as they were recorded without specific guidelines from the researcher (see SUBSECTION A1.3.3). As discussed in Dixon (2010) and Chelliah & DeReuse (2011: 359), elicitation did not precede or follow naturalistic text collection during fieldwork. Instead, both occurred concurrently and observations from the one informed the other. I discuss each in turn in order to illustrate the different methods used in data collection.

A1.3.1 Elicitation

Over the course of two field-trips, a number of elicitation sessions were held, typically with a single consultant, and less frequently with small groups of two or three. Audio

recorded sessions were held mainly in Bario, Pa' Dalih and the UK. Written examples were also elicited and discussed with a range of speakers in the Highlands, Miri, Kuching and the UK. There is a total of roughly 27 hours of audio recordings alongside several notebooks of fieldnotes and typed-up digital notes.

Elicitation is sometimes criticised as a method of collecting linguistic data since the data are by definition influenced by the researcher (cf. Lüpke 2009, Himmelmann 2006a, Mithun 2001, Dimmendaal 2001). However, it can be used to complement naturalistic text data in providing full paradigms of forms that may appear only rarely in a corpus (Seifart 2008: 63). Moreover, it allows us to discover the metalinguistic awareness of speakers and is useful in providing negative examples of structures that are not grammatical (Lüpke 2009). I mainly conducted elicitation sessions relating to Kelabit phonology, morphology and syntax, building on existing literature or observed practices during fieldwork. During the second trip, I also discussed patterns and analyses that arose from the data collected during the first trip.

Several different methods were used in elicitation. Initially, the primary method involved using written stimuli, schedules and questionnaires and translating from English to Kelabit.²⁵⁶ For example, the Swadesh (1952) 200 Word List was used as the basis for the phonological analysis presented in SUBSECTION 2.3. Similarly, I collected sentences that reflect different tense, mood and aspect configurations by adapting the Austronesian Elicitation Schedule, originally designed for Oceanic Languages (Johnston 1989).

A second method involved semi-structured or 'analytical' elicitation in the sense of Chelliah & DeReuse (2011). Typically, this involved using a Kelabit word to

²⁵⁶ All of the speakers that I worked with were highly proficient in English. After all, English is often used as a language of inter-ethnic communication in Sarawak (see Ting 2003 for discussion of attitudes towards English and other languages in Sarawak).

elicit an example sentence containing that form without specifying an exact English translation (cf. Mosel 2006: 77). In some cases, I would modify elements in a proposed sentence or clause and ask for a grammaticality judgement. For example, during elicitation sessions relating to morphology, I used the list of verbs presented in Haspelmath (1993: 97) and the Leipzig Valency Project to elicit basic verbs of different semantic classes. Once the Kelabit verb form had been elicited, I attempted to modify the form using a list of known verbal affixes in Kelabit, taken from Asmah (1983). I would offer a form and elicit a judgement as to whether this was a possible form in Kelabit, and if so what it would mean. Where modified forms were judged to be acceptable, example sentences were then elicited to illustrate the difference between different forms of the verb. Many example sentences in SUBSECTION 2.4 were collected in this manner.

In some cases, I created entire sentences in Kelabit and sought grammaticality judgements, following what Bower (2008) terms data-manipulation. One example of this was the elicitation of pronominal paradigms (see SUBSECTION 2.4.2.8). Having elicited the basic paradigms, I then created a list of sentences for each pronoun that contained the pronoun in different functions, different positions and different clause-types. Some of these, I understood to be grammatically incorrect but wanted to check the intuition. I presented these sentences in written form, without English translation. I then asked for a grammaticality judgement for each sentence. This method has been criticised as there is a possibility that sentences are too contrived, or that sentences are judged ungrammatical for pragmatic or sociolinguistic reasons (see Abbi 2001, Chelliah & de Reuse 2011). Nonetheless, it is useful in identifying systematically ungrammatical examples, which can be checked in other contexts and using other methods.

I also used video and/or picture stimuli in order to elicit sentences and narratives. The narratives are similar to narratives in the text corpus in that they also present language in context. However, rather than the speaker deciding what he/she will say, the content of the story is predetermined by the stimulus material. This has the advantage of allowing multiple versions of similar stories to be collected. The main stimulus that was used in this way was the Pear Story, as described in CHAPTER 5 (Chafe 1980). This was piloted on the first fieldtrip, with a consultant telling the researcher the story, and then conducted with 14 speakers, resulting in a small corpus of six pear stories in Kelabit. The pear story narratives are used in the analysis of word order in CHAPTER 5. Other picture stimuli used include the Topological Relations Picture Series (Bowerman & Pederson 1992) and the Circle of Dirt picture story²⁵⁷ (Eisenbeiss & McGregor 1999).

Finally, I collected some written examples by asking my primary consultant to write regular Kelabit language tests for me. This served the dual purpose of improving my understanding of the language, and collecting sentences which had not been unduly influenced by me. Tests typically involved passages for me to translate from Kelabit to English, which could be used as the basis for discussion of particular constructions, and fill-the-gap exercises (typically omitting verbs) in order to better understand paradigmatic relations within the clause.

A1.3.2 Experiment

The methodology used in the collection of data for the prosody experiment was outlined in CHAPTER 4. 26 paragraphs were developed to include variant pronouns in

²⁵⁷ This is a series of pictures with no obvious narrative structure that can nonetheless be used to elicit narratives. I asked the consultant to describe each picture in turn and subsequently to repeat the narrative, which was audio recorded.

different contexts, which are illustrated in APPENDIX 2. These were then read twice in succession by five speakers. Hence, there are five recordings of approximately 12 minutes long. The clips from these recordings that were analysed in Praat are also included in the corpus (see SUBSECTION 4.4.4, 4.4.5 and 4.4.6).

A1.3.3 Text

If elicitation represents a structured and systematic, but arguably less natural form of data collection, then text collection represents a more natural, but arguably less systematic or less controlled method. In Seifart's (2008) terms, the selection of texts recorded in the corpus was largely opportunistic. I identified speakers both on the basis of recommendations and those speakers who I knew well and had time to work with me. Before recording, I sought informed consent, explaining the purpose of the documentation and my intention to archive the materials. If I had access to the equipment, and speakers were willing, I aimed to video record texts, in order to document the extralinguistic context of the recording. I took time to familiarise participants with the recording equipment in order to limit the potential intrusiveness.²⁵⁸

Bowern (2008) recommends documenting audio, video and written materials in as many genres as possible in order to ensure maximal representativeness. In this documentation project, I have tried to include genres that are both culturally relevant and differ along the so-called 'spontaneity parameter' (Himmelman 1998: 117) in terms of how 'planned' the recording is (Ochs 1979).²⁵⁹ This was important not just

²⁵⁸ Of course, this is not always possible and in cases where it was felt video would be too intrusive, audio recording was preferred. In a number of recordings, I am an active part of the conversation rather than an onlooker. This helps to make the recordings more natural in some ways, but could also mean that speakers are accommodating to my level of fluency.

²⁵⁹ Seifart (2008) suggests that another method of ensuring representativeness in a documentary corpus would be to employ the methods of ethnography of communication (Hymes 1971). Following Hymes

for the sake of a more representative record of the language, but also because complex linguistic structures are known to be used in different ways in planned as opposed to spontaneous speech (Seifart 2008: 65, Ochs 1979, Biber 1995). In particular, genre has been shown to influence voice alternations in Austronesian languages (cf. Kroeger 2004, Wechsler & Arka 1998) and affected word order choice in Kelabit, as discussed in CHAPTER 5.

In total, I collected roughly 7 hours of recorded text, as well as some written materials, in Bario, Pa' Umur and Pa' Dalih. The audio and video recordings can be loosely divided into the following types: conversation, procedural text, personal histories, traditional narratives, news reports, formal speech and songs. The breakdown is shown in TABLE A1.1:

Table A1.1 Text Recordings

Text Type	Total Recordings
Conversation	1 hour 40 minutes
Procedural Text	55 minutes
Personal Histories	1 hour 30 minutes
Traditional Narratives	40 minutes
News Reports	1 hour 10 mins
Formal Speech	30 minutes
Song	35 minutes

These differ in terms of the number of speakers; the conversational purpose (i.e. to inform, to persuade etc.) and the degree of spontaneity. Conversations involve multiple speakers and are typically unplanned. They include conversations between groups of men, groups of women and mixed groups. Procedural texts involve a single speaker explaining a process or a recipe and are also relatively unplanned, though

(1971), communicative events are distinguished by SPEAKING factors (Scene, Participants, Ends, Act sequence, Key, Instrumentalities, Norms and Genre). The documentary corpus includes different genres, participants, scenes and purposes (see above) but was not arranged to systematically reflect differences, given the limitations of time, resources and intrusiveness (cf. Seifart 2008, Lüpke 2009).

determined to some extent by the topic and the speaker's knowledge of that topic. Personal histories involve speakers talking about their memories of the Kelabit Highlands or particular experiences in the past and are less spontaneous than conversation, but less planned than traditional narratives, which have been rehearsed in one way or another many times before. Traditional narratives include stories involving the characters *Palug I'it* 'little liar' and *Palug Rayeh* 'big liar', as well as other folk stories. News reports are broadcast twice daily during the week and once on Saturday morning. They are typically 20 minute segments in which local and national news is read in Kelabit, drawing from newspapers and online sources, written largely in English and Malay, and prepared in advance. Formal speeches in this corpus are acted rather than given to an audience. One explains the motivation behind the Education Unit of the Kelabit Association, *Rurum Kelabit*, and the other is a recreation of a motivational speech aimed at children sitting exams, which had been observed in a similar format the previous day. Finally, the songs collected include songs for dancing, children's songs, love songs, songs of praise, *sikih*, *ri lekuweh*, *sido*, *kuwab* and *laku* and probably contain the most conservative language forms. For more information on Kelabit songs see Saging & Bulan (1989).

Texts were collected in one of two ways. Either, a speaker was available to record a number of texts and discussed with the researcher the sort of things they might like to talk about, or a speaker was observed giving a speech or telling a story and asked to record a specific communicative act. In any case, all texts can be described as more naturalistic than elicitation sessions or the paragraphs recorded for the prosody experiment, since no more instructions were given than specifying the particular topic and the particular text type. They are, of course, not completely 'naturalistic', in that speakers are aware they are being recorded and I was always present during the

recording process. Nonetheless, coupled with systematic elicited data, text data can illustrate the use of different constructions in context, as shown in CHAPTER 3 and CHAPTER 5.

Finally, the corpus includes a collection of existing written materials, including older texts, such as a traditional creation myth (Galih 1965) and transcriptions of songs in Rubenstein (1973) and Talla (1979), children’s stories written in Kelabit for a story book produced by SIL, some riddles or *iniq-iniq* published in the programme for the Bario Food Festival or *Pesta Nukenen* 2014 and a selection of stories written by Kelabit children for a school project run by eBario and UNIMAS (Rethinasamy et al 2013b). It is hoped that these may enable future research into changes in Kelabit over time, as well as between Philippine-type and Indonesian-type languages (see CHAPTER 6).

A1.3.4 Speakers

There are 33 different speakers represented in the corpus, including both men (13) and woman (20) aged between 40 and over 80. The breakdown of speakers according to their place of birth is shown in TABLE A1.2.

Table A1.2 Speakers by Place of Birth

Place of Birth	Number of Speakers
Northern villages (Bario, Pa’ Umur, Pa’ Lungan & Pa’ Main)	20
Southern villages (Pa’ Mada, Pa’ Dalih, Remudu, Long Peluan)	8
Kelabit villages outside the Highlands (Long Seridan)	1
Urban centres outside the Highlands	2
Other villages (including villages in Indonesia)	2

Most of the speakers live in Bario, Pa' Umur or Pa' Dalih – the three locations where recordings were made. All of the speakers are multilingual and some speak more than eight languages. The most common languages spoken in addition to Kelabit are Malay, English and Lun Bawang.²⁶⁰ Speakers also had varying degrees of proficiency in other local languages, including Sa'ban, Kayan, Kenyah, Penan, Bidayuh, Iban, varieties of Apad Uat languages spoken in Indonesia and Chinese dialects.

A1.4 Summary

In this appendix, I reviewed previous literature on Kelabit and the nature of the documentary corpus on which the description in this thesis is based. The corpus includes audio, video and written materials. Some are elicited directly using translation prompts or stimuli and others represent relatively naturalistic data.

²⁶⁰ Malay and English are national languages in Malaysia (see Asmah 1993). Lun Bawang is not only closely related to Kelabit, and spoken in neighbouring villages, but also the language of the SIB Church or the Borneo Evangelical Mission. Many Kelabit speakers have a copy of the *Bala luk Do'* or the Lun Bawang Bible.

Appendix 2

Prosody Experiment

The following instructions were given to participants in the prosody experiment described in CHAPTER 4. The font size of examples is smaller than the document given to participants. I also indicate which test context the example represents (see SUBSECTION 4.4.4). Otherwise, the text is identical to that used in the experiment.²⁶¹

Please recite each of the following paragraphs leaving a short pause between each. Many of the short paragraphs are repeated.

1. Edto ma'un miney uih ngalap buaq kaber. Dooq pian kuh kuman buaq nuk inih. Dadan men uih na'am neh kuman dih kemuh. (context 2)
2. Ngarang tebeyq ideh na'an. Dooq teh ileh uih ngarang kadiq di'eyq uih mey ruyung deh. Mey ni'er deh tupu teh keduih. (context 1)

²⁶¹ As discussed in CHAPTER 4, the participants in the experiment were literate in Malay and English but may not have all been overly familiar with written Kelabit. For this reason, participants were given time to read through the sentences before recording.

3. Edteh edto miney uih nalan. Neh nuh neh ni'er ieh tudo liang buaq kiran sineh. Dooq pian neh tudo na'am naruq enun-enun. (context 7)
4. Murih ketuh tieh mala dih ngekuh. Kadiq keliq uih malem neh nuk midih sineh. Na'am buriq-buriq tuih. (context 1)
5. Miney uih nekap ieh ngi kedai. Iyuk-iyuk seni'er kuh ieh tudo sebuleng. Neh nuh miney tudo ruyung neh. (context 4)
6. Neh muliq uih ngimalem. Pu'un-pu'un kinan uih edteh buaq kabere nuk pelaba laam. Da'at ketuh teh ain dih kemuh. (context 3)
7. Inan duih buaq kabere nuk laak dih keyh. Neh madaq uih Lucy marih kuman si'it kadiq am tieh neh marih betoq. (context 5)
8. Ngudeh tieh na'am medting kekuh. Uih neh nuruq ieh ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 7)
9. Muliq uih mey Bario malem. Senibu uih dooq-dooq neh latiq tauh. Dooq mulaq bera kuh ridtuq inih netoq. (context 3)
10. Murih ketuh teh ieh mala dih ngekuh. Kadiq keliq kuh malem neh nuk midih sineh. Na'am buriq-buriq teh uih. (context 2)

11. Ngarang tebeyq ideh na'an. Dooq teh ileh kuh ngarang kadiq di'eyq uih mey ruyung deh. Mey ni'er deh tupu teh keduih. (context 2)

12. Ngudeh teh ieh na'am medting kekuh. Ieh neh nuruq uih ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 6)

13. Neh muliq uih ngimalem. Pu'un-pu'un neh kuman uih edteh buaq kaber nuk pelaba laam. Da'at ketuh teh ain dih kemuh. (context 5)

14. Edto ma'un miney uih ngalap buaq kaber. Dooq pian uih kuman buaq nuk inih. Dadan men uih na'am neh kuman dih kemuh. (context 1)

15. Miney uih nekap ieh ngi kedai. Iyuk-iyuk seni'er uih teh ieh tudo sebuleng. Neh neh uih miney tudo ruyung neh. (context 3)

16. Laq tebeyq Peter mey Miri edto riak keneh. Neh neh ieh muit uih mey mayaq ieh. Inan nuk tu'en kediwéh nangey turun. (context 6)

17. Muliq uih mey Bario malem. Senibu kuh dooq-dooq neh latiq tauh. Dooq mulaq bera kuh ridtuq inih netoq. (context 4)

18. Ngudeh teh ieh na'am medting kekuh. Senuruq uih tieh ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 3)

19. Neh muliq uih ngimalem. Pu'un-pu'un neh kuman edteh buaq kabere uih nuk pelaba laam. Da'at ketuh teh ain dih kemuh. (context 8)
20. Edteh edto miney ieh nalan. Neh neh ieh neh ni'er uih tudo liang buaq kiran sineh. Dooq pian kuh tudo na'am naruq enun-enun. (context 6)
21. Miney ieh nekap uih ngi kedai. Iyuk-iyuk seni'er neh tuih tudo sebuleng. Neh nieh neh marih tudo ruyung kuh. (context 9)
22. Muliq uih mey Bario malem. Neh nuh neh nibu latiq tauh dooq-dooq. Dooq mulaq bera kuh ridtuq inih netoq. (context 7)
23. Ngudeh tieh na'am medting kekuh. Senuruq kuh tieh ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 4)
24. Neh muliq uih ngimalem. Pu'un-pu'un kinan kuh edteh buaq kabere nuk pelaba laam. Da'at ketuh teh ain dih kemuh. (context 4)
25. Ngudeh tieh na'am medting kekuh. Senuruq neh uih ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 9)
26. Laq tebeyq Peter mey Miri edto riak keneh. Nuih neh uih mey mayaq ieh. Inan nuk tu'en kedihew nangey terun. (context 9)

27. Edto ma'un miney uih ngalap buaq kaber. Dooq pian kuh kuman buaq nuk inih. Dadan men uih na'am neh kuman dih kemuh. (context 2)
28. Ngarang tebeyq ideh na'an. Dooq teh ileh uih ngarang kadiq di'eyq uih mey ruyung deh. Mey ni'er deh tupu teh keduih. (context 1)
29. Edteh edto miney uih nalan. Neh nuh neh ni'er ieh tudo liang buaq kiran sineh. Dooq pian neh tudo na'am naruq enun-enun. (context 7)
30. Murih ketuh tieh mala dih ngekuh. Kadiq keliq uih malem neh nuk midih sineh. Na'am buriq-buriq tuih. (context 1)
31. Miney uih nekap ieh ngi kedai. Iyuk-iyuk seni'er kuh ieh tudo sebuleng. Neh nuh miney tudo ruyung neh. (context 4)
32. Neh muliq uih ngimalem. Pu'un-pu'un kinan uih edteh buaq kaber nuk pelaba laam. Da'at ketuh teh ain dih kemuh. (context 3)
33. Inan duih buaq kaber nuk laak dih keyh. Neh madaq uih Lucy marih kuman si'it kadiq am tieh neh marih betoq. (context 5)
34. Ngudeh tieh na'am medting kekuh. Uih neh nuruq ieh ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 7)

35. Muliq uih mey Bario malem. Senibu uih dooq-dooq neh latiq tauh. Dooq mulaq bera kuh ridtuq inih netoq. (context 3)
36. Murih ketuh teh ieh mala dih ngekuh. Kadiq keliq kuh malem neh nuk midih sineh. Na'am buriq-buriq teh uih. (context 2)
37. Ngarang tebeyq ideh na'an. Dooq teh ileh kuh ngarang kadiq di'eyq uih mey ruyung deh. Mey ni'er deh tupu teh keduih. (context 2)
38. Ngudeh teh ieh na'am medting kekuh. Ieh neh nuruq uih ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 6)
39. Neh muliq uih ngimalem. Pu'un-pu'un neh kuman uih edteh buaq kabur nuk pelaba laam. Da'at ketuh teh ain dih kemuh. (context 5)
40. Edto ma'un miney uih ngalap buaq kabur. Dooq pian uih kuman buaq nuk inih. Dadan men uih na'am neh kuman dih kemuh. (context 1)
41. Miney uih nekap ieh ngi kedai. Iyuk-iyuk seni'er uih teh ieh tudo sebuleng. Neh neh uih miney tudo ruyung neh. (context 3)
42. Laq tebeyq Peter mey Miri edto riak keneh. Neh neh ieh muit uih mey mayaq ieh. Inan nuk tu'en kediwah nangey turun. (context 6)

43. Muliq uih mey Bario malem. Senibu kuh dooq-dooq neh latiq tauh. Dooq mulaq bera kuh ridtuq inih netoq. (context 4)
44. Ngudeh teh ieh na'am medting kekuh. Senuruq uih tieh ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 3)
45. Neh muliq uih ngimalem. Pu'un-pu'un neh kuman edteh buaq kabeh uih nuk pelaba laam. Da'at ketuh teh ain dih kemuh. (context 8)
46. Edteh edto miney ieh nalan. Neh neh ieh neh ni'er uih tudo liang buaq kiran sineh. Dooq pian kuh tudo na'am naruq enun-enun. (context 6)
47. Miney ieh nekap uih ngi kedai. Iyuk-iyuk seni'er neh tuih tudo sebuleng. Neh nieh neh marih tudo ruyung kuh. (context 9)
48. Muliq uih mey Bario malem. Neh nuh neh nibu latiq tauh dooq-dooq. Dooq mulaq bera kuh ridtuq inih netoq. (context 7)
49. Ngudeh tieh na'am medting kekuh. Senuruq kuh tieh ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 4)
50. Neh muliq uih ngimalem. Pu'un-pu'un kinan kuh edteh buaq kabeh nuk pelaba laam. Da'at ketuh teh ain dih kemuh. (context 4)

51. Ngudeh tieh na'am medting kekuh. Senuruq neh uih ngelaak ngen tauh. Na'am temen ieh idih betoq. (context 9)

52. Laq tebeyq Peter mey Miri edto riak keneh. Nuit neh uih mey mayaq ieh. Inan nuk tu'en kediwelh nangey turun. (context 9)

Appendix 3

Example Texts

A3.1 Folk Narratives

This is the tale of Dayang Beladan, as told by Gerawat Riboh in Pa Dalih on the 10th November 2013. It corresponds to the recording PDA10112013CH_01.

Uih mala edteh sekunuh lem ayuq edteh dedtur sinulaq.
1SG.1 AV.tell one story in nature one woman UV.PFV.widow
'I'm going to tell a story about a widow.'

Ngadan neh Dayang Beladan.
name 3SG.2 Dayang Tortoise
'Her name was Dayang Beladan.'

Lem edteh edto Dayang Beladan nalan~nalan lem kebun nedih.
one day Dayang Beladan REDUP~AV.walk in garden 3SG.POSS
'One day, Dayang Beladan was walking around in her garden.'

Edteh kebun ba'ung neh, kebun ubih.
one garden banana 3SG.2 garden tapioca
'It was a banana garden, a tapioca garden.'

Mulaq na'an~na'an buaq lem kebun neh.
many REDUP~type fruit in garden 3SG.2
'There were many different types of fruit in her garden.'

Jadi lem edteh edto ieh nalan~nalan.
so on one day 3SG.1 REDUP~AV.walk'.
'So one day she was walking.'

Ni'er neh edteh lawa, lawa buaq ba'ung.
 AV.see 3SG.2 one trunk trunk fruit banana
 'And she saw a tree, a banana tree.'

Laak neh idih koq.
 ripe PT DEM PT
 'And it was ripe.'

Ni'er ieh keyh, dteh ngaley kuman pudo ba'ung.
 AV.see 3SG.1 PT a marten AV.eat ripe.fruit banana
 'And she saw a yellow-throated marten eating the ripe fruit of the banana.'

Nih ngelinuh ieh ken ngudeh teh ngaley sineh murih kuman
 DEM AV.think 3SG.1 Q why PT marten DEM often AV.eat

pudo ba'ung neh?
 ripe.fruit banana DEM
 'Then she thought, oh why does this yellow-throated marten keep eating those ripe bananas?'

Dooq tuih naruq edteh ebpung pengenep kuh ieh kenek.
 good PT=1SG.1 AV.do one trap IV.catch 1SG.2 3SG.1 PT
 'I'd better make a trap so that I can catch him, she thought.'

Adiq neh neh naruq edteh buluq, buluq matey koq dudur,
 so DEM PT AV.make one bamboo bamboo dead into post

koq dudur mey ngen buaq ba'ung neh.
 into post to with fruit banana DEM
 'So she made a bamboo, a dead bamboo into a post for vegetables to climb up, into a post up to the bananas.'

Senaruq neh edteh ruwing, ruwing ateb.
 UV.PFV.make 3SG.2 one trap trap
 'She made a marten trap.'

Neh nieh muliq neh mey rumaq, mey beruhmidang periak ieh.
 DEM PT=3SG.1 INTR.return 3SG.2 go home go morning next 3SG.1
 'Then she went home and came back the next morning.'

Ni'er neh keyh, neh ayuq teh ngaley mirat let dingi.
 AV.see 3SG.2 PT DEM PT PT marten INTR.appear from over.there
 'She looked around and exactly at that moment the marten appeared from over there.'

Muned~muned metoq ngaley sineh mayaq pu'un luun buluq ih.
 REDUP~true PT marten DEM follow first onto bamboo PT
 'And the marten indeed started to climb up the bamboo.'

Edteh neh... beruwing na'ah ih dih suk senaruq nedih.
 one PT trap before PT DEM REL UV.PFV.make 3SG.POSS
 'There was a... the trap that she made.'

Rengaq idih ngaley sineh nipa uluh nedih keleyh,
 fold DEM marten DEM AV.pack head 3SG.POSS PT.M

teka ebpung ih merek kedieh ri'er na'ah ih, ngaley sineh.
 teeth trap PT AV.squeeze 3SG.EMPH neck before PT marten DEM
 'As soon as the yellow-throated marten put his head [into the trap], the teeth of the trap squeezed shut around his neck, that marten's neck.'

Mirat eh buro ieh lem daan.
 INTR.appear eh run.out 3SG.1 of hut
 'So she ran out of the hut.'

Neh neh nuwit neh kayuh mupuq... mupuq ngaley sineh.
 DEM PT UV.PFV.take 3SG.2 stick AV.hit AV.hit marten DEM
 'And took a stick to hit the yellow-throated marten with.'

Matey ngaley sineh.
 INTR.die marten DEM
 'The marten died.'

Nuit neh neh ngaley sineh.
 UV.PFV.take 3SG.2 PT marten DEM
 'She took the marten.'

Neh nieh muwer ieh.
 DEM PT=3SG.1 AV.butcher 3SG.1
 'And she butchered it.'

Um neh nieh nanek ih.
 um DEM PT=3SG.1 AV.cook PT
 'And then she cooked it.'

Pengeh nieh nanek, neh nieh kuman medto.
 after PT=3SG.1 AV.cook DEM PT=3SG.1 AV.eat midday
 'After she cooked, she ate lunch.'

Adiq medting neh edteh linuh ngeneh.
 but INTR.arrive PT one thought to.3SG.2
 ‘And then it occurred to her.’

Eh dooq tuih naruq tulang, tulang segerang nedih keneh.
 eh good PT=1SG.1 AV.do bone bone rib 3SG.POSS PT
 ‘I should do something with the bones, the rib bones.’

Tu'en kuh koq edteh ruding.
 UV.IRR.do 1SG.2 into one jaw.harp
 ‘I’ll make them into a jaw harp.’

Deh dieh ... nalap neh tulang segerang.
 DEM PT=3SG.1 UV.PFV.pick.up 3SG.2 bone rib
 ‘Then she... she took the rib bone.’

Neh nieh ne-mudut dih koq edteh ruding lah.
 DEM PT=3SG.1 PFV-AV.shape DEM into one jaw.harp PT
 ‘And shaped it into a jaw harp.’

Pemetaso ieh koq.
 IV.CAUS.distract 3SG.1 PT
 ‘To pass the time.’

Kenep~kenep edto raut ruding sineh nieh.
 REDUP~every day play jaw.harp DEM PT=3SG.1
 ‘Every day she played that jaw harp.’

Lem edteh edto pengeh ieh kuman, neh nieh ngalap
 on one day after 3SG.1 AV.eat DEM PT=3SG.1 AV.pick.up

ru ding nedih, neh nieh naruq neh ta'ang nedih keyh.
 jaw.harp 3SG.POSS DEM PT=3SG.1 AV.put PT mouth 3SG.POSS PT
 ‘One day after she had eaten she took her ruding and put it in her mouth.’

Kinih unih ruding neh.
 like.this sound jaw.harp DEM
 ‘This is the sound the jaw harp made.’

ding ding tulang danging tulang labo keneh ruwing keneh
ding ding tulang danging tulang labo keneh ruwing keneh ateb sembinging
 [rhyme]

Neh unih ruding nedih lah.
 DEM sound jaw.harp 3SG.POSS PT
 ‘That was the tune of the jaw harp.’

Naruq ieh beruh keyh.

AV.do 3SG.1 again PT

‘She played it again.’

*ding ding tulang danging, tulang danging keneh ruwing, keneh ruwing sembiring
ding ding tulang danging, tulang danging keneh ruwing, keneh ruwing sembiring*
[rhyme]

Lem kumaq ieh raut ruding nedih neh edteh metoq kuyad
whilst 3SG.1 play jaw.harp 3SG.POSS DEM one PT monkey

metutun let ngi alad daan nedih bah.
peep from at wall hut 3SG.POSS PT

‘Whilst she was playing her ruding, a monkey peeped through the wall of her hut.’

Beken men alad daan rengaq ineh koq let ngen buluq men dih.
different PT wall hut fold DEM PT from with bamboo PT DEM

‘The walls were different then, they were made of bamboo.’

Tu'en deh milaq buluq ih naruq ih koq tesaag.
UV.IRR.do 3PL.2 AV.break bamboo PT AV.do PT into strip

‘They split the bamboo and make it into strips.’

Jadi tesaag neh alad daan nih.

so strip PT wall hut DEM

‘So the walls of the hut were strips (with holes between).’

Lem ieh raut ih beruh keyh, mirat neh kuyad sineh.
in 3SG.1 play PT again PT INTR.appear PT monkey DEM

‘Whilst she was playing it again the monkey appeared.’

Neh nieh ni'er.

DEM PT=3SG.1 AV.see

‘And he watched.’

Dooq neh pian kuyad sineh ngen unih ruding ih.
good PT wish monkey DEM to sound jaw.harp PT

‘That monkey really liked the sound of the jaw.harp.’

Pengeh Dayang Beladan raut ih lem edteh edto, rudap nieh,
after Dayang Beladan play PT on one day sleep PT=3SG.1

bawur men ieh koq.

full PT 3SG PT

‘One day after Dayang Beladan had played, she slept because she was full.’

Lem kumaq ieh rudap ken ruding sineh nginih nedih tekayang
 whilst 3SG.1 sleep ? jaw.harp DEM here 3SG.POSS flat.on.back

kinih.

like.this

‘Whilst she was sleeping, the jaw harp lay flat beside her.’

Mirat kuyad sineh, masuk lem rumaq, penidut neh ruding
 INTR.appear monkey DEM enter in house UV.PFV.steal 3SG.2 jaw.harp

sineh, buro ieh, menad edteh kayuh.
 DEM run.away 3SG.1 AV.climb one tree

‘The monkey appeared, went into the house, pinched the jaw harp and ran off up a tree.’

Ngi enaq ih iring daan sineh.
 at PRO PT next hut DEM

‘That was where, next to the hut.’

Ditaq ketuwey metoq kayuh sineh.
 high really PT tree DEM

‘And that tree was really high.’

Mey tieh lem kayuq limeh ngepuluq eh kaki ditaq turun.
 go PT=3SG.1 in like fifty eh feet high perhaps

‘As high as fifty feet high perhaps.’

Jadi neh neh kuyad sineh raut dih metoq, raut ruding ih.
 so DEM PT monkey DEM play DEM PT play jaw.harp PT

‘So then the monkey played the jaw harp.’

Ieh pun mileh tieh lah.
 3SG.1 PT INTR.able PT=3SG.1 PT

‘He too was good at it.’

ding ding tulang danging, tulang labo keneh ruwing, keneh ateb sembiring
ding ding tulang danging, tulang labo keneh ruwing, keneh ateb sembiring
 [rhyme]

Neh neh unih ruding sineh.
 DEM PT sound jaw.harp DEM

‘That was the sound of the jaw harp.’

Kekeliq Dayang Beladan.
 awake Dayang Beladan

‘Dayang Beladan woke up.’

Edteh teh lemulun raut ruding neh.
 one PT person play jaw.harp 3SG.2
 ‘Someone was playing her jaw harp.’

Ni’er ruding, am teh ruding idih lem tidtuq nedih.
 AV.see jaw.harp NEG PT jaw.harp present in hand 3SG.POSS
 ‘She looked for the jaw harp but the jaw harp wasn’t in her hands.’

Nekap~nekap luun tanaq, am ieh tidih.
 REDUP~AV.search on ground NEG 3SG.1 PT=present
 ‘She looked everywhere, it wasn’t to be found.’

Napu ieh, am ieh tidih.
 AV.sweep 3SG.1 NEG 3SG.1 PT=present
 ‘She swept up, it wasn’t there.’

Adiq nieh ninger keyh, nangey teh unih ih ngi ditaq.
 so PT=3SG.1 AV.hear PT over.there PT sound PT at high
 ‘So she listened, and there was the sound of it coming from up high.’

Buro ieh let lem daan ih.
 run.away 3SG.1 from in hut PT
 ‘So she ran out of the hut.’

Ni’er neh koq, nangey teh edteh kuyad.
 AV.see 3SG.2 PT there PT one monkey
 ‘And saw that there was a monkey.’

Neh men kuyad sineh raut ruding nedih.
 DEM PT monkey DEM play jaw.harp 3SG.POSS
 ‘And that monkey was playing her jaw harp.’

Eeh at lem burur neh.
 EXCL bad in body 3SG.2
 ‘That made her really sad.’

Nangey tu’uh~tu’uh tieh.
 AV.cry REDUP~real PT=3SG.1
 ‘She cried and cried.’

Kadiq dooq men buluh neh ngen ruding ih kan.
 because good PT love 3SG.2 to jaw.harp PT PT
 ‘Because she really loved that jaw harp, didn’t she.’

Edteh edto keyh, pu'un~pu'un am tieh da'at lem burur tu'uh.
 one day PT REDUP~first NEG PT=3SG.1 bad in body real
 'The first day she wasn't actually that sad.'

Edto keduweh ih, neh tun teh kuyad sineh raut ruding ngi
 day second PT DEM PT PT monkey DEM play jaw.harp at
luun ngi, udung kayuh ih.
 on.top at top tree PT
 'The second day the monkey was playing the jaw harp again up at the top of the tree.'

Iyuk da'at teh lem burur neh, nangey nieh.
 grow bad PT in body 3SG.2 AV.cry PT=3SG.1
 'She felt worse and she cried.'

Nangey nieh, nangey kadiq ieh... ruding nedih pino
 AV.cry PT=3SG.1 AV.cry because 3SG.1 jaw.harp 3SG.POSS UV.PFV.steal
 'She cried, cried because she... her jaw harp was stolen.'

Mirat nedteh reraq tumuh.
 INTR.appear PT=one ant tumuh.tree
 'A giant ant appeared.'

Rayeh reraq tumuh sineh, ken rayeh inih lah, ken rayeh edteh...
 big ant tumuh.tree DEM as big DEM PT as big one
 'It was big, that tumuh ant, as big as this, as big as a...'

Mumak reraq tumuh sineh ni'er ieh nangey.
 AV.climb ant tumuh.tree DEM AV.see 3SG.1 AV.cry
 'The tumuh ant climbed up and saw her crying.'

Kadiq neh reraq tumuh ne-mala ngeneh,
 so PT ant tumuh.tree PFV-AV.say to.3SG.2
 'So the tumuh ant says to her,'

"Ngudeh ko nangey Dayang Beladan?" keneh ngeneh.
 why 2SG.1 AV.cry Dayang Beladan say.3SG.2 to.3SG.2
 "'Why are you crying, Dayang Beladan?'," he said to her.'

"Nih men uih... nih men edteh ruding kuh," keneh koq.
 DEM PT 1SG.1... DEM PT one jaw.harp 1SG.2 PT PT
 "'I had... I had a jaw harp", she said.'

“*Nih neh kuyad sineh...*”
 DEM PT monkey DEM
 “‘And then that monkey...’”

Nih nieh mala kinih, “oh am susah,” keneh.
 DEM PT=3SG.1 AV.say like.this oh NEG worry say.3SG.2
 ‘So he said, “don’t worry”, he said.’

“*Uih, uih nulung ko,*” *keneh.*
 1SG.1 1SG.1 AV.help 2SG.1 say.3SG.2
 “‘I’ll help you”, he said.’

“*Kapeh niko kereb tulung uih, i’it tiko?*”
 how PT=2SG.1 can help 1SG.1 small PT=2SG.1
 “‘How can you help me, you’re so small?’”

“*Rayeh men enaq ngi, kuyad ih.*”
 big PT PRO there monkey PT
 “‘And it’s so big, that monkey.’”

“*Dooq am susah,*” *keneh,* “*na’it tupu ko,*” *keneh.*
 good NEG worry say.3SG2 AV.wait only 2SG.1 say.3SG.2
 “‘Don’t worry,” he said, “just wait,” he said.

“*Mey nuih,*” *keneh.*
 go PT=1SG.1 say.3SG.2
 “‘Off I go,” he said.’

“*Mo, mey niko,*” *keneh.*
 yes, go PT=2SG.1 say.3SG.2
 “‘Ok, go on then,” she said.

Mey nieh lah.
 go PT=3SG.1 PT
 ‘So off he went.’

Menad neh ruding... enaq... reraq sineh, mey ditaq ngi.
 AV.climb PT jaw.harp... PRO... ant DEM go high there
 ‘So the jaw.harp, no, the ant climbed the tree, up high.’

Medting ieh ngi luun keyh.
 INTR.arrive 3SG.1 at on.top PT
 ‘He got up to the top.’

Neh men kuyad neh.
 DEM PT monkey DEM
 ‘And the monkey was going.’

*ding ding tulang danging, tulang danging keneh ruing, keneh ateb sembiring
 ding ding tulang danging, tulang danging keneh ruing, keneh ateb sembiring*
 [rhyme]

Kineh keneh.
 like.that PT
 ‘That’s how it was.’

Mey neh muneng~muneng keliq ni’er neh keleyh.
 go 3SG.2 REDUP~close know AV.see 3SG.2 PT.M
 ‘He went up close so he could see.’

Libuh teh teruran kuyad ih koq.
 round PT testicles monkey PT PT
 ‘The monkey’s testicles were round.’

Kadiq neh neh reraq sineh, itep neh teruran kuyad sineh kah.
 so DEM PT ant DEM UV.PFV.bite 3SG.2 testicles monkey DEM PT
 ‘So then the ant, he bit the monkey’s testicle.’

“Eeep,” keneh kah.
 eeep say.3SG.2 PT
 ‘“Eeep,” he said.’

Ni’er ieh mey beneh, keliq edteh men bukung kayuh inan neh.
 AV.see 3SG.1 go low know one PT bulge tree exist 3SG.2
 ‘He looked down and saw that there was a bulge in the tree.’

Ni’er ieh koq, ooh ineh nedih tidih.
 AV.see 3SG.1 PT ooh DEM 3SG.POSS PT=DEM
 ‘He looked and thought that must be it.’

Adiq nieh mala, “bukung kayuh,” keneh
 so PT=3SG.1 AV.say bulge tree say.3SG.2
 ‘So he said, “ahh, the bulge in the tree.”’

Am tieh te-kidut lah... um... idih teh ruding ih.
 NEG PT=3SG.1 STAT-jolt PT... um... present PT jaw.harp PT
 ‘He wasn’t surprised and the jaw harp was still there.’

Mey reraq, “ngudeh ko uto~uto,” ken reraq.
 go ant why 2SG.1 REDUP~tease say ant
 ‘So the ant thought to himself, “why are you messing around?”’

“*Jagaq ko,*” *keneh.*
 watch.out 2SG.1 say.3SG.2
 ‘“Watch out,” he said.’

Mey reraq beruh.
 go ant again
 ‘So the ant went again.’

Itep neh teruran kuyad na’ah ih beruh.
 UV.PFV.bite 3SG.2 testicle monkey before PT again
 ‘He bit the very same monkey again.’

Ma’it ruka sineh.
 AV.hurt time DEM
 ‘This time it hurt.’

“*Eeek,*” *keneh.*
 eeek say.3SG.2
 ‘“Eeeek,” he said.’

Piliu neh enaq nedih.
 UV.PFV.let.go 3SG.2 PRO 3SG.POSS
 ‘So he let go of that thing of his.’

Ruding neh tutuq mey beneh let ditaq, tutuq mey beneh.
 jaw.harp DEM fall go low from high fall go low
 ‘And the jaw harp fell from up high to down low.’

Ni’er ieh keyh, neh neh Dayang Beladan na’it.
 AV.see 3SG.1 PT DEM PT Dayang Beladan AV.wait
 ‘She was watching, and Dayang Beladan was waiting.’

Ne-keliq neh teh ruding ih tutuq let dingi
 PFV-know 3SG.2 PT jaw.harp PT fall from over.there

ne-ngesu tutuq lem edteh takung.
 PFV-AV.continue fall in one pond
 ‘She saw the jaw harp falling from up high and straight into a pond.’

Mulaq metoq lebetuh lem takung neh bah.
 many PT tadpole in pond DEM PT
 ‘There were lots of tadpoles in the pond.’

Ngem~ngem teh dteh enaq lebetuh kuman ih
 REDUP~same.time PT one PRO tadpole AV.eat PT
 ‘At the same time a tadpole ate it.’

Ne-kuman ruding sineh su lem batek nedih.
 PFV-AV.eat jaw.harp DEM straight in stomach 3SG.POSS
 ‘It swallowed the jaw harp straight into its stomach.’

Neh neh men Dayang Beladan mey ngalap iyep nedih
 DEM PT PT Dayang Beladan go AV.pick.up net 3SG.POSS

laq ngiyep ngi nieh, lemangui kinih~kinih
 DESID AV.fish there PT=3SG.1 INTR.swim REDUP~like.this

mayaq lem erang abad ih.
 follow in between reed PT
 ‘So Dayang Beladan went to fetch her fishing net to trap the tadpole,
 swimming like this between the sharp reeds.’

Eeeh nerada kedeyq batek lebetuh neh ngen iring abad ih.
 eeeh AV.cut open stomach tadpole DEM with near reed PT
 ‘This cut open the stomach of the tadpole with the edge of the reeds.’

Bilaq, besit tebuut tidih.
 broken burst completely PT=DEM
 ‘It was broken, completely burst.’

Nalap neh neh ruding nedih na’ah ih.
 UV.PFV.pick.up 3SG.2 PT jaw.harp 3SG.POSS before PT
 ‘So she picked up her jaw harp.’

Liat nieh lah.
 happy PT=3SG.1 PT
 ‘And she was happy.’

Muliq nieh mey rumaq dooq.
 INTR.return PT=3SG.1 go house good
 ‘She went home.’

Neh nieh muroq ieh dooq~dooq.
 DEM PT=3SG.1 AV.clean 3SG.1 REDUP~good
 ‘Then she cleaned it thoroughly.’

Neh nieh naq ngunih dih beruh.
 DEM PT=3SG.1 PRO AV.make.noise DEM again
 ‘And what did she do, she played it again.’

Liat netoq Dayang Beladan kadiq nalap neh neh ruding
happy PT Dayang Beladan because UV.PFV.pick.up 3SG.2 PT jaw.harp

neh muliq lah.

DEM INTR.return PT

‘Dayang Beladan was happy again because she had got her jaw harp back.’

Kuman medto periak, dih tieh enaq neh tieh raut dih
AV.eat midday next.day DEM PT=3SG.1 PRO DEM PT=3SG.1 play DEM

beruh lah.

again PT

‘After lunch the next day, she played it again.’

Keliq neh neh kuyad ih.

see 3SG.2 PT monkey PT

‘She saw the monkey.’

Mey teh kuyad mupud dih.

go PT monkey AV.stalk DEM

‘And the monkey was stalking the jaw harp.’

Ieh keliq neh kuyad, layaq iat neh ngen kuyad dih.

3SG.1 know PT monkey soft breath 3SG.2 to monkey DEM

‘She saw the monkey and wasn’t happy about it.’

Nalap neh pupuq.

UV.PFV.fetch 3SG.2 hitting.implement

‘She fetched something to hit with.’

Nukab neh bubpuq daan.

UV.PFV.open 3SG.2 door hut

‘Opened the door to the hut.’

Nalap neh dteh kayuh.

UV.PFV.fetch 3SG.2 one stick

‘Picked up a piece of wood.’

Nulin neh kuyad sineh.

UV.PFV.throw 3SG.2 monkey DEM

‘And threw it at the monkey.’

Upun kuyad sineh buro.

run monkey DEM away

‘The monkey ran away.’

Am netoq kuyad sineh ne-kasau ieh mudtih lah.
 NEG PT monkey DEM PFV-bother 3SG.1 last PT
 ‘After that, the monkey didn’t bother her anymore.’

Kadiq nieh mulun dengan senang neh Dayang Beladan.
 so PT=3SG.1 INTR.live with happy PT Dayang Beladan
 ‘So Dayang Beladan lived happily ever after.’

Pingan nieh ne-ngalap ruding nedih muliq lah.
 after PT=3SG.1 PFV-AV.get jaw.harp 3SG.POSS INTR.return PT
 ‘After she got her jaw harp back.’

Am netoq ieh susah pingan idih.
 NEG PT 3SG.1 worry after DEM
 ‘She didn’t have any more troubles after that.’

Paad ineh neh cerita, cerita sineh lah.
 even DEM PT story, story DEM PT
 ‘And that’s the end of the story.’

A3.2 News Reports

This is a news report, recorded from Radio Bario on 2nd September, 2014, and read by Connie Aping. It corresponds to recording BAR02092014CH_03.

Tauh laq ninger si’it karuh pedingeren let ngen studio tauh
 1PL.INCL DESID AV.hear little word IV.hear from to studio 1PL.INCL

alem sinih.
 evening DEM
 ‘Now we are going to hear the news from our studio this evening.’

Lawa karuh
 stem word
 ‘The headlines’

Sarawak FA, ideh mala na’am kedikamih pegamung lem tuseh
 Sarawak FA 3PL.1 AV.say NEG 1PL.EXCL.EMPH CAUS.mix in difficulty

riot GB 13.
 riot GB 13
 ‘Sarawak FA says that they are not mixed up in the GB 13 riot.’

Mawan ne-mala perlu narih liteh atau teminaq kidih~kidih.
 Mawan PFV-AV.say need IMPERS watch.out or prepare REDUP~like.that
 ‘Mawan says we need to watch out or be prepared all the time.’

Epat case a’it ALS lem Sarawak lem duweh laak ken Pengarah Kesehatan.
 four case illness ALS in Sarawak in two year say director health
 ‘There have been four cases of ALS in Sarawak in the last two years according to the Director of Health.’

Mulaq ideh nuk mudeng lem bandar peringudan mey ninger puisi
 many 3PL.1 REL INTR.stay in town REFL.rain go AV.hear poetry

atau nuk belaan.
 or REL UV.IRR.say
 ‘Many people in town sat in the rain listening to poetry.’

Merey tanaq NCR kamih muliq kedeh.
 AV.give land NCR 1PL.EXCL INTR.return say.3PL.2
 ‘They say give our NCR (Native Customary Rights) land back.’

Let Kuching
 from Kuching
 ‘From Kuching’

Ideh nuk dooq kail sukung Sarawak FA, Football Association GB
 3PL.1 REL good strong support Sarawak FA, Football Association GB

13 ne-mala ngimalem, mala kedideh na’am pegamung,
 13 PFV-AV.say yesterday, AV.say 3PL.EMPH NEG CAUS.mix

na’am pedeket lem tuseh riot suk senaruq dulun
 NEG CAUS.stick in difficulty riot REL UV.PFV.do other.people

pingan raut dih pengeh ngi State Stadium edto kenem malem.
 after game DEM finished at State Stadium day sixth last
 ‘Supporters of Sarawak FA, Football Association GB 13, said yesterday that they were not mixed up or involved in the riot that was started after the game at the State Stadium last Saturday.’

La'ih suk ne-nekaruh utung ideh, Awang Hairur Azar,
 man REL PFV-AV.speak behalf 3PL.1 Awang Hairur Azar

ne-mala ideh nuk lem GB 13 keteng lem stadium
 PFV-AV.say 3PL.1 REL in GB 13 still in stadium

ngi terrace inan deh tudo lem... lem stadium rengaq inan
 at terrace EXIST 3PL.2 sit in in stadium when EXIST

tuseh, rengaq inan kedaluh ngi lai stadium.
 difficulty when EXIST fight at outside stadium
 'The man who spoke in their behalf, Awang Hairur Azar, said that those of them in GB 13 were still in the stadium on the terrace where they were sitting in the stadium, when the riot took place, when there was fighting outside the stadium.'

Keneh mala kedikamih ngesu~ngesu menani tu'uh peh score
 say.3SG.2 AV.say 1PL.EXCL.EMPH REDUP~continue sing true PT score

Sarawak diweh Perak paad ieh ineh edteh-edteh.
 Sarawak 2DU Perak even EQUATIVE one-one
 'He said, we were still singing, even though the score for Sarawak versus Perak was a draw, that is one all.'

Mulaq dulun nutud iluq atau apui nuk pengedtang lem stadium.
 many other.people AV.burn torch or fire REL IV.light in stadium
 'Many other people started the torches or fire to set the stadium alight.'

Na'am dooq, atau na'am adil ideh mala GB 13 ne-nepu'un tuseh sineh.
 NEG good or NEG fair 3PL.1 AV.say GB 13 PFV-AV.start difficulty DEM
 'It is not fair to say that GB 13 started the riot.'

Kineh karuh nedih rengaq BP ne-...ne-pekaruh ngedeh,
 like.that word 3SG.POSS when BP PFV-RECP.discuss to.3PL.2

Borneo Post ne-nekaruh ngedeh ngen itun ideh.
 Borneo Post PFV-AV.talk to.3PL.2 to question 3PL.1
 'That was what he said when BP discussed it with them, when Borneo Post spoke to them, in answer to their question.'

Mayaq Social Media *nuk ne-ngebpen GB 13 ne-nepu'un*
 follow Social Media REL PFV-AV.accuse GB 13 PFV-AV.start

tuseh atau riot ngi lai ngi nuk inan tuduq kerita polis ne-tasaq
 difficulty or riot at outside at REL EXIST seven car police PFV-damage

penengan edteh ne-meseb.
 in.addition one PFV-INTR.burn.

‘It was social media that accused GB 13 of starting the riot outside where seven police cars were damaged and another one was burnt.’

Edteh nuk um ne-belekad atau tekayang.
 one REL um PFV-overturn or STAT-on.back
 There was one that was overturned or on its back.’

Awang Hairur ngesu ne-mala
 Awang Hairur continue PFV-AV.say

pu'un~pu'un tuseh sineh ieh ineh ngi Ipoh
 REDUP~first difficulty DEM EQUATIVE at Ipoh

rengaq peminat atau penyokong Perak diweh... ideh nuk sokong
 when fan or supporter Perak 2DU 3PL.1 REL support

Perak ne-naruq tuseh ngen limeh ideh nuk sokong panen Sarawak,
 Perak PFV-AV.make difficulty for five 3PL.1 REL support team Sarawak

ne-mala karuh~karuh nuk da'at ngedeh.
 PFV-AV.say REDUP~word REL bad to.3PL.2

‘Awang Hairur continued to say that the riot started in Ipoh when the fans or supporters of Perak starting picking on five Sarawak fans and saying bad things to them.’

Abi~abi lemulun lem stadium ne-ninger, ne-ni'er
 REDUP~all people in stadium PFV-AV.hear PFV-AV.see

ideh neh sokong Sarawak nawar ideh nuk sokong Perak
 3PL.1 PT support Sarawak AV.call 3PL.1 REL support Perak

nuruq ideh remuat.

AV.ask 3PL.1 INTR.leave

‘All the people in the stadium heard and saw the Sarawak supporters calling the Perak supporters and asking them to leave.’

Ken na'am teh lun mulaq ne-ninger... ne-ninger ineh?
 Q NEG PT people many PFV-AV.hear... PFV-AV.hear DEM
 'Was there no-one who heard that?'

Kineh karuh Awang Hairur dih.
 like.that word Awang Hairur DEM
 'That's what Awang Hairur said.'

Keneh men rengaq FAM, Football Association of Malaysia kineh teh
 say.3SG.2 PT when FAM, Football Association of Malaysia and

media nuk ngi Peninsula keteng naruq Sarawak ko'ayug lem pengeh
 media REL at Peninsula still AV.do Sarawak like in finish

na'am tuseh sinih kereb mabi.
 NEG difficulty DEM can INTR.over
 'He also said, if the FAM or Football Association of Malaysia continue to treat Sarawak like in the past the troubles will never end.'

Keteng teh inan tuseh riak neh.
 still PT EXIST difficulty future DEM
 'There will be more problems to come.'

FAS, Football Association of Sarawak, ne-mala kedideh pengeh
 FAS, Football Association of Sarawak PFV-AV.say 3PL.EMPH finish

ne-ngaduq inih ngen FAM.
 PFV-AV.report DEM to FAM
 'The FAS or Football Association of Sarawak said that they have already reported this to the FAM.'

Ineh edteh aduq tupu, na'am teh nuk senaruq ideh.
 DEM one report only NEG PT REL UV.PFV.do 3PL.1
 'But it was just a complaint and nothing has been done.'

Keneh men, ideh nuk sokong Sarawak kineh teh abi~abi
 say.3SG.2 PT 3PL.1 REL support Sarawak and REDUP~all

lun Sarawak layaq iat.
 people Sarawak soft breath
 'He also said the fans of Sarawak as well as all the people of Sarawak are unhappy.'

Ideh pengeh megkul ngen FAM suk na'am ngeremu'uh Sarawak.
 3PL.1 finish AV.give.up with FAM REL NEG AV.bother.with Sarawak
 'They have already given up on the FAM who haven't followed up on Sarawak.'

Ken Hairur men ideh nuk sokong Sarawak dooq galih ngen fans
 say Hairur PT 3PL.1 REL support Sarawak good respect to fans

atau peminat atau penyokong nuk marih let mado.
 or fan or supporter REL come from far

‘Hairur also said the Sarawak fans are respectful to fans who come from afar.’

Pengitap neh keneh JDT, Terengganu FA, Selangor FA, Kelantan FA.
 example DEM PT JDT, Terengganu FA, Selangor FA, Kelantan FA
 ‘For example, JDT, Terengganu FA, Selangor FA, Kelantan FA.’

Rengaq ideh marih mey Sarawak, peruyung, peruyung teh narih
 when 3PL.1 come go Sarawak, RECP.together RECP.together PT IMPERS

pegaber.

RECP.photo

‘When they come to Sarawak, they take pictures together.’

Peruyung teh narih ideh kayuq tudo pesiwa scarf.
 RECP.together PT IMPERS 3PL.1 like sit RECP.exchange scarf
 ‘They are happy together for example they sit together and exchange scarves.’

Neh rengaq ideh marih ne-madaq ideh dooq rurum.
 DEM when 3PL.1 come PFV-AV.show 3PL.1 good company
 ‘That’s what happens when they come to show they get along well.’

Lem pengeh~pengeh tu’en deh, tu’en ideh mekaaq teh
 in REDUP~finish UV.IRR.do 3PL.2 UV.IRR.do 3PL.1 AV.change PT

karuh kamih, karuh kamih dih.
 word 1PL.EXCL word 1PL.EXCL DEM

‘What’s happened recently is that they have changed our words, our words.’

Udung~udung neh kamih teh nuk tuseh.
 REDUP~end PT 1PL.EXCL PT REL difficulty

‘In the end, we are the ones who suffer.’

Perlu inih tu’en ngudtuq keneh.
 need DEM UV.IRR.do AV.stop say.3SG.2

‘He says this needs to stop.’

Kekamih sediaq laq kerja paad~paad ngen FAS pingan inih
 1PL.EXCL.EMPH ready DESID work REDUP~even with FAS after DEM

kinéh téh kekamih repet ngen nuk ko'ayuq inih dih
 and 1PL.EXCL.EMPH hope that REL like DEM DEM

na'am tu'en dulun beruh ken Hairur.
 NEG UV.IRR.do other.people again say Hairur
 'We are ready to work together with the FAS from now on and hope that things like this do not happen again, said Hairur.'

Ken Awang Hairur tuseh suk edto kenem malem edteh tuseh
 say Awang Hairur difficulty REL day sixth last one difficulty

riot nuk let, nuk leng~leng da'at lem sejarah raut
 riot REL from REL REDUP~very bad in history game

ebol lem negara tauh ih.
 ball in nation 1PL.INCL PT
 'Awang Hairur said that the riot last Saturday was one of the worst riots in the history of foot in our nation.'

Let ngen edteh tuseh ngi Ipoh, let ngineh saget neh video dih
 from to one difficulty at Ipoh from there fast PT video DEM

senaruq mayaq social media.
 UV.PFV.do follow social media
 'From the troubles in Ipoh, from there, videos were quick to appear on social media.'

Rengaq ineh ideh nuk pekedaluh ineh ken mulaq duweh ngeribuh
 when DEM 3PL.1 REL RECP.fight DEM as many two thousand

penyokong ngelawan public, ngelawan Public Disorder Riot Unit, PRU,
 fans against public against Public Disorder Riot Unit, PRU,

diweh Light Strike Force (LSF).
 2DU Light Strike Force (LSF)
 'At that time, there were as many as 2000 fans fighting against the Public Disorder Riot Unit (PRU) and the Light Strike Force (LSF).'

Pengeh ineh mulaq tear gas atau gas pemedtiq mateh senaruq deh
 after DEM many tear gas or gas IV.salt eye UV.PFV.do 3PL.2

paad lun mulaq dih nedteh atau buro let ngineh.
 until people many DEM AV.leave or run.away from there
 'After that, lots of tear gas was used until the crowd left the stadium.'

Suk pingan ineh, Mawan ne-mala perlu narih liteh atau
REL after DEM Mawan PFV-AV.say need IMPERS watch.out or

teminaq kidih~kidih.
ready REDUP~like.that

‘After that, Mawan said we need to watch out or be prepared at all times.’

Tuseh atau riot suk ngi Stadium Negeri edto kenem malem
difficulty or riot REL at State Stadium day sixth last

pingan raut ebol lem erang Sarawak diweh Perak
after game ball in between Sarawak 2DU Perak

mayu na'am ne-jadiq rengaq ideh nuk inan kuasa lem ineh pengeh
likely NEG PFV-arise if 3PL.1 REL EXIST power in DEM finish

liteh atau teminaq tu'uh~tu'uh.
watch.out or prepared REDUP~true

‘The riot at the State Stadium last saturday night after the game between Sarawak and Perak probably wouldn’t have happened if those in charge had watched out or been prepared properly.’

Na'am pekeneh inan tuseh kineh ken karuh menteri pembangunan sosial
in.case EXIST difficulty like.that say word minister health social

Tan Seri William Mawan.

Tan Seri William Mawan

‘This is what the Social Development Minister Tan Seri William Mawan said in regards to any troubles.’

Keneh men memang tuseh sineh na'am tetamen
say.3SG.2 PT truly difficulty DEM NEG UV.IRR.expect

kineh peh tuseh nuk ko'ayuq ineh ngi peh~peh lem inan mulaq
like.that PT difficulty REL like dem at REDUP~where in EXIST many

lemulun pemung kereb jadiq.
people RECP.gather can arise

‘He also said, even if that riot was unexpected, this sort of things can happen wherever there are big crowds.’

Kadiq neh tauh teminaq kidih~kidih ko'ayuq lem alem sineh dih.
so PT 1PL.INCL ready REDUP~like.that like in evening DEM DEM

‘That’s why we should always be prepared for nights like that.’

Tulu ideh nuk litesh~litesh lem ineh mageh ne-ngabat tuseh ineh,
if 3PL.1 REL REDUP~watch.out in DEM soon PFV-AV.prevent difficulty DEM

mayu teh tuseh ineh na'am saget ne-mirud.
likely PT trouble DEM NEG fast PFV-INTR.get.worse
'If those who are concerned with these things had tried to prevent the riot earlier,
then the riot probably wouldn't have got out of hand so quickly.'

Kineh keneh ngen ideh pemberita.
like.that say.3SG.2 to 3PL.1 reporter
'That's what he said to the reporters.'

Tuseh dih ne-jadiq lem pukul puluq edteh dedtem.
difficulty DEM PFV-arise in hour eleven night
'The riot started at 11pm.'

Puluq edteh burur emung limeh lun polis ne-murat.
eleven body including five people police PFV-INTR.injured
'11 people, including five policemen, were injured.'

Tuduq kerita polis ne-tasaq emung edteh nuk ne-belakad atau nuk
seven car police PFV-damage including one REL PFV-overturn or REL

tekayang, edteh nuk ne-meseb.
STAT.on.back one REL PFV-INTR.burn
'Seven police cars were damaged, including one that was overturned and one on
fire.'

Limeh dela'ih beruh lem libun umur puluq tuduq medting duweh ngepuluq.
five men again in circle age seventeen until twenty
'Five men between 17 and 20.'

Kadiq ineh, kadiq ineh nepu'un tuseh atau riot sineh.
because DEM because DEM AV.start difficulty or riot DEM
'They were the ones who started the riot.'

Ideh ridtuq inih tu'en polis litesh paad epat edto bulan
3PL.1 fold DEM UV.IRR.do police watch.out until four day month

September, bulan iwaq nih.
September, month nine DEM
'They are now in police custody until the 4th September this month.'

Duweh burur ideh let Kuching, duweh let ngi Kota Semerahan.
two body 3PL.1 from Kuching two from at Kota Semerahan
'Two of them are from Kuching and two from Kota Semerahan.'

Kedeh mala pu'un tuseh sineh let ideh, let ngen ideh
 say.3PL.2 AV.say start riot DEM from 3PL.1 from to 3PL.1

penyokong ngelinuh atau mala referee idih pesalag nuk senaruq
 fan AV.think or AV.say referee DEM wrong REL UV.PFV.do

kadiq ieh ne-terimaq goal suk senipa Perak ngen Sarawak.
 because 3SG.1 PFV-receive goal REL UV.PFV.pack Perak with Sarawak
 'They say the start of the riot was when the fans thought or said the referee's
 decision was wrong as he accepted the goal that Perak scored against Sarawak.'

Kekiped neh score raut sineh edteh-edteh.
 at.end DEM score game DEM one-one
 'In the end, the score of the match was one all.'

Ngen ineh Sarawak pengeh pemug let lem Campaign Malaysia Cup suk lem
 for DEM Sarawak finish out from in Campaign Malaysia Cup REL in

laak sinih.
 year DEM
 'That means that Sarawak have been kicked out of the Campaign Malaysia Cup this
 year.'

Mawan ngesu ne-mala ieh respect ieh repet ngen na'am tuseh nuk
 Mawan continue PFV-AV.say 3SG.1 respect 3SG.1 hope that NEG trouble REL

ko'ayug ineh pingan inih.
 like DEM after DEM
 'Mawan went on to say he respects, he hopes that there are no more troubles like
 this.'

Kadiq na'am lun nuk ko'ayug inih lem pengeh~pengeh.
 because NEG people REL like DEM in REDUP~past
 'Because people weren't like this in the past.'

Ieh ne-mala edteh nuk pengitap ieh ineh rengaq masa
 3SG.1 PFV-AV.say one REL example EQUATIVE when time

'ngap sayot' malem.
 buy vegetables last
 'He gave an example, namely when the slogan was 'ngap sayot' (buy vegetables in
 Sarawak Malay).'

Memang kamih keneh negara gelatey ngen mulaq nuk midih kadiq
 truly 1PL.EXCL say.3SG.2 nation excited for many REL INTR.present because

na'am metoq paad koq inan tuseh ko'ayuq inih dih.
 NEG PT even PT EXIST difficult like DEM DEM
 'We, as a nation, were excited for many things but now we are not happy when things like this happen.'

Kadiq nieh mala ideh nuk inan kuasa lem raut ineh liteh
 so PT=3SG.1 AV.say 3PL.1 REL EXIST power in game DEM watch.out

atau ni'er tu'uh~tu'uh.
 or AV.see REDUP~true

'So he said those in charge should watch out and oversee things properly.'

Ideh ngalap tindakan rengaq perlu.
 3PL.1 AV.take action when need
 'They will take action if necessary.'

Keneh men lun Sarawak na'am patut sokong nuk midih
 say.3SG.2 PT people Sarawak NEG prepared support REL INTR.present

ko'ayuq ineh.
 like DEM

'He also said, Sarawakians are not prepared to support things like that.'

Tuseh riot, ngen tuseh riot, ngen itun deh ngeneh,
 difficulty riot to difficulty riot to question 3PL.2 to.3SG.2

Mawan ne-mala rengaq perlu ideh nuk ne-naruq tuseh ineh
 Mawan PFV-AV.say when need 3PL.1 REL PFV-AV.do difficulty DEM

na'am berey mey lem stadium pingan inih.
 NEG UV.IRR.give go in stadium after DEM

'As for the riot, to their question to him, Mawan said if necessary they would prevent those who started the riot from returning to the stadium.'

Tulu leng~leng perlu kereb teh ideh nuk nepu'un tuseh ineh
 if REDUP~very need can PT 3PL.1 REL AV.start difficulty DEM

tu'en ngasuk lem jail.
 UV.IRR.do AV.enter in jail

'And if really necessary the troublemakers could be put in jail.'

Epat case ALS lem Sarawak lem lem duweh laak,
four case ALS in Sarawak in in two year

kinéh ken Pengarah Kesehatan.
like.that say director health

‘There have been four cases of ALS in the last two years, according to the Director of Health.’

Let Kuching
from Kuching
‘From Kuching’

Epat case ALS pengeh kekeliq ideh Jabatan Kesehatan
four case ALS finish ABIL.see 3PL.1 department health

lem Sarawak lem laak duweh ribuh puluq duweh
in Sarawak in year two thousand and twelve

duweh ribuh puluq teluh.
two thousand and thirteen

‘Four cases of ALS have been detected according to the Health Department in Sarawak in 2012 and 2013.’

Ieh ineh duweh case lem manid laak.
EQUATIVE two case in share year
‘That is two cases per year.’

Pengarah Kesehatan Negeri, Datuk Dr Zulkifli Jantan, ne-mala
director health state Datuk Dr Zulkifli Jantan PFV-AV.say

na'am mulaq lemulun keneh ngen a'it sineh.
NEG many people caught to illness DEM

‘The Director of Health for Sarawak, Datuk Dr Zulkifli Jantan, said not many people have caught the disease.’

Ideh nuk keneh ngen a'it sineh peped-peped ne-da'at kadiq na'am
3PL.1 REL catch to illness DEM REDUP~end PFV-bad because NEG

nuk penabat idih, tu'uh peh ideh, tu'uh pideh tu'en nabat
REL IV.cure DEM true PT 3PL.1 true PT=3PL.1 UV.IRR.do AV.cure
‘It’s a terminal illness as there is no cure even if they are treated.’

Ideh um... mudtih~mudtih da'at ayuq nideh kadiq... kadiq na'am
 3PL.1 um... REDUP~last bad PT PT=3PL.1 because because NEG

tabat ngen a'it sineh.
 medicine for illness DEM

‘In the end, they will die because there is no cure.’

Na'am mulaq lemulun nuk kereb tu'en ngelawa... nuk kereb taan
 NEG many people REL can UV.IRR.do AV.counter REL can last

ngelawan a'it sineh.
 against illness DEM

‘Not many people can... can survive against the disease.’

Ideh nuk ma'it miyuk~miyuk mey kereng.
 3PL.1 REL AV.hurt REDUP~INTR.increase go thin

‘Sufferers get thinner and thinner.’

Kadiq wang dedih miyuk~miyuk layaq
 because muscle 3PL.POSS REDUP~INTR.increase soft

um kadiq... kenep~kenep edto ngen a'it sineh.
 um because REDUP~every day from illness DEM

‘Because their muscles become weaker every day from the disease.’

Na'am tun tideh nuk ma'it ineh paad~paad tu'uh.
 NEG PT PT=3PL.1 REL AV.hurt DEM REDUP~even true

‘However, not everyone suffers from the disease in the same way.’

Ibal ideh perlu ngen kerusi roda kekiped neh.
 some 3PL.1 need for chair wheel at.end DEM

‘Some of them need a wheelchair in the end.’

Edteh dalan, edteh lawey lun mulaq kereb tulung ieh ineh naruq
 one road one path people many can help EQUATIVE AV.make

edteh teripun rigit.
 one donation money

‘One way in which people can help is to donate money.’

Abi~abi nuk berey dulun mey lem teripun sineh
 REDUP~all REL UV.IRR.give other.people go in donation DEM

tu'en pakai penulung ideh nuk ma'it.
 UV.IRR.do use IV.help 3PL.1 REL AV.hurt

‘All the contributions to this fund are used to help those who are suffering.’

Mulaq ideh nuk mudeng lem bandar perengudan mey ninger puisi
many 3PL.1 REL INTR.live in town REFL.rain go AV.hear poetry

atau pun nuk belaan.
or REL UV.IRR.say

‘Many people who live in town sat in the rain to listen to poetry.’

Nih let Kuching.
DEM from Kuching
‘This one from Kuching.’

Leng~leng dooq neh pian lun mulaq lem bandar mey ngen malem
REDUP~very good PT wish people many in town go to evening

pengucapan puisi bandar raya Kuching.
presentation poetry city Kuching
‘People in town really enjoyed going to an evening of poetry in Kuching city.’

Laak duweh ngeribuh puluq epat paad ideh perengudan mey ninger ineh
year 2014 even 3PL.1 REFL.rain go AV.hear DEM

ngi Godang Amphitheatre alem migu nuk ne-laba.
at Godang Amphitheatre evening Sunday REL PFV-pass.
‘In 2014, they sat in the rain to hear poems in Godang Amphitheatre last Sunday night.’

Tema pengucapan puisi sineh ieh ineh Malaysia di sini,
theme presentation poetry dem equative Malaysia at here

lahirnya sebuah cinta.
birth one love

‘The theme for the poetry reading was Malaysia here, the birth of love.’

Alem ineh dewan bandar Kuching utara DBKU ne-ngerayeh
evening DEM hall city Kuching north DBKU PFV-AV.big

Hari Merdeka.
day independence

‘That evening, the north Kuching city hall DBKU celebrated National Day.’

Ideh metoq ne-ngeramai ulang tahun suk keduweh ngepuluq enem
3PL.1 PT PFV-AV.busy anniversary REL twenty sixth

Kuching senaruq koq bandar raya atau city status.
Kuching UV.PFV.make into city or city status
‘They also celebrated the 26th anniversary of Kuching achieving city status.’

Pu'un~pu'un ideh ne-ni'er edteh, edteh arang atau raut senaruq
 REDUP~first 3PL.1 PFV-AV.see one one dance or play UV.PFV.do

anak sekolah Encik Buyong.
 child school Encik Buyong
 'First, they watched a dance or play by the students of Encik Buyong school.'

Pengeh ineh ideh ne-muit, ne-ngupa edteh cake.
 after DEM 3PL.1 PFV-AV.bring PFV-AV.slice one cake
 'After that they brought, they cut a cake.'

Suk pengupa cake, suk ne-ngupa cake ieh ineh Menteri
 REL INS.slice cake REL PFV-AV.slice cake EQUATIVE minister

Perumahan dan Pelancungan Datuk Amar Abang Johari Tun Openg.
 housing and tourism Datuk Amar Abang Johari Tun Openg
 'The one who cut the cake was the Minister for Housing and Tourism, Datuk Amar Abang Johari Tun Openg.'

Pengeh ineh ieh ne-mala edteh nuk belaan ieh ineh
 after DEM 3SG.1 PFV-AV.say one REL UV.IRR.say EQUATIVE

'Mencipta Malaysia Berjaya' ieh ineh 'Creating a Successful Malaysia'.
 creating Malaysia successful EQUATIVE creating a successful Malaysia
 'After that he read a poem entitled, 'Creating a Successful Malaysia'.'

Pingan ineh Tarah Menteri Pelancungan Datuk Talib zul Philip
 after DEM assistant minister tourism Datuk Talib zul Philip

suk ne-masaq 'Siapa merdeka?', 'Who is independent?'
 REL PFV-AV.read who independent who is independent
 'After that was assistant minister of Tourism, Datuk Talib zul Philip, who read 'who is Independent?'

Pingan ineh Datuk Bandar, Datuk Abang Abdul Wahab Abang Julai, ne-masaq
 after DEM city mayor Datuk Abang Abdul Wahab Abang Julai, PFV-AV.read

'Bandar raya ku gemilang', 'My glorious city'.
 city my glorious my glorious city
 'After that, the city mayor, Datuk Abang Abdul Wahab Abang Julai, read 'my glorious city'.'

Suk peped ne-masaq nuk pendingeren lem alem sineh ieh ineh
 REL last PFV-AV.read REL IV.hear on evening DEM EQUATIVE

Datuk Dr Awang Serion ne-masaq 'Pohon merdeka' atau 'Independence tree'
 Datuk Dr Awang Serion PFV-AV.read tree independence or independence tree

nuk senulis edteh, edteh burur ideh, nuk senulis burur ieh.
 REL UV.PFV.write one one body 3PL.1 REL UV.PFV.write body 3SG.1
 'The last person to read that evening was Datuk Dr Awang Serion who read
 'Independence tree' which was written by one of them, which was written by
 himself.'

Leng~leng dooq neh nuk senaruq deh pedingeren lun
 REDUP~very good DEM REL UV.PFV.do 3PL.2 IV.hear people

mulaq nuk ne-ninger dih ngi Godang Amphitheatre alem
 many REL PFV-AV.hear DEM at Godang Amphitheatre evening

migu malem.

Sunday last

'They did a really good job with the poems for the people who were listening in
 Godang Amphitheatre last Sunday night.'

Suk peped ieh ineh...

REL last EQUATIVE

'And finally...'

Merey muliq tanaq NCR kamih kedeh.
 AV.give back land NCR 1PL.EXCL say.1PL.2

'Give our NCR land back, they say.'

Let Kuching, lemulun nuk lem kampung Sungai Linkau
 from Kuching people REL in village Sungai Linkau

kampong Melanjuk, kampung Sekandu ngi Semunjan mutuh muliq
 village Melanjuk village Sekandu at Semunjan AV.request back

tanaq NCR dedih.

land NCR 3PL.POSS

'From Kuching, people in the Sungai Linkau village, the Melanjuk village and the
 Sekandu village in the Semunjan district ask for their NCR land back.'

Mayaq karuh la'ih suk nekaruh utung deh, utung ideh ieh ineh
 follow word man REL AV.speak behalf 3PL.2 behalf 3PL.1 EQUATIVE

Senabung Sampai limeh ngeribuh hectres tanaq NCR dedih
 Senabung Sampai five thousand hectres land NCR 3PL.POSS

tu'en company Kelapak Sawid pakai.
 UV.IRR.do company Kelapak Sawid use

‘According to the spokesman, Senabung Sampai, 5000 hectres of their NCR land is being used by the Kelapak Sawid company.’

Kadiq nideh lun lem kampung ineh laq taruq ideh~ideh
 so PT=3PL.1 people in village DEM DESID deed REDUP~3PL.1

lem ayuq ineh.
 about DEM

‘So the people in those villages want to take matters into their own hands about this.’

Let lem bulan waluh ideh ne-naruq edteh, ideh ne-naruq
 from in month eight 3PL.1 PFV.AV.make one 3PL.1 PFV-AV.make

etaa ngi dalan lun... lun mulaq suk mey ngi tanaq
 fence at road people people many REL go at land

NCR dedih... kamih ken la'ih neh.
 NCR 3PL.POSS 1PL.EXCL say man DEM

‘From August they have made... they have made a fence on the main road that leads to their... our NCR land, said the man.’

Paad company ineh udtuq mey lem tanaq kamih, na'am
 until company DEM stop go in land 1PL.EXCL NEG

kereb mey lem tanaq kamih.
 can go in land 1PL.EXCL

‘So that they company stops going onto our land or can't get onto our land.’

Kamih ngesu~ngesu laq ngetaa dalan paad kamih kereb setuju
 1PL.EXCL REDUP~continue DESID AV.fence road until 1PL.EXCL can agree

ngen nuk petuwen kamih.
 to REL UV.IRR.request 1PL.EXCL

‘We will continue to block the road until they (here ‘we’ in text) agree to our demands.’

Na'am kamih pian ngen tanaq kamih inan merey inan merey nuk
 NEG 1PL.EXCL wish that land 1PL.EXCL EXIST AV.give EXIST AV.give REL

koq siwa dih.
 into exchange DEM
 'We don't want our land to be replaced.'

Senabung ne-mala tanaq dedih ineh let ngen lun merar
 Senabung PFV-AV.say land 3PL.POSS DEM from to people big

dedih ngilad.
 3PL.POSS in.past
 'Senabung said the land belonged to their elders in the past.'

Let let lun merar ngilad~ngilad sebelum teh inan atau
 from from people big REDUP~past before PT EXIST or

edteh perinteh negeri idih.
 one government state present
 'It was used by their ancestors long before there was a state government.'

Tanaq ineh emung tanaq nuk inan sibu dedih, inan deh
 land DEM include land REL exist plant 3PL.POSS EXIST 3PL.2

nibu paraq, tanem~tanem ma'un dedih
 AV.plant rubber REDUP~burial.site original 3PL.POSS

tembawai emung nuk pengeh ne-tasaq senaruq company
 burial.site (Iban?) include REL finish PFV-damage UV.PFV.do company

Kelapak Lawid.
 Kelapak Lawid
 'That land includes land where they plant, where they plant rubber trees, old burial sites, tembawai, including ones that have already been damaged by the company Kelapak Lawid.'

Iwaq ngepuluq teluh lubang rumaq let teluh rumaq kadang nuk tegiuq
 ninety three hole house from three house long REL shaken

senaruq company kadiq ideh pakai tanaq kamih kedeh.
 UV.PFV.do company because 3PL.1 use land 1PL.EXCL say.3PL.2
 '93 families from three long houses have been affected by the company because they are using our land, they say.'

Kekamih periput, perepimet ngen tanaq kamih kadiq tanaq
1PL.EXCL.EMPH ? REFL.hold to land 1PL.EXCL because land

ineh koq ulun kamih kedeh.
DEM into life 1PL.EXCL say.3PL.2

‘We want to hold on to our land because it is our way of life they say.’

Senabung ne-mala case dedih pengeh nuit mey ngi, mey ngi
Senabung PFV-AV.say case 3PL.POSS finish AV.take go at go at

court atau court besara lem laak duweh ngeribuh puluq duweh.
court or court high in year 2012

‘Senabung said their case had already been taken to court, or High Court, in 2012.’

Tapi High Court na'am neh ne-ngakim lipa ngen ideh lem
because High Court NEG 3SG.2 PFV-AV.judge side to 3PL.1 in

bulan April lem laak duweh ngeribuh puluq teluh.
month April in year 2013

‘But the High Court did not rule in their favour in April 2013.’

Pingan ineh pengeh neh kamih ne-appeal.
after DEM finish PT 1PL.EXCL PFV-appeal

‘After that we appealed again.’

Kadiq paad ridtuq inih na'am teh nuk senaruq, na'am teh
but until fold DEM NEG PT REL UV.PFV.do NEG PT

nuk seninger kamih let ngedeh lem ayuq nuk midih sinih
REL UV.PFV.hear 1PL.EXCL from to.3PL.2 about REL thing DEM

paad edto kinih.
until day today

‘But to this day nothing has been done and we have heard nothing from them about this.’

Kamih repet ngen perintah kereb tulung kereb nengan kamih
1PL.EXCL hope that government can help can help 1PL.EXCL

ngen nuk midih sinih.
with REL thing DEM

‘We hope that the government can help in this matter.’

Paad ineh karuh pendingeren tauh lem alem sinih mayaq
over DEM word IV.hear 1PL.INCL in evening DEM follow

Radio Bario. Terimakasih.

Radio Bario thank you

‘That was the news from Radio Bario tonight. Many thanks.’