THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

COPULAR CLAUSES IN MALAY:

Synchronic, Diachronic, and Typological Perspectives

Amir Rashad Mustaffa

Thesis submitted for the degree of
Doctor of Philosophy

School of Philosophy, Psychology & Language Sciences
The University of Edinburgh

2022

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

Signed

Date……16/05/2022……

Abstract

This thesis examines the copulas *ialah* and *adalah* in Malay on different levels of linguistic analysis, in different periods in time, and against different genetically related languages. Addressing the scarcity of research on copular clauses in Malay in all three areas, namely synchrony, diachrony, and typology, this thesis aims to serve as a point of reference for future study on nonverbal predication in Malay and beyond.

The synchronic portion of the thesis begins with a demonstration of the monomorphemic nature of the two copulas, which no longer exhibit the morphosyntax, semantics, and information structure of the morphemes that they appear to comprise, viz. 3rd person *ia*, existential verb *ada*, and focus marker *lah*. Following that, several syntactic and semantic phenomena, including extraction from copular clauses, copular inversion, and overt vs. zero encoding of the copula, are investigated. Lastly, the derivation of clefts in Malay is examined, which I reveal to be a type of copular construction despite the absence of an overt copula. I then show that the derivation of a cleft feeds the further derivation of a pseudocleft via remnant movement.

In the history of Malay, *ialah* and *adalah* are shown to have emerged relatively recently, that is towards the end of the Classical Malay era, circa the 18th to 19th century. *Ialah* grammaticalised from the combination of 3rd person pronoun *ia* and comment marker *lah* in a topical construction that involved left dislocation. Specifically, the topic was reanalysed as the canonical subject, which subsequently forced the resumptive pronoun to undergo Spec-to-Head reanalysis, resulting in *ialah* grammaticalising into a copula heading TP. Meanwhile, *adalah* grammaticalised from semantically vacuous support auxiliary *ada*, also in combination with comment marker *lah*. Both copulas originally developed from the need to provide a host for the comment marker as a way of avoiding a violation of the stray affix filter.

The typological survey of copular clauses in Austronesian reveals that syntactic alignment and word order play a central role in the emergence of copulas in a language. Of the 40 languages examined, all the 19 languages that have overt copulas are accusatively aligned, except the ergatively aligned Formosan language Puyuma, which entails that ergative-absolutive and split ergative languages within Austronesian are statistically very unlikely to have overt copulas. In addition to that, 20 of the 25 accusatively aligned languages have SVO word order, whilst all of the 9 ergatively aligned languages have VSO word order. The word order of the language is relevant as all but two of the 19 languages with overt copulas have SVO word order. In consideration of these findings, I argue that the correlation among the three factors is such that change from ergative to accusative alignment triggers change in word order from verb-initial order to verb-medial order, and that this is conducive to the emergence of overt copulas. Furthermore, word order plays a crucial role in the emergence of overt copulas as they may develop in topical constructions following reanalysis of the left-dislocated topic as the canonical subject, as argued in the diachronic portion of the thesis. Given this path of development, I argue that pronominal copulas have not been able to develop in the ergatively aligned Philippine-type languages due to the lack of the notion of subject and the absence of the canonical subject position, which prevents reanalysis of left-dislocated topics as canonical subjects and subsequently resumptive pronouns as copulas, as undergone by the Malay copula *ialah*.

In addition to that, verbal copulas cannot develop from posture verbs in the Philippine-type languages because of the clash between the unergative nature of posture verbs and the unaccusative nature of the copula, which presents a problem in the Philippine-type languages due to the encoding of the agent argument on the verb in the actor voice. Besides, the strict intransitive nature of the copular clause is incompatible with other voice alternations such as the benefactive and the locative, as the trigger in these voice alternations is encoded as an applied argument, making the clause transitive. Verbs of becoming also cannot copularise in the Philippine-type languages via semantic bleaching of the inchoative aspect, due to the robust morphological marking of aspect on the verb.

Lay Summary

This thesis is an examination of the following words in Malay: *ialah* and *adalah*. I investigate the properties of these two words as copulas, akin to English *be*, in sentences with nonverbal predicates of the likes of "John *x* doktornya" (John *x* the doctor), in which the copula *x* intervenes between the subject *John* and the nonverbal predicate *doktornya*. I investigate how the two copulas behave in the current stage of the language (synchrony), how they developed in the history of the language (diachrony), and how they compare with copulas in other genetically related languages (typology).

I first demonstrate some very interesting properties and behaviours of the two copulas in the current stage of the language. For example, in spite of their appearance, the string *lah* may no longer be separated from the copulas *ialah* and *adalah*, which means that the pronoun *ia*, the verb *ada*, and the particle *lah* have fused together to form monomorphemic words (i.e. words that consist of a single component such as *cat*).

Historically, the two copulas started out as words that reflected their appearance. Specifically, the copula *ialah* developed from the pronoun *ia*, whilst the copula *adalah* developed from the verb *ada*. These roots of the copulas were used in combination with the particle *lah* to prevent the particle from being stranded, considering that it could not stand on its own and requires other material to latch onto.

The survey of 40 different Austronesian languages reveals that the word order Subject-Verb-Object (SVO) is important in the emergence of copulas in a language. Considering this, many of the languages examined lack pronominal copulas (those that develop from pronouns, e.g. Malay *ialah*) due to deviations from this word order. In addition to that, verbal copulas (those that develop from verbs, e.g. Malay *adalah*) also cannot develop in many of the languages examined due to morphological markings on the verb that are incompatible with certain properties of copulas.

Acknowledgements

I cannot possibly fit each and every person to thank on a couple of pages as so many people have helped me towards completing this thesis. Know that, even if I have absent-mindedly excluded your name here, your contribution means a great deal to me, no matter how little it might be.

First and foremost, my supervisors: Dr. Robert Truswell, Prof. Caroline Heycock and Prof. Nikolas Gisborne. I would like to thank Rob, especially for keeping me on a long leash throughout my studies. Other than the freedom that he has allowed for me to work on my own accord, the academic support by him has allowed me to learn a great deal, not just about linguistic theory but also about argumentation, presentation, and other valuable, transferable skills. Thanks to Caroline for all the advice that she has given with respect to my research on copular clauses in Malay. I have had the privilege of being taken under the wing of a renowned figure in the field of nonverbal predication such as her. Also, thanks to Nik for steering me into information structure in the early stages of the PhD. Although Nik had only been my supervisor for the 1st year of my PhD, my research had grown more rounded under his supervision, without which it would have been too syntax-centric and lacked a great deal of interesting data. All three of these brilliant people have contributed so much to my growth as a researcher and linguist.

Also deserving of recognition are my parents, Mustaffa Mohamed Zain and Rashidah Mohamad. Without the precedent that they have set as successful academics and their enthusiasm towards my education, I would not have aspired to climb to the pinnacle of the ivory tower and dared to delve into academia. Although our academic disciplines differ, my respect for them and the work that they do has endowed me with an appreciation for scholarship and the effort poured into creating knowledge. The environment in which they have raised me emphasised the importance of learning, and I have benefitted from their constant encouragement to learn new things, to learn from my mistakes, and to just learn.

I would like to give special thanks to all the consultants who have had to painfully answer the single most annoying question time and again *ad nauseam*: “pelik tak ayat ni?” (is this sentence weird?) *Bitinis, uols memang melets! Chrissy, maraming salamat po!* Without them, I would not have been able to obtain the grammaticality judgements required for me to make conclusions about some of the most basic things related to the copulas *ialah* and *adalah*.

Lastly, I am grateful to the University of Malaya and the Malaysian Ministry of Higher Education for the generous financial support that has afforded me this degree. The trust and support shown by Prof. Stefanie Pillai, Assoc. Prof. Surinderpal Kaur, and Dr. Soh Bee Kwee from the Faculty of Languages and Linguistics, in and for my research on copular clauses in Malay made it possible for me to land the sponsorship. Also, thanks to my eldest brother Mohd Ashraf and my dear friends Lincoln and Joanne for all the help getting the paperwork for the sponsorship done whilst I was in the UK.

Demi kesinambungan bahasa tercinta!

Table of Contents

Declaration iii
Abstract v
Lay Summary vii
Acknowledgements ix
Table of Contents xi
List of Tables xv
List of Figures xvi
List of Glosses xvii

Chapter 1 : Introduction 1

1.1 Background 3
1.1.1 Nonverbal Predication 3
1.1.2 Copulas 6
1.2 Research Questions 9
1.3 Framework and Methodology 11
1.3.1 Synchrony 11
1.3.2 Diachrony 13
1.3.3 Typology 14
1.4 Overview of the Malay Language 14
1.4.1 Standard Malay vs. KL Malay 16
1.4.2 Malay and Austronesian Alignment 18
1.5 Basic Properties of Malay Grammar 22
1.5.1 General Morphology 22
1.5.2 Word Order and the Preverbal Subject 24
1.5.3 Syntactic Categories 27
1.5.4 Tense, Aspect, and Mood 32
1.5.5 Finiteness 36
1.5.6 Negation 38
1.5.7 Verbs and Verbal Morphology 39
1.5.8 Voice 42
1.5.9 Demonstratives 44
1.5.10 NP Modifiers 44
1.5.11 Pronouns 46
1.5.12 Ā-Movement 47
1.5.13 Extraction 48
1.5.14 *Wh*-Interrogatives 51
1.5.15 Polar Interrogatives 53
1.6 Delimiting the Notion of Copula 56
1.6.1 The Copular Status of *Ialah* and *Adalah* 56
1.6.2 Linking Verbs 59
1.6.3 Other Non-Copulas 63

Chapter 2 : The Malay Copulas 66

2.1 *Ialah* 67
2.1.1 Semantics 67
2.1.2 Information Structure 70
2.1.3 Morphology 74
2.1.4 Distribution 76

2.2 *Adalah* 79
2.2.1 Semantics 79
2.2.2 Information Structure 80
2.2.3 Morphology 81
2.2.4 Distribution 83

2.3 *Itu* 86
2.3.1 Semantics 86
2.3.2 Morphology 87
2.3.3 Distribution 88

2.4 Coexistence of the Monomorphemic and Bimorphemic Forms 90

2.5 Summary 92

Chapter 3 : The Syntax of Copular Clauses 94

3.1 Misconceptions about Copular Clauses in Malay 95
3.1.1 Categorical Selection 95
3.1.2 Co-occurrence of Copulas and Verbs 99
3.1.3 The Internal Structure of the Copular Clause 103

3.2 The Category and Position of the Copulas 106

3.3 Subjects in Copular Clauses in Malay 109
3.3.1 Selection of 3rd Person Subjects by *Ialah* 110
3.3.2 Specificational Copular Clauses 112
3.3.3 Copular Inversion 115
3.3.4 Extraction from Copular Clauses 122

3.4 Summary 130

Chapter 4 : Overt Encoding of the Copulas 131

4.1 Beyond Economy and Parsing 132

4.2 Aspect and Eventiveness of the Predicate 135
4.2.1 Individual-Level vs. Stage-Level 135
4.2.2 States vs. Events 138

4.3 Coercion and Grammatical Aspect 140

4.4 Inner AspP as the Main Factor 145
4.4.1 Motivation for Inner AspP 148

4.5 A Further Problem: Verbalisation of the Predicate 153

4.6 Verbal Encoding of Temporary States 157
4.6.1 Differentiating Adjectives and Verbs 162

4.7 Summary 168

Chapter 5 : Cleft Constructions 170

5.1 On Cleft Constructions in Malay 171
5.1.1 The Structure of the Cleft Clause 175
5.1.2 The Malay Cleft as a Copular Construction 179

5.2 The Derivation of Cleft Constructions in Malay 185
5.2.1 Against Extraposition of the Cleft Clause to Derive Clefts 186
5.2.2 For Intraposition of the Cleft Clause to Derive Pseudoclefts 189
5.2.3 Connectivity as a Corollary of Intraposition 192
5.2.4 On the Split-CP Hypothesis and Phasehood 194
5.2.5 Topic as the Trigger for Movement 198
5.2.6 *Yang* and Topicality 201

5.3 Against the Cleft Clause as a Complex DP 202
5.3.1 Syntactic Islands 204
5.3.2 Inversion 206
5.3.3 Relativisation vs. (Pseudo)Clefting 207
5.3.4 Optionality of the NP 211
5.3.5 Definiteness and Referentiality 212
5.3.6 Licensing of NP-Ellipsis 218
5.3.7 Information Structure 221

5.4 Summary 225

Chapter 6 : The Diachrony of the Copulas 226

6.1 Change in Word Order 227

6.2 Copular Clauses throughout the History of Malay 235
6.2.1 The Grammaticalisation of *Adalah* 240
6.2.2 The Grammaticalisation of *Ialah* 245

6.3 The Potential Role of Language Contact 253

6.4 From Diachrony to Synchrony 257
6.4.1 Verum Focus 257
6.4.2 3rd Person Subject Selection 260
6.4.3 Predication vs. Specification 261
6.4.4 Atemporality 261
6.4.5 The Structure of Clefts 263

6.5 Summary 265

Chapter 7 : The Copulas in Typological Perspective 267

7.1 Alignment, Word Order, and the Emergence of Copulas 268
7.1.1 Change in Alignment from Ergative to Accusative 274
7.1.2 Change in Word Order and Reanalysis of the Trigger as the Subject 280

7.2 Pronominal Copulas 286
7.2.1 No Subject, No Pronominal Copula in Tagalog 288

7.3 Verbal Copulas 291
7.3.1 Posture Verbs and Light Verbs 292
7.3.2 Verbs of Becoming 294

7.4 Summary 296

Chapter 8 : Conclusion 298

8.1 Revisiting the Research Questions 299

References 302

List of Tables

Table 1: The Contemporary Sources Used 11
Table 2: The Historical Texts Analysed 14
Table 3: The Absence of Case in the Nominal Category 22
Table 4: The Absence of Subject-Verb Agreement 23
Table 5: The Auxiliaries in Malay 33
Table 6: The Heads of VoiceP 41
Table 7: The Difference in Distribution between Copula 'Ialah' and PRON-FOC Form 76
Table 8: The PRON-FOC Forms as Clefted Constituent 77
Table 9: The Distributions of the Different Forms of 'Ada-lah' 83
Table 10: Summary of Comparison between Copula 'Ialah' and the PRON-FOC Forms 92
Table 11: Summary of Comparison between Copula 'Adalah' and the Ada-FOC Forms 93
Table 12: Summary of Comparison between Copula and Demonstrative 'Itu' 93
Table 13: Referentiality According to Type of Copular Clause 96
Table 14: The Malay Copulas and Their Predicates According to Karim et al. (2014) 97
Table 15: The Nonverbal Predicates Compatible with 'Adalah' According to Scholar 97
Table 16: Predicates Following 'Adalah' in DBP Corpus of Utusan Newspapers (Mustaffa, 2018) ... 99
Table 17: The Subjects of Copular Clauses with 'Ialah' in DBP Corpus of Modern Novels 111
Table 18: The Subjects of Copular Clauses with 'Adalah' in DBP Corpus of Modern Novels 112
Table 19: The Spell-Out of 'Yang' in Copular Clauses Following Extraction 123
Table 20: The Correspondence among Predicate Type, Eventuality, and Encoding of the Copula ... 135
Table 21: Aspectual Interaction and the Resulting Eventiveness 144
Table 22: The Spell-Out of the Copula According to Predicate Type 146
Table 23: Predicates According to Their Occurrence with the Copula and Inchoative Verb 155
Table 24: The Spell-Out Patterns of the Copula and the Inchoative Verb 155
Table 25: The Spell-Out of the Copula and the Inchoative Verb According to Category 157
Table 26: The Similarities between Cleft Constructions and Extra- and Intra-position 186
Table 27: The Subject-like Behaviour of the Clausal Constituent 192
Table 28: The Functions of 'Yang' 201
Table 29: Comparison between Argument and Adjunct Relative Operators 208
Table 30: Comparison of Wh-Operators and 'Yang' between Relative Clauses and Cleft Clauses .. 211
Table 31: Classical Malay Constituent Order (Cumming, 1988) 229
Table 32: The Number of Occurrences of Auxiliary 'Ada-lah' 242
Table 33: The Number of Occurrences of 'Ia-lah' 251
Table 34: Comparison of Austronesian Languages 270
Table 35: ±COP vs. ±AGR Contingency 271
Table 36: ±COP vs. ±TAM Contingency 271
Table 37: Alignment vs. Word Order 272
Table 38: Word Order vs. Overt Copula 272
Table 39: The Stages towards Copula Emergence 274
Table 40: Austronesian Languages with Demonstrative Copulas 287
Table 41: The Origins of the Copulas in Austronesian Languages 291

List of Figures

Figure 1: Flowchart for Identifying Copulas 8
Figure 2: The Lectal Continuum in Kuala Lumpur 12
Figure 3: The Timeline of the Emergence of Proto-Malayic based on Bellwood (2007) 15
Figure 4: The Austronesian Language Tree 15
Figure 5: Map of Areas Where Malayic Isolects are Spoken (Adelaar, 1993) 16
Figure 6: The Ergative-Absolutive Alignment in Tagalog 20
Figure 7: Languages Evolving from Ergative to Accusative (Aldridge, 2007) 21
Figure 8: The Nominative-Accusative Alignment in Malay 22
Figure 9: Movement of the Subject to SpecTP 25
Figure 10: The Syntactic Structure of a Transitive VP in Malay 41
Figure 11: The Structure of a Polar Interrogative in Malay 55
Figure 12: The Supposed Morphology of the Copulas 66
Figure 13: The Full Syntactic Structure of the Copular Clause 109
Figure 14: The Derivation of a Specificational Copular Clause 114
Figure 15: The Derivation of an Inverted Verbal Predicate 120
Figure 16: The Derivation of an Inverted Adjectival Predicate 121
Figure 17: The Scale of Referentiality 129
Figure 18: The Structure of Copular Clauses with Inner AspP 148
Figure 19: The Movement of 'Jadi' to Asp^0 and $Voice^0$ 151
Figure 20: The Verbalisation of a Temporary State 156
Figure 21: The Revised Time-Stability Scale Adapted from Givón (1984) 160
Figure 22: The Structure of an Interrogative Cleft According to Aldridge (2007) 172
Figure 23: The Structure of a Cleft According to Kader (1981) 174
Figure 24: The Structure of a Cleft According to Cole and Hermon (2000) 175
Figure 25: The Structure of the Cleft Clause in Malay 179
Figure 26: The Structure of the Matrix Copular Clause of Clefts in Malay 179
Figure 27: The Structure of a Cleft in Malay 184
Figure 28: The Structure of a Pseudocleft in Malay 186
Figure 29: Extraposition in the Derivation of Clefts (Percus, 1997) 187
Figure 30: The Reconstruction and Connectivity Caused by Movement 193
Figure 31: The [uTop] Feature in Pseudoclefts in Malay 200
Figure 32: One Dead Person in View 217
Figure 33: The Difference in the Antecedents of the Gaps 221
Figure 34: The Factors that Changed the Word Order of Malay 227
Figure 35: The Structure of the 'Pun-Lah' Construction 233
Figure 36: The Functions of 'Ialah' and 'Adalah' at Different Stages of Malay 240
Figure 37: The Frequency of Auxiliary 'Ada-lah' 243
Figure 38: The Grammaticalisation of Copula 'Adalah' 245
Figure 39: The Structure of a Copular Clause in Classical Malay with 'Ia-lah' 248
Figure 40: The Spec-to-Head Reanalysis of 'Ialah' 250
Figure 41: The Frequency of 'Ia-lah' 252
Figure 42: The Grammaticalisation of Copula 'Ialah' 253
Figure 43: The Cline of the Emergence of Pronominal Copulas in Austronesian 273
Figure 44: The Passive-Patient Split 278
Figure 45: The Position of the Patient in SpecTP in the Ni- Construction in Old Malay 282
Figure 46: The Structure of an Intransitive Negative Polar Interrogative in Tagalog 284
Figure 47: The Detailed Cline of the Emergence of Pronominal Copulas in Austronesian 285

List of Glosses

Gloss	Meaning	Gloss	Meaning
1	First person	IMPRF	Imperfective aspect
2	Second person	INAN	Inanimate
3	Third person	INCEP	Inceptive aspect
ABS	Absolutive	INCL	Inclusive
ACT	Active voice	INF	Infinitive
ADV	Adverb	INTR	Intransitive
AFF	Affirmative	LIG	Ligature
AG	Agentive	LNK	Linker
AN	Animate	LOC	Locative
APPL	Applicative	MED	Medial
ART	Article	N	Noun
ASP	Aspect	NEG	Negator
ATTR	Attributive	NDEF	Indefinite
AUX	Auxiliary	NH	Non-human
AV	Actor voice	NMZ	Nominaliser
Capital letter.	Name (e.g. A. = *Alex*)	NVOL	Non-volitional
CAUS	Causative	OBJ	Object marker
CLF	Classifier	OBL	Oblique
CNTR	Contrastive marker	PASS	Passive voice
COM	Comment marker	PL	Plural
COMP	Complementiser	PRED	Predicate marker
COND	Conditional	PREP	Preposition
CONT	Continuative aspect	PRF	Perfective aspect
COP	Copula	PRS	Present tense
CTOP	Contrastive topic	Pro	Null pronoun
DEF	Definite	PRO	Null pronoun in control
DET	Determiner	PROG	Progressive aspect
DISC	Discourse marker	PROH	Prohibitive mood
DISJ	Disjunctive marker	PROS	Prospective aspect
DIST	Distal	PROX	Proximal
EMPH	Emphatic	PRT	Particle
ERG	Ergative	PST	Past tense
EXCL	Exclusive	Q	Interrogative marker
EXIST	Existential	RED	Reduplication
EXP	Experiential aspect	REL	Relativiser
EXPL	Expletive	SG	Singular
FOC	Focus	STAT	Stative
FUT	Future tense	SUP	Superlative
GEN	Genitive	TOP	Topic
HON	Honorific	TR	Transitive
H	Human		

Chapter 1: Introduction

This thesis explores one aspect of the grammar of Malay that has received little attention, copular clauses.[1] Although there have been some studies on copular clauses, they are disparate, narrow, and too few to allow for a comprehensive view of the grammar of these constructions in Malay. This general scarcity of information pertaining to copular clauses in the language forms the intellectual gap which this thesis seeks to fill.

What this thesis endeavours to accomplish is not just an extensive description of copular clauses in Malay, but a deeper understanding of nonverbal predication in the language. I examine copular clauses in Malay from synchronic, diachronic, and typological perspectives in order to provide a descriptively adequate account of these constructions. Several issues concerning the copulas encompassing different levels of linguistic analysis, which have escaped satisfactory explanation by linguists, are identified and explicated. The following are the issues addressed according to chapter:

- The morphosyntactic features of the Malay copulas
- The syntax of copular constructions in Malay
- The semantic rules that govern overt encoding of the copulas
- The derivation of cleft constructions in Malay
- The historical development of the Malay copulas
- The typology of copulas in Austronesian

First, I identify the basic morphosyntactic and interpretive properties of the putative copulas, *ialah* and *adalah*, to demonstrate that they are monomorphemic items that are no longer associated with the morphemes that they appear to be comprised of, namely 3rd person *ia*, existential verb *ada*, and their combinations with focus marker *lah*. I also provide a description of the use of distal demonstrative *itu* in copular clauses as a form of copula.

[1] Malay is also formally known as *Bahasa Malaysia, Bahasa Indonesia, Bahasa Brunei* in Malaysia, Indonesia, and Brunei respectively (*bahasa* = 'language'). Although there are regional differences between these varieties in terms of lexis, they are syntactically similar, and speakers share mutual intelligibility; they will henceforth be referred to as *Malay*.

Furthermore, I examine the syntax of copular clauses in Malay and identify the structural position of the copulas in the syntax; they are auxiliaries that head TP. Then I develop a principled account of several phenomena that show the privileged status of the subject in copular clauses, such as the selection of 3rd person subjects by *ialah*, the correspondence of subject with topic in specificational copular clauses, and the possibility of clefting only the subject of predicational copular clauses.

The syntactic analysis then leads to an elucidation of the semantic rules that govern whether the copula *adalah* may be realised overtly in a copular clause, as opposed to the misinformed general view that the copulas are always optional. Although it has been stated that nonverbal predication in Malay is affected by the permanence of the predicate, as described by Pustet (2003) and Stassen (1997), this restriction only happens to be one part of a greater generalisation, which involves both lexical and grammatical aspect. Overt encoding of the copula is only possible in copular clauses that have an atemporal interpretation.

To conclude the synchronic portion of the thesis, I provide an analysis of cleft constructions in Malay and demonstrate that clefts and pseudoclefts share the same derivation to a certain point. This is made possible by the finding that the clausal constituent of both constructions – the cleft clause of a cleft and the clausal subject of the pseudocleft – is a bare CP, as opposed to it being a complex DP, as argued by scholars such as Kader (1981) and Cole and Hermon (2000). The clausal constituent exhibits non-DP properties such that it is not a syntactic island, it does not invert with the focus, it does not allow the postulated null NP to surface, etc.

The diachronic portion of the thesis involves an account of the historical development of copular clauses in Malay and the grammaticalisation of the copulas *ialah* and *adalah*. I examine copular clauses in different stages of the language and conclude that *ada-lah* and *ia-lah* emerged in the Classical Malay period (14th ~ 19th century) as combinations of dummy auxiliary *ada* and 3rd person *ia* with comment marker *lah* to host *lah* as a last resort to avoid a violation of the stray affix filter. They later grammaticalised into copulas towards the end of the Classical Malay period due to the change in the word order from VSO to SVO and the decline in presentational *ada-lah* in verbal clauses, topic marker *pun*, and 3rd person pronoun *ia*.

Finally, for the typological portion of the thesis, I conduct a crosslinguistic comparison of copular clauses in 40 Austronesian languages and investigate the aspects of a language that make it possible for copulas to emerge. The comparison yields the finding that syntactic alignment and word order, specifically accusative alignment and SVO order, are very conducive to copulas. Almost all of the languages that have overt copulas are accusatively aligned and verb-medial.

1.1 Background

1.1.1 Nonverbal Predication

A predicate is commonly referred to as the constituent that provides an assertion about the subject. It can be said to be the part that carries the propositional content of an utterance, which becomes complete once it combines with the appropriate argument(s) and is saturated, as in Frege (1892). As an assertion, a predicate can therefore be negated. To illustrate, *merokok* in the following example, discernible as a V(erb) by the active voice prefix *meŋ-*, combines with and is saturated by the argument *Zul* to yield a complete proposition, and it may be negated by *tidak*. As such, *merokok* is a verbal predicate that participates in verbal predication.

Malay[2] (Austronesian – Malayic)

(1) *Zul* {*meŋ-rokok* /*tidak* *meŋ-rokok*.} (Verbal)
Z. ACT-smoke NEG ACT-smoke
'Zul smokes/doesn't smoke.'

On the other hand, nonverbal predication, as its name suggests, is a relation that holds between a subject and a predicate of a syntactic category other than V, e.g. N(oun), A(djective) or P(reposition). As shown in each of the examples below, a nonverbal constituent is employed as a negatable predicate in the expression of an assertion. They are complete propositions despite the absence of a verb and appear to be nothing but juxtapositions of the subject and a (negated) nonverbal constituent.

[2] Throughout the thesis, all non-English examples are in Malay, unless explicitly indicated.

(2) a. *Zul* {*guru* /*bukan guru.* } (Nominal)
Z. teacher CNTR teacher
'Zul is/isn't a teacher.'

b. *Zul* {*kreatif* / *tidak kreatif.* } (Adjectival)
Z. creative NEG creative
'Zul is/isn't creative.'

c. *Zul* {*seperti bapa=nya* /*tidak seperti bapa=nya.* } (Prepositional)
Z. like father=3 NEG like father=3
'Zul is/isn't like his father.'

Nonverbal predicates are commonly accompanied by an element traditionally classified as a copula, such as *be* in "Zul is a teacher". In fact, copula *adalah* may optionally surface in the examples above between the subject and the nonverbal predicate, as also shown below:

(3) *Zul* (*adalah*) *kreatif.*
Z. COP creative
'Zul is creative.'

Among the 1200 or so languages in the very large Austronesian family, Malay is unusual in that nonverbal predication in the language makes use of overt copulas, as opposed to the majority of the Austronesian languages, in which a visible copula is absent. As illustrated in the comparison below, the nonverbal predicate may combine with an overt copula, albeit optional. Conversely, most of the other Austronesian languages, such as Malagasy, do not appear to utilise a copula in the same environment.

	Malay (Austronesian – Malayic)	Malagasy (Austronesian – Barito)
(4) a.	*Langit* (*adalah*) *biru.* sky COP blue 'The sky is blue.'	b. *Manga ny lanitra.* blue DET sky 'The sky is blue.'

Remarkably, Malay does not just employ one, but two copulas. Compared to the copula in the previous examples, the one in the example below has a different form. Yap (2007) is of the opinion that *adalah* originated from *ada-lah* (EXIST-FOC), whilst *ialah* originated from *ia-lah* (3.SG-FOC).

(5) *Guru besar sekolah saya* (*ialah*) *Zul.*
teacher big school 1.SG COP Z.
'The headmaster of my school is Zul.'

By virtue of there being a copula in these clauses, they are appropriately called copular clauses. One might then assume that clauses that involve nonverbal predication but do not include a visible copula, e.g. (2a-c), should not be called copular clauses. On the contrary, they should be treated no differently, especially considering that the copula is optional in examples (3)-(5). Therefore, I shall assume that copular clauses without a visible copula contain instead a zero copula, as illustrated below:[3]

(6) a. *Zul Ø guru.*
Z. COP teacher
'Zul is a teacher.'

b. *Zul Ø kreatif.*
Z. COP creative
'Zul is creative.'

c. *Zul Ø seperti bapa=nya.*
Z. COP like father=3
'Zul is like his father.'

From the brief exposé thus far, there arise several acute questions pertaining to nonverbal predication and copular clauses in Malay that extend across different levels of linguistic analysis. They concern the morphological properties of the two copulas that appear to be composed as *ada-lah* (EXIST-FOC) and *ia-lah* (3.SG-FOC), the syntax of copular clauses in Malay as mere juxtapositions or having more complex structure, the semantic rules that govern the use of overt vs. zero copulas, the historical development of the two copulas from existential verb *ada* and 3rd person *ia*, and the typological differences among the Austronesian languages with regard to copulas. By covering the breadth and depth spanned by these topics, this thesis endeavours to be a point of reference for studies in, about, and beyond nonverbal predication in Malay.

[3] This should not be confused with small clauses, which do not contain a copula, be it overt or null. Despite the term small *clause*, they are not clauses according to the widely accepted Pr(ed)P analysis: [$_{IP}$ *They consider* [$_{PrP}$ *John* [$_{Pr'}$ [$_{Pr}$ e][$_{AP}$ *crazy*]]]] (Bowers, 1993). Although small clauses do not constitute copular clauses, they are instances of nonverbal predication.

1.1.2 Copulas

Before anything, one must ask the fundamental question: what is a copula? Different definitions exist in the literature. Some of the most salient ones listed by Pustet (2003) identify copulas as the following:

- a linker between subject and predicate
- a syntactic 'hitching post' to which verbal inflectional categories can be attached
- a predicator which is added to lexemes that do not form predicates on their own

However, these definitions are not without problems. As argued by Pustet (2003), they raise more questions than they answer. First, the linker hypothesis would not be able to answer the question why a copula is still required in cases where there is no subject to link the predicate to. For instance, in the Irish sentence "tá daoine ann" (there are people), *tá* cannot be characterised as a linker as no subject is present for *tá* to link *daoine* (people) with. McCloskey (1996) argues that expletives, be they overt or null, do not exist in Irish and that *ann* is an existential predicate, hence no linking. Next, the "Dummy Hypothesis" (in reference to the copula as a dummy element for inflectional categories to attach to), as Stassen (1997) calls it, would not be able to account for the use of copulas in languages that lack verbal inflectional categories, as is the case of the copula *shi* in Mandarin. As an analytic language, Mandarin does not make use of any morphological inflections on the verb but expresses information related to tense, aspect, and mood (TAM) periphrastically. Finally, the predicator hypothesis would need to address the use of a predicator (or its absence) with both verbal and nonverbal predicates, as in the case of the *i* particle in Tok Pisin, as shown below. Verbs, being the quintessential predicate, should not require some additional morpheme to indicate that it is a predicate, according to the predicator hypothesis.

Tok Pisin (English Creole – Melanesian Pidgin)

(7) a. *Em i kam.* (Verbal)
3.SG PRED come
'S/he comes.'

b. *Madang i bik.* (Nonverbal)
M. PRED big
'Madang is big.'
(Smith, 2008)

Perhaps the three traditional accounts are more suitable for specific languages, not crosslinguistically. Understandably, it is difficult to be inclusive of copulas across languages as they take the guise of many different things. Copulas overlap with different grammatical categories such as verbs, pronouns, and others, so it is difficult to identify a common set of features that are universally shared by what can be considered copulas across different languages. According to Arche, Fábregas, and Marín (2019), the "copula is not a distinct grammatical category, but rather the label that has been given to a number of distinct objects in different languages" (p. 6). There also exist elements of other syntactic categories which are considered copulas. Stassen (2013) notes that, other than the typical verbal form, a copula can come in the form of a pronominal element (*pro-copula*) or a particle (*particle copula*).

Pustet (2003) does decide on the following definition for her crosslinguistic investigation: "A copula is a linguistic element which co-occurs with certain lexemes in certain languages when they function as predicate nucleus. A copula does not add any semantic content to the predicate phrase it is contained in". However, imposing semantic vacuousness on the copula would lead to undergeneration, as copulas in some well-studied cases are known to carry semantic content, albeit little. For instance, the Spanish copulas *ser* and *estar* make an aspectual distinction. In the following examples by Arche et al. (2019), the predicate is interpreted as a temporary state with the use of *estar* – the subject looks handsome at the time of utterance, e.g. after a haircut. On the other hand, it is interpreted as a permanent state with the use of *ser* – the subject is always handsome, e.g. has nice facial features.

Spanish (Indo-European – Romance)

(8) a. *Roberto Alcázar* *está* *guapo.* (Temporary)
R. COP^{estar} handsome
'Roberto Alcázar is handsome. (now)'

b. *Roberto Alcázar* *es* *guapo.* (Permanent)
R. COP^{ser} handsome
'Roberto Alcázar is handsome (always).'

Clearly, there is great difficulty in ascribing a definition to copulas across languages, especially because they have different functions in each language. In fact, Stassen (2013) does not provide a clear definition for copulas, but often refers to them

as auxiliaries or supportive items. Coupled with the fact that they do not form a homogenous group, defining them based on functions and grammatical properties whilst expecting to be crosslinguistically inclusive might prove futile. Thus, it is best to be agnostic to the forms and functions of a copula and, instead, focus on other aspects of their use, such as their syntactic distribution.

Therefore, I shall loosely delimit the notion of copula according to the following criteria:

- It accompanies a nonverbal predicate in a non-modifying manner.[4]
- It semantically differs when used with or as a verbal predicate, if possible.

The first criterion is purposely broad to allow a wide range of linguistic material (even modals, negators, etc. in certain languages), whilst the second criterion acts upon the first one implicationally to eliminate all extraneous material. Below is a flowchart of how a copula is identified according to the criteria:

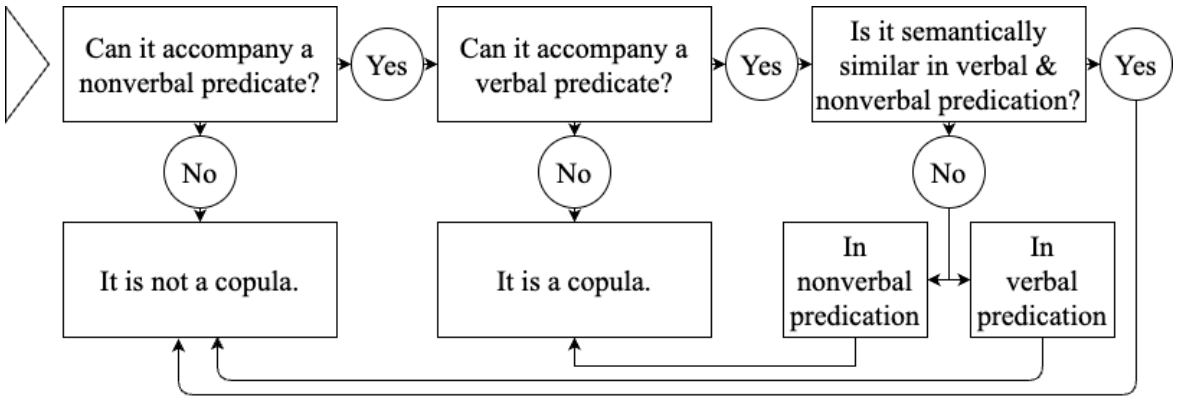


Figure 1: Flowchart for Identifying Copulas

The filter above is used throughout this thesis for the identification of copulas cross-linguistically (see Chapter 7 for a more detailed description of the methodology). As for the copulas in Malay, I shall return to them in Section 1.6 to make way for an overview of the grammar of Malay that is requisite for the identification of copulas and the analysis of *ialah* and *adalah* as copular auxiliaries.

[4] *Modifying* here refers to the addition of optional linguistic material such as adjectives, adverbs, etc. to provide non-essential meaning to the modificand. For example, *fast* in "he drives the fast car" and "he drives the car fast" is an optional modifier that provides non-essential meaning to the object "car" in the former and the action "drives the car" in the latter. It may be omitted from both sentences.

1.2 Research Questions

To what extent are the copulas predictable from their component parts and the way they are combined?

It is not clear whether the two reputed copulas are verbal, pronominal, or something entirely different. Putatively, each copula formed from the compounding of a different lexeme with focus marker *lah*: the root of *adalah* is believed to be existential verb *ada*, whereas the root of *ialah* is 3rd person *ia* (Yap, 2007). I demonstrate that the copulas do not transparently reflect the morphemes that they are believed to have evolved from. Specifically, *ialah* is no longer a focused pronoun, whilst *adalah* is no longer a focused existential verb. The inseparability of *lah* from *ia* and *ada* indicates that the two copulas are monomorphemic in their current form.

What is the syntactic structure of a copular clause in Malay?

Although it is obvious that the copulas intervene between the subject and its predicate, where precisely the copulas occur in the structure between those constituents – within the IP or VP layer – is not totally clear. Also wanting is the analysis of various syntactic phenomena observed of copular clauses, such as extraction and inversion. Based on word order facts, e.g. the possibility of the co-occurrence of *adalah* with semantically vacuous linking verbs and the position of the copulas relative to negators and auxiliaries, I conclude that the copulas head TP.

What governs overt encoding of the copulas in Malay?

There are many environments in which copulas are impossible, contra the widely held belief that they are always optional. Karim, Onn, Musa, and Mahmood (2014) claim that they are "restricted and actually not encouraged in the formation of Malay sentences" (p. 264), but there has not been a principled account of the exact restrictions on their use. To say that they are somehow not encouraged is a prescriptive and dismissive statement that is not at all helpful to understanding how they are restricted. I show that the copulas interact with the notion of aspect, which determines whether they can be encoded overtly. It is only in copular clauses that carry an atemporal interpretation, as opposed to a temporally bound interpretation, that overt encoding of is possible.

How are cleft constructions in Malay derived?

A Malay cleft does not resemble a copular clause as it has no copula. However, a pseudocleft is undoubtedly a copular clause. One can then ask whether a cleft is a copular construction and how these two constructions are syntactically related, given that one is a paraphrase of the other. Upon closer inspection, there is a phonologically null matrix copular clause above the visible part of the cleft. I argue that the matrix copular clause allows raising of the constituent introduced by *yang* in the cleft, i.e. the cleft clause, to its subject position to derive a pseudocleft.

When did the copulas emerge in Malay and how did they grammaticalise?

No attempt has been made to look into the diachrony of the copulas. Consequently, it is unknown when exactly the copulas emerged and what circumstances led to their emergence. Whilst no visible copula was used in Old Malay (7^{th} ~ 14^{th} century), what have now become copulas *adalah* and *ialah* had only emerged in the beginning of the Classical Malay period. Although *ia-lah* and *ada-lah* were used in copular clauses then, they functioned as an argument and a presentational auxiliary respectively. It was only towards the end of the Classical Malay period that their copular use became apparent, as the presentational use of *ada-lah* and the argument status of *ia-lah* declined. Topical constructions underwent changes to result in the topic to be reanalysed as the canonical subject, which promoted VSO to SVO change, subsequently allowing *ia-lah* and *ada-lah* to grammaticalise into copulas.

What factors condition the distribution of copulas across Austronesian?

Starosta, Pawley, and Reid (1982) state that no copulas were present in nonverbal clauses in Proto-Austronesian, which is why most modern languages across Austronesian lack overt copulas altogether. As for those that do have copulas, an investigation into how copulas have been able to emerge is wanting. In concord with the finding that the Malay copulas *ialah* and *adalah* developed in line with change in word order, the survey reveals that almost all the Austronesian languages with overt copulas have SVO word order and accusative alignment. Word order especially is therefore an important factor in the emergence of copulas in an Austronesian language.

1.3 Framework and Methodology

In this thesis, observations are made with regard to the phenomena related to them and hypotheses are formulated, which are then tested. Through this approach, an explanatory account of the unexplained phenomena related to copular clauses in the language may be achieved.

Although largely descriptive, especially concerning the morphology, diachrony, and typology of the copulas, this thesis utilises a fairly standard generative Chomskyan framework. The reason for this choice is that many linguistic phenomena require *formalisms* (Dixon, 1997) to allow the analysis of underlying structures. For example, the postulation of null expletives in Malay by Mustaffa (2020), based on the theory-specific Extended Projection Principle by Chomsky (1981), allows a principled account of subjectless existential clauses. A purely descriptive view of the phenomenon might simply posit *a posteriori* that there is no subject in such a construction, which, although appeals to Occam's razor, presents an exception to the rule that all clauses in Malay must have a subject, be it overt or null.

1.3.1 Synchrony

The synchronic portion of this thesis deals with copular clauses in the current state of Malay using a mixed method approach. It focuses on Standard Malay, as well as the spoken variety of Malay in and around the capital, Kuala Lumpur, also called *KL Malay* by Abu Bakar (2009) or *Cakap KL* (KL speech) by locals. It makes use of corpus analyses of Standard Malay using data gathered from authentic sources such as newspapers, as appropriately cited, but also benefits from grammaticality judgements by 11 native speakers of KL Malay. Those not cited are examples analogous to the ones cited or examples elicited from consultants who are native speakers.

Official Sources	Corpora
Online Malay news portals: https://www.utusan.com.my/ https://www.bharian.com.my/ https://www.astroawani.com/ https://www.hmetro.com.my/ https://www.mstar.com.my/	Corpus of the Malay language by DBP: http://sbmb.dbp.gov.my/korpusdbp/Login.aspx Malay Concordance Project by ANU: http://mcp.anu.edu.au/Q/mcp.html

Table 1: The Contemporary Sources Used

Considering the linguistic landscape of KL as the capital of Malaysia, Standard Malay, KL Malay, and many more varieties of Malay can be said to exist in a lectal continuum (Holm, 1988). KL Malay forms the mesolect between the basilectal varieties of Malay and Standard Malay, the acrolect, as suggested by Abu Bakar (2009). In the basilect, many features of spoken language such as simplification and omission are used by speakers, whereas in the acrolect, the language is richer with respect to grammatical features. Meanwhile, the mesolect would incorporate features of the lexifying standard language whilst retaining features of the substrate.

Figure 2: The Lectal Continuum in Kuala Lumpur

The inclusion of KL Malay provides valuable insights to the study of copular clauses in Standard Malay as the grammaticality judgements by native speakers of KL Malay reflect the findings from the corpus studies in Standard Malay, considering that KL Malay is the variety that most closely resembles Standard Malay in terms of grammar and lexis, in contrast to more divergent dialects such as Bazaar Malay, Kelantanese Malay, etc. (see Section 1.4.1 for details). However, caution must be exercised not to misconstrue the two varieties to be one and the same, as the former is a spoken vernacular that is natively acquired, whereas the latter is a codified language that is learnt in school. Standard Malay is characterised by pronunciation that is different from that of KL Malay, called *sebutan baku* (standard pronunciation), which is never used in spoken Malay in Malaysia (Asraf, 1988).

To contextualise the lectal continuum with regard to copular clauses, the basilect features simplification in copular clauses through the use of zero copulas, whilst the use of overt copulas is more frequent in Standard Malay, the acrolect. For example, a speaker of Bazaar Malay may move up the lectal continuum to accommodate their speech to speak KL Malay (e.g. with someone from a different or unknown linguistic background) by changing the word order of certain phrases or to approximate Standard Malay (e.g. in a formal context) by incorporating more features from the acrolect such as affixes and overt copulas, as shown below:

(9) a. *Itu budak banyak cakap dia punya kakak.* (Bazaar)
DIST child a.lot speak 3.SG own sister
'That child who talks a lot is his sister.'

b. *Budak banyak cakap itu kakak dia.* (KL)
child a.lot speak DIST sister 3.SG
'That child who talks a lot is his sister.'

c. *Kanak-kanak yang banyak ber-cakap itu adalah kakak=nya.* (Standard)
child-RED REL a.lot INTR-speak DISTCOP sister=3
'That child who talks a lot is his sister.'

In consultations with native speakers of KL Malay, I have ensured that the pronunciation was according to KL Malay, and not the *sebutan baku* of Standard Malay, so as to encourage use of KL Malay, albeit slightly more formal in register (as in the language in parliamentary debates) to promote use of overt copulas.

1.3.2 Diachrony

On the other hand, the diachronic portion of the thesis employs a corpus analysis of several epigraphic texts in Old Malay written on stone inscriptions and literary texts in Classical Malay. The texts of the Old Malay inscriptions are obtained from references such as Cœdès (1930), Boechari (1966), Suhadi (1983), Postma (1992), Kozok (2015), Griffiths (2018), and Griffiths (2020). Isolated examples from Classical Malay are drawn from the Malay Concordance Project by the Australian National University (Proudfoot, 1991). The Classical Malay texts are chosen based on genre and the number of words. The genre is Classical Malay prose, which commonly comes in the form of the *hikayat* (story). The texts from the 18th century being many is to allow a total number of words that is approximately proportionate to the total number of words of the texts in the other centuries, ≈ 140,000 words.[5]

[5] Unfortunately, there is only one text available from the 14th century, and it is a manual, rather than a form of prose, which does not match the genre of the other texts. Nonetheless, the language used in both genres is similar. There are only three *hikayat* in the 20th century, so they fall only slightly short of the 140,000-word count. Also, only the two texts from the 14th century can be used, and their word count is half of that of the texts in the other centuries. Although there are two more accessible on the MCP website – Hikayat Bayan Budiman and Hikayat Amir Hamzah – I have chosen not to include them due to doubts on the language used. The latter is noted on the website to have been edited freely for personal taste by the author of the 1987 edition by the publisher Dewan Bahasa dan Pustaka *Hikayat Amir Hamzah* 1987).

Old Malay		Classical Malay	
		Terengganu Inscription Stone	(14th century)
		Hikayat Raja Pasai	(14th century)
		Hikayat Muhammad Hanafiah	(14th century)
		Undang-Undang Melaka	(15th century)
		Hikayat Pandawa Lima	(16th century)
Kedukan Bukit inscription	(683 AD)	Hikayat Indera Putera	(16th century)
Talang Tuwo inscription	(684 AD)	Hikayat Sang Boma	(17th century)
Kota Kapur inscription	(686 AD)	Hikayat Banjar & Kota Waringin	(17th century)
Sojomerto inscription	(7th century)	Hikayat Sang Bima	(18th century)
Telaga Batu inscription	(700 AD)	Hikayat Patani	(18th century)
Mañjuśrīgṛha inscription	(793 AD)	Hikayat Putera Jaya Pati	(18th century)
Gandasuli inscription	(832 AD)	Hikayat Syah Mardan	(18th century)
Bukateja inscription	(840 AD)	Hikayat Nakhoda Muda	(18th century)
Laguna Copperplate inscription	(900 AD)	Hikayat Hasanuddin	(18th century)
Tanjung Tanah manuscript	(14th century)	Hikayat Pahang	(19th century)
		Hikayat Merong Mahawangsa	(19th century)
		Hikayat Merpati Mas & Perak	(19th century)
		Hikayat Johor Serta Pahang	(20th century)
		Hikayat Seri Kelantan	(20th century)
		Hikayat Kerajaan Sikka	(20th century)

Table 2: The Historical Texts Analysed

1.3.3 Typology

In the typological portion, the comparative method is employed in examining copular clauses in Malay, as well as genetically related languages, to identify possible proto-forms.[6] This method is also useful in gaining an understanding of the way in which Malay and similar languages diverged from the majority of Austronesian languages in terms of nonverbal predication to allow the emergence of the copulas.

1.4 Overview of the Malay Language

It is generally accepted by linguists that Malay originated from the western parts of Borneo and spread westwards to the Malay peninsula and the Indonesian archipelago (Adelaar, 2004). Of course, this occurred after the expansion of Proto-Malayo-Polynesian from the Philippines in the north, which ultimately originated from the Proto-Austronesian language spoken in what is now Taiwan.

[6] References for languages with overt copulas: Urak Lawoi' (Hogan & Pattemore, 1988); Moklen (Chantanakomes, 1980); Batak (Woollams, 2005); Rejang (McGinn, 1998); Daakaka (von Prince, 2012); Lelepa (Lacrampe, 2014); Bierebo (Budd, 2009); Erromangan (Crowley, 1998); Ambel (Arnold, 2018); Biak (van den Heuvel, 2006); Irarutu (Jackson, 2014); Puyuma (Teng, 2007); Dusun (Price, 2007); Cham (Blood, 1977); Jarai (Jensen, 2013); Tsat (Thurgood & Li, 2002).

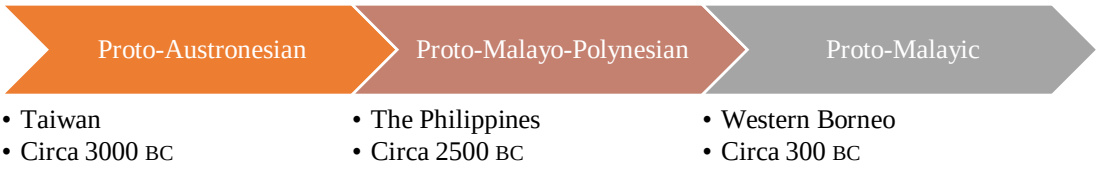


Figure 3: The Timeline of the Emergence of Proto-Malayic based on Bellwood (2007)

It is a member of the Malayic language group, which "consists of (literary, standard) Malay and all dialects and languages that are sufficiently close to Malay in order to form an exclusive subgroup with it within the Austronesian language family" (Adelaar, 1993, p. 566). On a broader typological scale, the subgroup of Austronesian languages that is most closely related to Malayic is Chamic. This is attested by the high degree of similarity between Old Malay and Old Cham, as documented in written form on several inscriptions, such as the Dong Yen Chau inscription (circa 4th century AD) written in Old Cham and the Talang Tuwo inscription (circa 7th century AD) written in Old Malay. According to Thurgood (1999), Malayic and Chamic form the Malayo-Chamic subgroup based on a set of lexical and phonological innovations. The innovations shared among these languages gave rise to the hypothesis by Adelaar (2005) that Chamic, Malayic, and the Balinese-Sasak-Sumbawa group form a larger group called Malayo-Sumbawan.

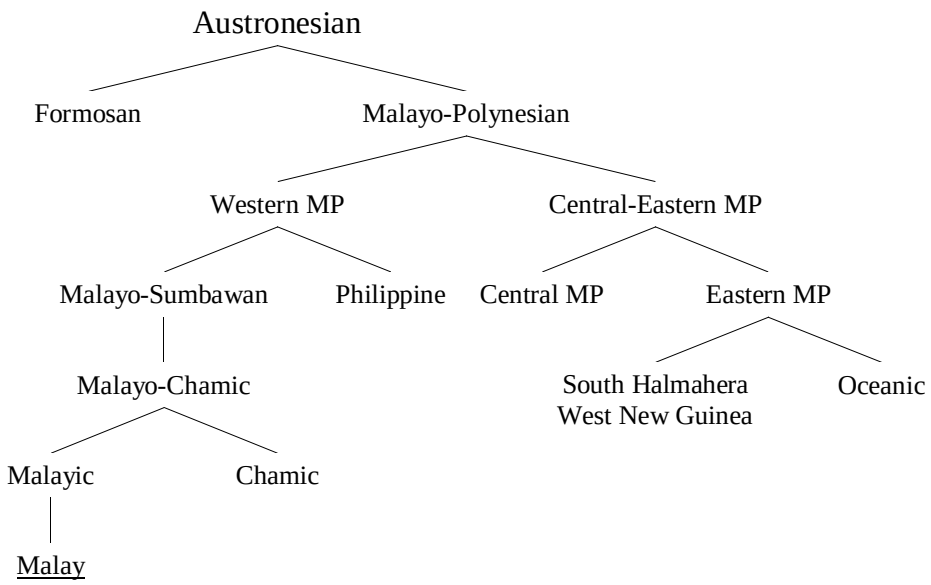


Figure 4: The Austronesian Language Tree

Malayic is the largest member of the Austronesian family by number of speakers. In fact, Indonesian (the standard variety of Malay spoken in Indonesia) alone is spoken by 199 million people, according to the 2022 list of most spoken languages by Ethnologue.[7] As it is such a large group of languages, Malayic spans across a vast geographical area, as shown in the shaded areas in the map below.

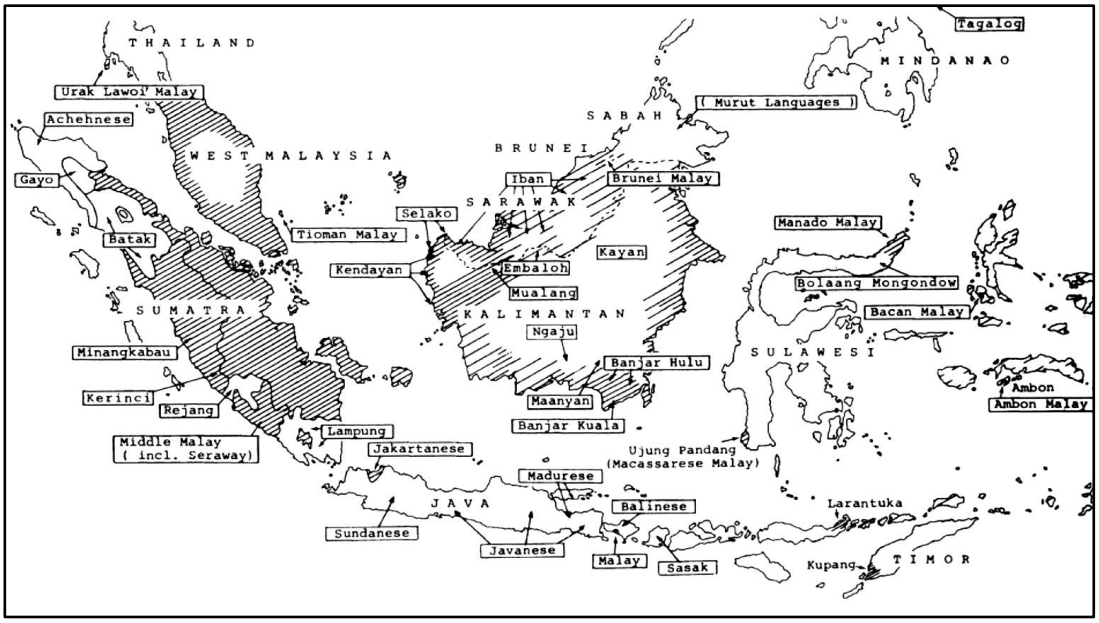


Figure 5: Map of Areas Where Malayic Isolects are Spoken (Adelaar, 1993)

1.4.1 Standard Malay vs. KL Malay

Malay is an official language in Malaysia, Indonesia, Brunei and Singapore, and although it is called Bahasa Malaysia, Bahasa Indonesia, and Bahasa Brunei in the former 3 countries, it is by and large the same language (Steinhauer, 2005). In fact, Malaysia, Indonesia, and Brunei have formed a collaborative effort to develop and codify the Malay language through an organisation called *Majlis Bahasa Brunei-Indonesia-Malaysia* (MABBIM) or the Language Council of Brunei-Indonesia-Malaysia, with Singapore as an observer. As such, the standard form can be used in all four countries without problems in intelligibility; however, it should be cautioned that there exist many dialects of the language within each country that have significant variation from the standard form.

[7] https://www.ethnologue.com/guides/ethnologue200

Certain discrepancies exist between the codified language and the vernaculars that are in common, informal use. Even more noticeable are the differences between Standard Malay and the different dialects that exist in each country, such as Jakartanese, Sarawakian Malay, Pattani Malay, etc. KL Malay, which this thesis is concerned with, can be characterised as a less standard version of the prescribed language, due to the prevalent use of contractions, omissions, slang, and informal vocabulary in general, but not colloquial enough to be what is known as Bazaar Malay, the low variety spoken as a pidgin amongst people of different races in Malaysia and Singapore (for a discussion on Bazaar Malay, see Zhiming & Aye, 2010).

Abu Bakar (2009) contends that *KL Malay* stands between Standard Malay and Bazaar Malay, the former being very constrained, whilst the latter very broad. Consider the contrast between Standard and KL Malay in (10), which shows that KL Malay exhibits simplification in several different respects: Pro-drop has applied as the subject is omitted; the string *tidak hendak* is contracted to *tak nak*; the intransitive verbal prefix *ber-* is omitted; the more colloquial forms *sebab* and *tengah* are used, instead of *kerana* and *sedang*. Nevertheless, use of the rich grammatical features that are characteristic of Standard Malay in KL Malay speech is normal, especially in formal settings; the higher the register, the closer it is to the standard variety.

(10) a. *Saya tidak hendak ber-lari kerana saya sedang sakit.* (Standard)
1.SG NEG want INTR-run because 1.SG PROG ill
'I do not want to run because I am ill.'

b. *Ø Tak nak lari sebab Ø tengah sakit.* (KL)
Pro NEG want run cause Pro middle ill
'I do not want to run because I am ill.'

Furthermore, various grammatical markers are less frequently used in KL Malay, but omission is optional. In (11b), each word appears in its bare form.

(11) a. *Siapa-kah yang akan meŋ-jaga pe-makan-an dia?* (Standard)
who-Q COMP PROS ACT-take.care NMZ-eat-NMZ 3.SG
'Who will take care of his diet?'

b. *Siapa yang akan jaga makan dia?* (KL)
who COMP PROS take.care eat 3.SG
'Who will take care of his diet?'

With regard to the main subject of this thesis, the use of *ialah* and *adalah* is most prominent in formal registers such as in formal writing and speech. In KL Malay, they are largely omitted wherever possible, but omission is optional; no ungrammaticality arises when an overt copula is used in KL Malay where it is expected in the same environment in Standard Malay. To illustrate, the examples below could be used in both Standard Malay and KL Malay. The only difference is the likeliness of the copulas to be omitted in KL Malay.

(12) a. *Penulis buku itu* (*ialah*) *Zul.*
writer book DIST COP Z.
'The writer of that book is Zul.'

b. *Penulis buku itu* (*adalah*) *seorang yang jujur.*
writer book DIST COP someone LIG honest
'The writer of that book is someone honest.'

1.4.2 Malay and Austronesian Alignment

Austronesian alignment is a syntactic alignment characterised by a multiple-voice system that encodes different constituents as what Austronesianists interchangeably call *Trigger*, *Pivot*, *Subject*, *Topic*, *Focus*, etc. (henceforth *Trigger*), depending on what can be loosely construed as the "topic" of the clause or what the clause is "about" in information-structural terms. The semantic role of the trigger is reflected by different affixes on the verb. To illustrate, consider the following examples by Aldridge (2007) in which the *ang*-marked constituent is the trigger and the marking on the verb changes according to the semantic role of the trigger.

Tagalog (Austronesian – Philippine)

(13) a. *B‹**um**›ili ang babae ng isda.* (Agent)
‹INTR.PRF›buy ANG woman OBL fish
'The woman bought a fish.'

b. *B‹**in**›ili ng babae ang isda.* (Patient)
‹TR.PRF›buy ERG woman ANG fish
'The woman bought the fish.'

c. *B‹in›ilh-**an*** *ng babae ng isda ang tindahan=ko.* (Locative)
‹TR.PRF›buy-APPL ERG woman OBL fish ANG store=1.SG.GEN
'The woman bought a/the fish <u>at my store</u>.'

d. ***I**-b‹in›ili* *ng babae ng isda* ang *lalaki*. (Benefactive)
APPL-‹TR.PRF›buy ERG woman OBL fish ANG man
'The woman bought the fish <u>for the man</u>.'

A mismatch between the trigger and the voice affix causes ungrammaticality. For instance, the following example is ungrammatical because the verb is marked with the actor-voice affix, but the *ang*-marked constituent corresponds to the patient, which results in a voice-trigger mismatch.

(14) **B‹um›ili* *ng babae ang isda.* (Actor Voice – Patient Trigger)
‹INTR.PRF›buy ERG woman ANG fish
(The woman bought <u>a fish</u>.)

Languages that strongly obey Austronesian alignment are categorised as Philippine-type languages, as many of the languages of the Philippines have inherited this alignment from Proto-Austronesian, and are thus argued by Blust (2013) to be conservative. These languages show an ergative-absolutive alignment such that the object of a transitive verb patterns with the subject of an intransitive verb. This pattern is taken to be the basic difference between absolutive-ergative languages and nominative-accusative languages, whose agent of a transitive verb it is that patterns with the subject of an intransitive verb, e.g. via nominative case marking on the two arguments. To illustrate, the patient of the transitive clause patterns with the subject of the intransitive clause in the following examples by being marked by the same *ang* marker, which is believed to mark absolutive case:

(15) a. *Na-matay ang babae.* (Intransitive)
PRF-die ANG woman
'The woman died.'

b. *B‹in›ili* *ng babae ang isda.* (Transitive)
‹TR.PRF›buy ERG woman ANG fish
'The woman bought <u>the fish</u>.'

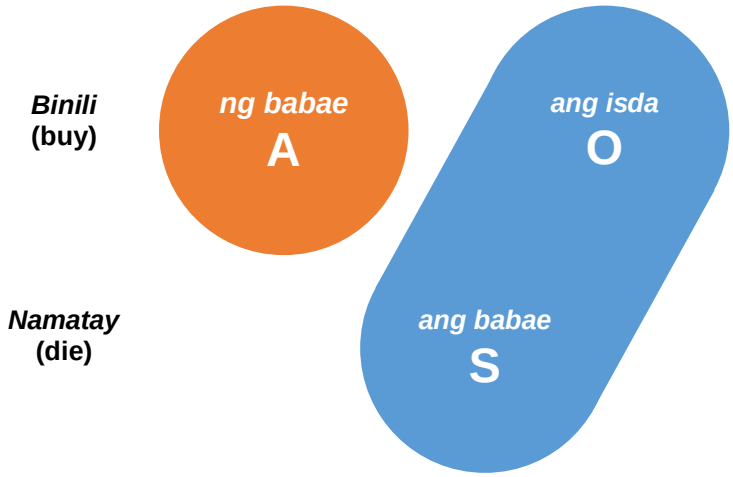


Figure 6: The Ergative-Absolutive Alignment in Tagalog

On the other hand, languages in which Austronesian alignment has significantly diminished are called Indonesian-type languages. These languages are most recognised by their true passive construction, which resembles the passive voice of nominative-accusative languages like English. Malay is one such language as it has drifted from Austronesian alignment to a certain extent. Although it is considered a nominative-accusative language, as exemplified by the true passive construction as in (16b), it retains the patient voice as in (16c), which is a vestige of Austronesian alignment. This makes the number of voice alternations in Malay three in total: the active voice, the passive voice, and the patient voice.

Malay (Austronesian – Malayic)

(16) a. *Dia telah **meŋ**-beli ikan itu.* (Active Voice)
3.SG PRF ACT-buy fish DIST
'He has bought the fish.'

b. *Ikan itu telah **di**-beli (oleh=nya.)* (Passive Voice)
fish DIST PRF PASS-buy by=3
'The fish has been bought by him.'

c. *Ikan itu telah dia beli.* (Patient Voice)
fish DIST PRF 3.SG buy
'He has bought the fish.'

The active voice in Malay is signalled by the nasal *meŋ-* prefix on the verb, whereas the passive voice is signalled by the *di-* prefix.[8] The patient voice, or what is called *pasif semu* (*semu* passive) in traditional Malay grammars, is bare in form. This instantiation of voice parallels the patient voice in Philippine-type languages in that both agent and patient are arguments of the verb, as opposed to the passive voice, in which the subject is demoted to an optional adjunct *oleh*-phrase (*by*-phrase).

The patient voice in Malay is argued by Aldridge (2007) to be a remnant of an earlier stage of Malay in which the syntax was ergative. The active-passive distinction is regarded to be a sign that Malay has evolved from ergative-absolutive to nominative-accusative and is a defining characteristic of Indonesian-type languages. However, the retention of the patient voice is said to be evidence that the language has not completed the transition, as is also believed of Malagasy.

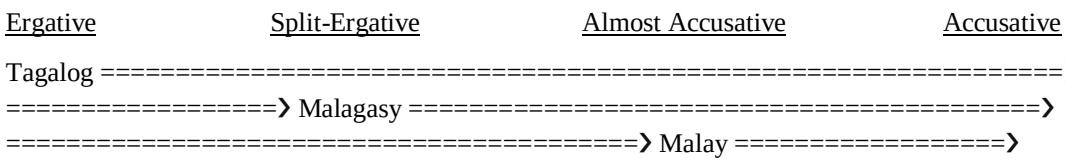


Figure 7: Languages Evolving from Ergative to Accusative (Aldridge, 2007)[9]

Malay is said to be an "almost accusative" language. The agent of a transitive verb and the subject of an intransitive verb consistently occur in preverbal position, classifying the language as nominative-accusative, as shown below. Also, because Modern Malay is morphologically caseless, it relies on word order to differentiate subject from object, in addition to the verbal voice affixes. This is made even more apparent in the case of utterances in which the affixes are dropped, as is characteristic of spoken Malay.

---

[8] The nasal *meŋ-* prefix undergoes assimilation with the following phoneme.

[+Nasal] → [αPlace] / __ [αPlace]

meŋ-karang (to compose)	→	*mengarang*
meŋ-basuh (to wash)	→	*membasuh*
meŋ-sapu (to sweep)	→	*menyapu*
meŋ-tarik (to pull)	→	*menarik*

[9] Aldridge (2007) originally describes Indonesian as almost accusative in the cline.

(17) a. *Dia mati.* (Intransitive)
3.SG die
'He died.'

b. *Dia meŋ-beli ikan itu.* (Transitive)
3.SG ACT-buy fish DIST
'He bought the fish.'

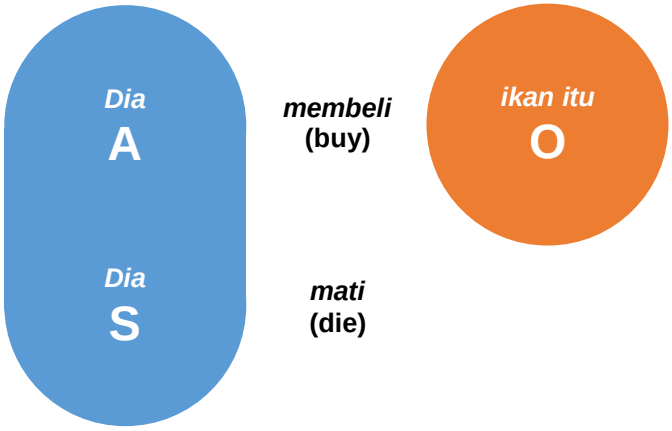


Figure 8: The Nominative-Accusative Alignment in Malay

To sum, Malay is a nominative-accusative language that has retained a vestige of Austronesian alignment. The patient voice occurs alongside the active and passive voices and bears witness to the Austronesian roots of the Malay language.

1.5 Basic Properties of Malay Grammar

This section provides a sketch of the grammar of the language. It is intended to only include parts of the grammar that are relevant to the thesis.

1.5.1 General Morphology

Generally, Malay can be described as an analytic language with a grammar that makes little use (if at all) of inflectional morphology such as case and agreement to distinguish grammatical relations. As in the table below, neither pronouns nor full DPs inflect for case. Therefore, word order is the main means of distinguishing subject from object.

Subject	Object
Saya/awak/dia/etc. *cakar kucing itu.* 1.SG/2.SG/3.SG/etc. scratch cat DIST 'I/you/he scratched that cat.'	*Kucing itu cakar saya/awak/dia*/etc. cat DIST scratch 1.SG/2.SG/3.SG/etc. 'That cat scratched me/you/him.'

Table 3: The Absence of Case in the Nominal Category

The verb does not exhibit any φ-feature agreement with its arguments:

	1st Person	2nd Person	3rd Person
Singular	*Saya baca buku.* 1.SG read book 'I read a book.'	*Awak baca buku.* 2.SG read book 'You read a book.'	*Dia baca buku.* 3.SG read book 'He/she/it reads a book.'
Plural	*Kita/kami baca buku.* 1.PL.INCL/EXCLread book 'We read a book.'	*Awak baca buku.* 2.PL read book 'You read a book.'	*Mereka baca buku.* 3.PL read book 'They read a book.'

Table 4: The Absence of Subject-Verb Agreement

Although Malay is an analytic language, it also makes use of derivational morphology. Morphemes are affixed to a root to form complex words. To illustrate, the following examples are complex one-word sentences whose root is host to various other morphemes. Furthermore, the roots themselves are complex: *tanggungjawab* (responsibility) is a compound *tanggung* and *jawab*, whilst *kotak-katik* is reduplicated.

(18) a. *Ku=di-per-tanggung-jawab-kan=nya-lah.*
1=PASS-CAUS-bear-answer-APPL=3-FOC
'I was made responsible for it.'

b. *Ku=meŋ-per-kotak-katik-kan=nya-lah.*
1=ACT-CAUS-chaos-RED-APPL=3-FOC
'I made it chaotic.'

Derivational morphology is also used to change the syntactic category of a root. In example (19a), the noun *ketidakberperikemanusiaan* (inhumanity) undergoes nominalisation twice to yield its final form. It is derived from the root *manusia* (human) via twice the application of the nominalising circumfix *ke-an*, but with deletion of the last *-an* suffix as a result of contiguous repetition. In example (19b), the root *umbang* is reduplicated for iterative aspect before being affixed by the involuntary or accidental verbal *ter-* prefix and finally nominalised by the *ke-an* circumfix.

(19) a. *Ke-tidak-ber-peri-ke-manusia-an-~~an~~=nya-lah yang saya benci.*
NMZ-NEG-INTR-about-NMZ-human-NMZ-NMZ=3-FOC COMP 1.SG hate
'It is his inhumanity that I hate.'

b. *Ke-ter-umbang-ambing-an kapal=nya meŋ-buat=nya mabuk.*
NMZ-NVOL-rock-RED-NMZ ship=3 ACT-make=3 sick
'The rocking of his ship makes him sick.'

Inflectional morphology plays a minimal role, whilst derivational morphology has a more secure footing in the language.

1.5.2 Word Order and the Preverbal Subject

Malay has a basic SVO word order, as opposed to the majority of the languages in Austronesian, which are predominantly verb-initial. The use of non-canonical word orders, such as VSO and OSV, can be attributed to different pragmatic or information-structural functions, such as focalisation, topicalisation, interrogative formation, etc.

(20) a. *Ali telah habis-kan novel itu.* (Basic – SVO)
A. PRF finish-APPL novel DIST
'Ali has finished that novel.'

b. *Novel itu, Ali telah habis-kan.* (Pragmatic – OSV)
novel DIST A. PRF finish-APPL
'That novel, Ali has finished.'

According to Mustaffa (2020), the subject position of a clause in Malay must always be occupied, as Malay obeys the Extended Projection Principle (EPP) by Chomsky (1981). Considering that subjects in Malay do not exhibit agreement, there should be no motivation for the subject to move to the preverbal position, other than the EPP. Consider the following examples:

(21) a. *Ø di-saran-kan bahawa kita kerap meŋ-cuci tangan.*
EXPL PASS-encourage-APPL COMP 1.PL.INCL often ACT-wash hand
'It is encouraged that we often wash our hands.'

b. *Kita di-saran-kan untuk kerap meŋ-cuci tangan.*
1.PL.INCL PASS-encourage-APPL for often ACT-wash hand
'We are encouraged to often wash our hands.'

c. **Kita di-saran-kan bahawa kerap meŋ-cuci tangan.*
1.PL.INCL PASS-encourage-APPL COMP often ACT-wash hand
(We are encouraged that often wash our hands.)

In example (21a), a null expletive is merged in the matrix clause as no DP is able to move to the subject position, since the subordinate subject *kita* has been satisfied of all its featural needs within the subordinate clause. In example (21b), the same DP is allowed to raise to the matrix clause as it cannot receive case within the non-finite subordinate clause (see Section 1.5.4 for finiteness in Malay). In example (21c), raising of the subject despite having its features checked within the finite subordinate clause is illicit. Therefore, it is analysed that finite T^0 possesses a strong EPP feature that triggers movement of the closest constituent that can be identified as a subject to SpecTP.

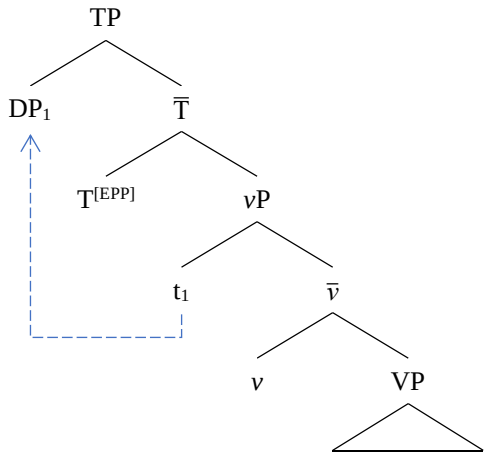


Figure 9: Movement of the Subject to SpecTP

This movement contrasts with several verb-initial Austronesian languages, the analysis of some of which involves T^0 having some feature that triggers movement of the predicate to SpecTP. For example, Cole and Hermon (2008) argue that, for the closely related Toba Batak language, the VP moves to SpecTP and the subject moves to a clause-peripheral topic position. However, this analysis is not suitable for Malay clause structure, as the subject remains in preverbal position in environments in which topicalisation is not possible, such as relative clauses. To illustrate, in example (22a), the beneficiary PP may not be preposed to the front of the relative clause, which shows that the topic position is not available for topics to move into. Given that movement to the left periphery of a relative clause is not permitted, it can be concluded that the subject of the relative clause in example (22b) occupies SpecTP.

(22) a. **Novel yang* [*untuk Zul*]$_1$ *dia telah beli* t$_1$ *belum di-baca.*
novel REL for Z. 3.SG PRF buy IMPRF PASS-read
(The novel that for Zul he has bought has not been read.)

b. *Novel yang dia telah beli untuk Zul belum di-baca.*
novel REL 3.SG PRF buy for Z. IMPRF PASS-read
'The novel that he has bought for Zul has not been read.'

Chung (1976) argues that the subject in the patient voice may undergo long-distance passivisation to a matrix passive clause, as shown in example (23). This finding demonstrates that movement of the patient in the subordinate patient-voice clause constitutes A-movement, which can feed passivisation in the matrix clause. The patient moves to the subordinate TP to satisfy the EPP feature on the subordinate T^0 head before raising to the matrix TP for the same reason.

(23) *Buku ini di-anggap oleh mereka sudah saya baca.*
book PROX PASS-believe by 3.PL already 1.SG read
'This book is believed by them to have been read by me.'
(Chung, 1976, p. 65)

Embedding the clause in a relative clause confirms the position of the subject as SpecTP, given that topicalisation is banned in relative clauses in Malay:

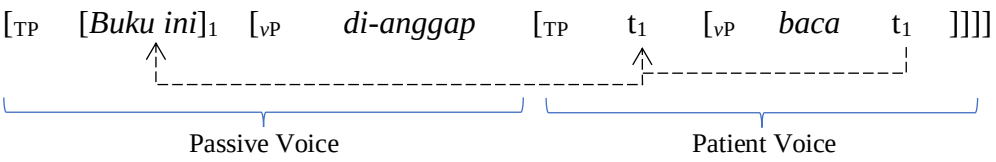
(24) *Waktu buku ini di-anggap oleh mereka sudah saya baca telah ber-lalu.*
time book PROX PASS-believe by 3.PL already 1.SG read PRF INTR-pass
'The time this book was believed by them to have been read by me has passed.'

Also, the patient voice is signalled by the position of auxiliaries between the patient and the agent, as shown below in (25a). The grammaticality of (25b), which illustrates relativisation of the patient and the resulting position of the auxiliary preceding the agent, further indicates that the word order in the patient voice is not obtained via Ā-movement.

(25) a. *Buku ini telah saya baca.*
book PROX PRF 1.SG read
'I have read this book.'

b. *Buku yang telah saya baca…*
book REL PRF 1.SG read
'The book that I have read…'

Therefore, the patient in the patient voice in Malay can be analysed to occupy the highest A-position, which is SpecTP, whereas the agent can be said to remain in Spec*v*P.



The subject in Malay is concluded to move from where it is merged (Spec*v*P in the active voice and CompVP in the passive and patient voices, respectively) to SpecTP, as shown in Figure 9.

1.5.3 Syntactic Categories

Malay is known to be a language whose syntactic categories are difficult to identify, owing to the fact that the same lexical roots may be used in different distributions. For example, *marah* could be used in a possessive phrase, in combination with a degree expression, and with an object transitively, corresponding to a nominal, adjectival, and verbal predicate to mean "anger", "angry", or "to scold" respectively.

(26) a. *Marah dia belum reda.* (Nominal)
MARAH 3.SG IMPRF subside
'His anger has not subsided.'

b. *Dia sangat marah.* (Adjectival)
3.SG very MARAH
'He is very angry.'

c. *Dia marah saya.* (Verbal)
3.SG MARAH 1.SG
'He scolded me.'

However, it is possible to demonstrate a distinction between verbs and non-verbs in Malay using procliticisation – it is only possible for proclitics to attach to verbs. In examples (90), repeated below, the 1st person proclitic may attach to the dynamic verb *lari* (run) and the stative verb *tahu* (know) but not to *guru* (teacher), indicating that it is not a verb.

(27) a. *Ku=lari.*
1=run
'I run.'

b. *Ku=tahu.*
1=know
'I know.'

c. **Ku=guru.*
1=teacher
(I am a teacher.)

Jadi (become) may be used to test if a word is verbal or nonverbal. The passive causative form of *jadi* cannot select a verbal predicate, as illustrated below:[10]

(28) a. **Ali di-jadi-kan lari.* (Dynamic Verbal)
A. PASS-become-APPL run
(Ali was made to run.)

b. **Ali di-jadi-kan tahu.* (Stative Verbal)
A. PASS-become-APPL know
(Ali was made to know.)

c. *Ali di-jadi-kan guru.* (Nonverbal)
A. PASS-become-APPL teacher
'Ali was made a teacher.'

Depending on the (non)verbal status of a word that combines with *jadi*, the interpretation changes. In combination with a dynamic verbal constituent, *jadi* gains an agentive interpretation that implies deliberate execution of an action. On the other hand, when it combines with a nonverbal constituent, it encodes inchoative aspect and induces a change-of-state interpretation. For example, *lari* can only be verbal as it only

[10] However, when in the active, *jadikan* does allow stative verbs as well as nonverbal predicates:

(i) a. **Pengalaman itu meŋ-jadi-kan Ali lari.*
experience DIST ACT-become-APPL A. run
(That experience made Ali run.)

b. *Pengalaman itu meŋ-jadi-kan Ali tahu.*
experience DIST ACT-become-APPL A. know
'That experience made Ali know.'

c. *Pengalaman itu meŋ-jadi-kan Ali guru.*
experience DIST ACT-become-APPL A. teacher
'That experience made Ali a teacher.'

caries an agentive interpretation when combined with *jadi*. Meanwhile, *guru* (teacher) can only be nonverbal, considering that it only receives a change-of-state interpretation with *jadi*. Stative verbs receive an inchoative interpretation, whilst ambiguous words like *marah* could either be verbal or nonverbal. Therefore, when it combines with *jadi*, the meaning is ambiguous between an inchoative and agentive interpretation.

(29) a. *Dia jadi lari.* (Agentive)
3.SG become run
'He (proceeded/decided to) run.'

b. *Dia jadi tahu.* (Inchoative)
3.SG become know
'He became aware.'

c. *Dia jadi guru.* (Inchoative)
3.SG become teacher
'He became a teacher.'

d. *Dia jadi marah.*
3.SG become MARAH
i. 'He became angry.' (Inchoative)
ii. 'He proceeded/decided to rant.' (Agentive)

Although (29a-d) do not tell us much about the (non)verbal status of a predicate, the non-volitional prefix *ter-* disallows verbal predicates. It also dispels the agentive interpretation associated with the verbal form of the ambiguous *marah*.

(30) a. **Dia ter-jadi lari.* (Verbal)
3.SG NVOL-become run
(He accidentally proceeded/decided to run.)

b. **Dia ter-jadi tahu.* (Verbal)
3.SG NVOL-become know
'He accidentally became aware.'

c. *Dia ter-jadi guru.* (Nonverbal)
3.SG NVOL-become teacher
'He accidentally became a teacher.'

d. *Dia ter-jadi marah.*
3.SG NVOL-become MARAH
i. 'He accidentally became angry.' (Nonverbal)
ii. #(He accidentally proceeded/decided to rant.) (Verbal)

In addition to that, the imperative can be used as a diagnostic to test whether a word constitutes the verbal class. Dynamic imperatives are ungrammatical with *jadi*. In addition, stative verbs cannot be used in the imperative, regardless of the use of *jadi*.[11] Conversely, if the use of inchoative verb *jadi* is obligatory, it is nonverbal. The apparent optionality of *jadi* with ambiguous words like *marah* is a reflection of its dual verbal/nonverbal status. These results are illustrated below:

(31) a. [(**Jadi*) *lari* *sekarang!*] (Dynamic Verbal)
become run now
'Run now!'

b. *[(*Jadi*) *tahu* *sekarang!*] (Stative Verbal)
become know now
(Know now!)

c. [*(*Jadi*) *guru* *sekarang!*] (Nonverbal)
become teacher now
'Become a teacher now!'

d. [(*Jadi*) *marah* *sekarang!*]
become MARAH now
i. 'Become angry now!' (Nonverbal)
ii. 'Rant now!' (Verbal)

Furthermore, word classes can often be distinguished via the affixation of derivational morphology, as done by Tjia (2015). For instance, *marah* in example (26a) may be modified by the nominalising circumfix *ke–an* to form the noun *ke-marah-an* (NMZ-angry-NMZ). The nominal form of the word may no longer be used in the same environments in examples (26b-c), as shown below:

(32) a. *Ke-marah-an* *dia* *belum* *reda.*
NMZ-MARAH-NMZ 3.SG IMPRF subside
'His anger has not subsided.'

[11] Obviously, there are many utterances that may be rendered ungrammatical in the imperative, such as prepositional phrases, adverbials, etc.; however, for the purpose of telling verbs, adjectives, and nouns apart, the imperative is sufficient. Following the application of the imperative diagnostic and an ungrammatical result, simply reasoning based on the semantics of a word encoding a state, as opposed to a spatiotemporal location, manner of action, etc. should be enough to tease stative verbs apart from prepositions, adverbs, and other categories.

b. **Dia sangat ke-marah-an.*
3.SG very NMZ-MARAH-NMZ
(He is very angry.)

c. **Dia ke-marah-an saya.*
3.SG NMZ-MARAH-NMZ 1.SG
(He scolded me.)

Similarly, it may be derived into a verb via affixation by the applicative *-i* suffix and the active *meŋ-* prefix, no longer allowing the derived verb to occur in the same environments as (26a-b). The ungrammaticality of *marah* in (32b) and (33b) as a noun and a verb suggests a separate class of predicate. Given that it is intransitive, it can only be adjectival (more on the verb-adjective distinction in Section 4.6.1).

(33) a. **Meŋ-marah-i dia belum reda.*
ACT-MARAH-APPL 3.SG IMPRF subside
(His anger has not subsided.)

b. **Dia sangat meŋ-marah-i.*
3.SG very ACT-MARAH-APPL
(He is very angry.)

c. *Dia meŋ-marah-i saya.*
3.SG ACT-MARAH-APPL 1.SG
'He scolded me.'

As verbal predicates, *lari* and *tahu* therefore cannot combine with copulas. When used in verbal clauses, *adalah* corresponds to a focal perfective aspectual auxiliary, which is separate from the non-aspectual copula, as shown below:

(34) a. *Lelaki itu ada-lah lari.*
man DIST AUX-FOC run
#(That man runs.)
'That man HAS run.'

b. *Lelaki itu ada-lah tahu.*
man DIST AUX-FOC know
#(That man knows.)
'That man HAS known.'

c. *Lelaki itu adalah guru.*
man DIST COP teacher
'That man is a teacher.'
#(That man HAS been a teacher.)

1.5.4 Tense, Aspect, and Mood

In keeping with the analytic nature of the Malay language, tense, aspect, and mood are not morphologically marked on the verb. Rather, the language relies on lexical resources to convey such information, such as auxiliary verbs and adverbials. Information pertaining to tense and time reference is mainly conveyed using temporal adverbials, wherever appropriate. Otherwise, it is inferred from context, without which the time reference of a clause would be ambiguous. For instance, the tense in the isolated example below is ambiguous between the present, past or future tense.

(35) *Saya meŋ-lukis komik.*
1.SG ACT-draw comic
'I draw/drew/will draw comics.'

In order to be explicit, either a context needs to be provided or some temporal adverbial needs to be used. In examples (36a-b), the first clause sets the context for the interpretation of the tense in the following clause, whereas in examples (36c-d), an adverbial is used.

(36) a. *Hobi saya satu sahaja. Saya meŋ-lukis komik.*
hobby 1.SG one only 1.SG ACT-draw comic
'My hobby is just one thing. I draw comics.'

b. *Saya sangat kreatif dahulu. Saya meŋ-lukis komik.*
1.SG very creative previously 1.SG ACT-draw comic
'I was very creative previously. I drew comics.'

c. *Saya meŋ-lukis komik tadi.*
1.SG ACT-draw comic just.now
'I drew comics just now.'

d. *Nanti saya meŋ-lukis komik.*
later 1.SG ACT-draw comic
'Later I will draw comics.'

The auxiliaries in Malay include aspectual markers, deontic modals, and epistemic modals (Abdullah, 1993; Karim et al., 2014; Kroeger, 2014; Salleh, 1993). They are used to convey information regarding aspect and modality, and they occur in the preverbal position – also known as the inflectional layer by scholars such as Rizzi (1997), in contrast to the verbal/lexical layer, where the lexical verb resides, and the complementiser layer, where complementisers reside.

Aspectual	Deontic	Epistemic
Sedang (progressive)	*Mesti* (necessity)	*Mesti* (certainty)
Telah (perfective)	*Boleh* (possibility)	*Mungkin* (probability)
Belum (imperfective)	*Dapat* (ability)	
Akan (prospective)	*Patut* (advisability)	
Pernah (experiential)		

Table 5: The Auxiliaries in Malay[12]

Auxiliaries may be stacked. Based on the example below, the order of co-occurring auxiliaries in Malay is epistemic-aspectual-deontic.

(37) *Dia mungkin belum boleh lukis komik dengan baik.*
3.SG might IMPRF can draw comic with well
'He might not yet be able to draw comics well.'

In addition to that, epistemic modals may be separated from the other auxiliaries by the negator *tidak*, making them the head of the highest projection in the inflectional layer, TP:

(38) *Dia* (*mungkin*/ *mesti*) *tidak akan boleh lukis komik lagi.*
3.SG might must NEG PROS can draw comic again
'He might/must no longer be able to draw comics again.'

There also exists an ambiguous auxiliary that serves a range of purposes. In terms of function, auxiliary *ada* is used in several ways that are similar to English *do*. For example, it is used in signalling an eventive interpretation of the predicate (as opposed to a habitual interpretation), in asserting the truth-value of a proposition, in hosting interrogative marker *kah* in polar questions, and in negation.

[12] It must be noted that there are certain items that occur between the subject and the verb but appear to be ambiguous. These items can be used as auxiliaries, adjectives, adverbs, normal verbs and/or raising predicates: *tetap* (still), *masih* (still), *perlu* (must), *wajib* (must), *harus* (must), *haram* (forbidden), *tentu* (definitely), *pasti* (definitely). Also, although *sudah* is commonly construed as a perfective marker, it can occur sentence-finally, which is not characteristic of auxiliary verbs in Malay. It is more appropriately identified as an adverb.

(ii) *Dia mati sudah.*
3.SG die already
'He died already.'

(39) a. *Ali ada baca novel.* (Eventive Aspect)
A. AUX read novel
'Ali read a novel.'

b. *Ali ada-lah baca novel.* (Assertion)
A. AUX-FOC read novel
'Ali DID read a novel.'

c. *Ada-kah Ali baca novel?* (Interrogative Formation)
AUX-Q A. read novel
'Did Ali read a novel?'

d. *Ali ti-ada baca novel.* (Negation)
A. NEG-AUX read novel
'Ali did not read a novel.'

Consider the following examples:

(40) a. **-Kah dia habis-kan novel itu?*
-Q 3.SG finish-APPL novel DIST
(Did he finish the novel?)

b. *Boleh$_1$-kah dia* t_1 *habis-kan novel itu?*
can-Q 3.SG finish-APPL novel DIST
'Can he finish the novel?'

c. *Ada-kah dia habis-kan novel itu?*
AUX-Q 3.SG finish-APPL novel DIST
'Did he finish the novel?'

During the formation of a yes/no question, interrogative marker *kah* is merged in CP. However, when there is no auxiliary available or nothing to move to CP, e.g. (40a), *kah* would be left without a host. Considering that *kah* is a clitic-like bound morpheme, this scenario would violate the stray affix filter, if no repair is made. Fortunately, *ada* can be merged as a last resort to rescue the stray interrogative marker and act as its host. Mustaffa (2018) calls the use of *ada* in such a support role *ada*-support, analogous to *do*-support in English.

Although auxiliaries are merged in the inflectional layer, *ada* can be merged either in the TP or directly in the CP. When merged in a string of auxiliary verbs in TP, *ada* must occupy the lowest position, as shown in (41a). It may pick up auxiliaries above it via head movement on its way to C^0, as in (41b).

(41) a. *Ali mesti ada baca novel.*
A. must AUX read novel
'Ali must have read a novel.'

b. *Mesti ada-kah Ali baca novel?*
must AUX-Q A. read novel
'Must've Ali read a novel?'

Based on (42), it is possible for *ada* to be merged in CP. If it were merged in TP, its skipping the auxiliary verb *mesti* above it to move to CP would violate the head movement constraint by Travis (1984). However, its grammaticality confirms that *ada* is merged in C^0, and not moved. (42) also suggests that subject-auxiliary inversion in Malay is optional; instead of moving *mesti* to host *kah*, *ada* is merged.

(42) *Ada-kah Ali mesti baca novel?*
AUX-Q A. must read novel
'Must Ali read a novel?'
#'Must Ali have read a novel?'

Also notice the difference in interpretation between (41) and (42). In examples (41), an eventive reading is obtained when *ada* is merged in the inflectional layer, which indicates that it carries some semantic content. Conversely, the auxiliary is meaningless and merely plays a support role for interrogative marker *kah* in example (42). This entails that the semantic content of *ada* depends on where it is merged – meaningful when merged in TP but meaningless when merged in CP. The interpretation of the sentence with *mesti* moved to C^0 in (43a) is similar to the one with *ada* base-generated in C^0 in (42), whereas the one in (43b) is similar to (41b).

(43) a. *Mesti-kah Ali baca novel?*
AUX-Q A. read novel
'Must Ali read a novel?'

b. *Mesti-kah Ali ada baca novel?*
must-Q A. AUX read novel
'Must Ali have read a novel?'

Lastly, information conveying mood is expressed using additional lexical items (or the omission of the subject):

(44) a. *Lukis komik!* (Imperative)
draw comic
'Draw comics!'

b. *Jangan lukis komik!* (Prohibitive)
PROH draw comic
'Do not draw comics!'

c. *Jom lukis komik.* (Hortative)
HORT draw comic
'Let's draw comics.'

d. *Semoga dia lukis komik.* (Optative)
OPT 3.SG draw comic
'May he draw comics.'

e. *Ada-kah dia lukis komik?* (Interrogative)
AUX-Q 3.SG draw comic
'Does he draw comics?'

1.5.5 Finiteness

Finiteness is not a grammatical notion that is recognised in Malay reference grammars, presumably due to the absence of overt morphological realisation of finiteness within a clause in Malay. However, there are noticeable characteristics shown by clauses that are often used in non-finite contexts (e.g. control clauses) that are disjoint from main clauses. For example, *untuk* (for) is commonly used to introduce control clauses. As illustrated below, the subordinate clauses may not be introduced by the complementiser *bahawa* or relativiser *yang*, which are only used in finite complement or relative clauses:

(45) a. *Saya pilih* {*untuk* /**bahawa*} *ber-senam.*
1.SG choose for COMP INTR-exercise
'I choose to exercise.'

b. *Pilih-an saya* {*untuk* /**yang*} *ber-senam adalah muktamad.*
choose-NMZ 1.SG for REL INTR-exercise COP final
'My choice to exercise is final.'

Nak – contraction of *hendak* (want) – is also commonly used in non-finite clauses in spoken Malay. It does not carry any semantics concerning wanting or desire, as shown in the example below with a non-sentient inanimate subject.

(46) a. *Buku itu susah* {*untuk* / *nak*} *ter-apung dalam air*.[13]
book DIST difficult for to NVOL-float in water
'That book is difficult to float on water.'

b. *Buku-buku* {*untuk* / *nak*} *di-lupus ada di atas meja.*
book-RED for to PASS-dispose EXIST LOC on table
'The books to be disposed are on the table.'

It is also evolving into a kind of aspectual marker meaning "about to" or "almost", akin to English *will*, as shown in the examples below:

(47) a. *Buku itu nak ter-koyak.*
book DIST NAK NVOL-tear
'That book is almost torn.'

b. *Pokok itu nak mati.*
tree DIST NAK die
'That tree is almost dead.'

Salleh (1995) argues that certain auxiliaries cannot occur in non-finite clauses in Malay. To illustrate, the bracketed constituents in examples (48b-d) below are control clauses that cannot accommodate aspectual markers.

(48) a. *Saya* {*akan* / *telah* / *sedang*} *ber-senam.*
1.SG PROS PRF PROG INTR-exercise
'I will/have/am exercise(ed/ing).'

b. *Saya perlu* [(**akan*) *ber-senam.*]
1.SG need PROS INTR-exercise
'I need to exercise.'

c. *Dia ber-senam* [*tanpa* (**telah*) *meŋ-ambil sarapan=nya.*]
3.SG INTR-exercise without PRF ACT-write breakfast=3
'He is exercising without having taken his breakfast.'

d. *Susah* [*untuk* (**sedang*) *ber-senam di sini.*]
difficult for PROG INTR-exercise LOC here
'It is difficult to be exercising here.'

---

[13] This use of *nak* as a non-finite marker is also attested in Classical Malay in its non-contracted form, as shown below:

(iii) *Ada-pun ke-datang-an hamba ini, ia-lah hendak meŋ-rebut negeri tuanhamba.*
EXIST-TOP NMZ-come-NMZ 1.SG PROX 3.SG-COM to ACT-seize country 2.SG
'As for my arrival, it is to seize your country.'
(Hikayat Sang Boma – 17th century AD)

1.5.6 Negation

There exist two negators in Malay: *tidak* and *bukan*. Kroeger (2014) argues that *tidak* is a marker of internal (predicate) negation that is commonly used to negate verbal predicates. On the other hand, *bukan* is a marker of external (sentential) negation and is believed to introduce an opposition, alternative, or contradiction to whatever was uttered in the preceding discourse. It can be said to be more pragmatically sensitive as it has a contrastive connotation (thus glossed as CNTR henceforth), which is why a proper context is required for its use. For example, the use of *bukan* in example (49b) would be infelicitous without the context provided in the disjunct that follows it.

(49) a. *Dia tidak meŋ-tangis.*
3.SG NEG ACT-cry
'He did not cry.'

b. *Dia bukan meŋ-tangis, tetapi ber-pura-pura saja.*
3.SG CNTR ACT-cry but INTR-pretend-RED just
'He did not cry but was just pretending.'

With regard to nonverbal predicates, an asymmetry is observed such that adjectival and prepositional predicates are negated by *tidak*, whereas nominal predicates can only be negated by *bukan*. This finding is interesting as such predicates do not require a context for the use of *bukan*, as opposed to the need for a context by verbal predicates, as in example (49b). Kroeger (2014) states that the use of *bukan* with nominal predicates "is a purely grammatical requirement. The factors that license the use of *bukan* in verbal clauses, by contrast, seem to be largely pragmatic" (p. 138).

(50) a. *Dia tidak pandai.*
3.SG NEG smart
'He is not smart.'

b. *Dia tidak seperti orang lain.*
3.SG NEG like person other
'He is not like other people.'

c. *Dia* {*bukan* /**tidak* } *se-orang genius.*
3.SG CNTR NEG one-CLF genius
'He is not a genius.'

It is possible for the two negators to co-occur. As shown below, it intervenes between the epistemic modal and *tidak*.

(51) *Ali mungkin bukan tidak pandai. Dia hanya malas.*
A. might CNTR NEG smart 3.SG just lazy
'It might not be that Ali isn't smart. He's just lazy.'

Examples with two instances of *bukan* – one pragmatic and the other grammatical – in copular clauses with nominal predicates are possible. However, it seems that the contrastive one must occur in the periphery.

(52) *Bukan-nya Ali bukan se-orang genius. Dia hanya belum ter-kenal lagi.*[14]
CNTR-EMPH A. CNTR one-CLF genius 3.SG just IMPRF NVOL-known yet
'It's not that Ali isn't a genius. He's just not known yet.'

Following from that, *tidak* can never take scope over the *bukan*.

(53) **Ali tidak bukan pandai. Dia hanya malas.*
A. NEG CNTR smart 3.SG just lazy
(It's not that Ali isn't smart. He's just lazy.)

In relation to auxiliaries, the negators precede all but epistemic modals.

(54) *Ali mesti tidak akan boleh habis-kan novel itu.*
A. must NEG PROS can finish-APPL novel DIST
'Ali must not be able to finish that novel.'

1.5.7 Verbs and Verbal Morphology

The Malay VP is syntactically complex. For example, several different affixes can simultaneously modify the verb, e.g. the voice-marking prefixes, the applicative suffix, the causative suffix, etc. Each morpheme should correspond to a different head within the split-VP, in accordance with the Mirror Generalisation by Baker (1985): "Morphological derivations must directly reflect syntactic derivations (and vice versa)" (p. 375). It can be analysed that the lexical core of the verb starts out in V^0 and

[14] The emphatic marker *-nya* must not be confused with the 3rd person enclitic of the same form. Its use as an emphatic marker is neither referential nor deictic. Therefore, it does not agree with the subject of the example below. Also, it only occurs in the left periphery on moved constituents.

(iv) a. *Pandai-nya* { *saya* / *kamu* / *dia.* }
smart-EMPH 1.SG 2.SG 3.SG
'So smart, I am/you are/he is.'

b. { *Saya* / *kamu* / *dia* } *pandai* (**-nya.*)
1.SG 2.SG 3.SG smart -EMPH
'I am/you are/he is smart.'

it undergoes head movement to v^0 before finally landing in Voice^0 to obtain the voice prefix, as per Cole, Hermon, and Yanti (2008), who argue that the *v*P is immediately dominated by a VoiceP projection in which the voice markers are merged.

For a further elaboration of the Mirror Generalisation, consider examples (55). They demonstrate how the polymorphemic nominal *ketidakberhasilan* reflects morphemes that are known to head their own projections, such as *tidak* (head of NegP) and *berhasil* (head of VP), the latter of which can further be decomposed into the possessive verb *ada* (head of VP) and *hasil* (head of NP).[15] As illustrated below, the complex nominal in (55a) corresponds to the complex clause in (55b), and the complex verb in (55b) corresponds to the VP in (55c). Thus, the complex lexeme is assembled in a systematic fashion that reflects the hierarchy of syntactic projections that is expected in a transitive clause, NegP-VP-DP.

(55) a. *Saya benci ke-tidak-ber-hasil-an=nya.*
1.SG hate NMZ-NEG-INTR-result-NMZ=3
'I hate his lack of results.'

b. *Saya benci bahawa dia tidak ber-hasil.*
1.SG hate COMP 3.SG NEG INTR-result
'I hate that he does not have results.'

c. *Saya benci bahawa dia tidak ada hasil.*
1.SG hate COMP 3.SG NEG have result
'I hate that he does not have results.'

Therefore, the various verbal affixes can be assumed to head their own projections. In the case of there being extra morphemes on the verb, e.g. the causative and applicative affixes, as in example (56), the causative affix can be analysed to be a manifestation of the v^0 head, as is commonly assumed, whereas the applicative affix can be analysed to head an additional functional projection, e.g. ApplP.

(56) *Ali meŋ-per-ingat-kan saya tentang novel=nya.*
A. ACT-CAUS-remember-APPL 1.SG about novel=3
'Ali reminded me about his novel.'

[15] One of the functions of the intransitive *ber-* prefix is to encode states and properties. By attaching to nominals, it encodes a reading that involves possession or properties. For example, *ber-suami* (INTR-husband), *ber-akal* (INTR-intellect), and *ber-duit* (INTR-money) mean having a husband/being married, having intellect/being sentient, and having money/being rich, respectively.

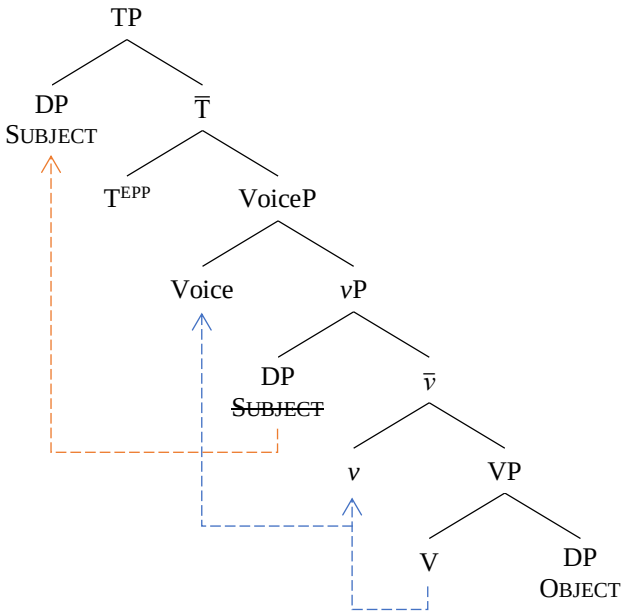


Figure 10: The Syntactic Structure of a Transitive VP in Malay

The following prefixes occur in complementary distribution, which entails that they head the same projection, VoiceP. Examples (57) illustrate their use.

Prefix	Usage
Meŋ-	Active voice marker
Di-	Passive voice marker
Ter-	Accidental/non-volitional marker
Ber-	Intransitive marker

Table 6: The Heads of VoiceP

(57) a. *Saya meŋ-tumbuk Ali.*
1.SG ACT-punch A.
'I punched Ali.'

b. *Saya di-tumbuk Ali.*
1.SG PASS-punch A.
'I was punched by Ali.'

c. *Saya ter-tumbuk Ali.*
1.SG NVOL-punch A.
'I (accidentally) punched Ali.'

d. *Saya dan Ali ber-tumbuk.*
1.SG and A. INTR-punch
'Ali and I fought.'

1.5.8 Voice

As covered in Section 1.4.2, Malay has three instantiations of voice: the active voice, the passive voice, and the patient voice. The active voice is marked by the nasal *meŋ-* prefix, whereas the passive voice is marked by the *di-* prefix. As for the patient voice, although it is not marked by an overt morpheme, its use is indicated by word order, as illustrated below:[16]

(58) a. *Dia telah meŋ-beli ikan itu.* (Active Voice)
3.SG PRF ACT-buy fish DIST
'He has bought the fish.'

b. *Ikan itu telah di-beli (oleh=nya.)* (Passive Voice)
fish DIST PRF PASS-buy by=3
'The fish has been bought by him.'

c. *Ikan itu telah dia beli.* (Patient Voice)
fish DIST PRF 3.SG buy
'He has bought the fish.'

Additionally, it is possible to omit the active voice prefix in spoken Malay, but not the passive voice prefix:

(59) a. *Dia telah (meŋ-)beli ikan itu.* (Bare Active Voice)
3.SG PRF ACT- buy fish DIST
'He has bought the fish.'

b. *Ikan itu telah *(di-)beli (oleh=nya.)* (Passive Voice)
fish DIST PRF PASS- buy by=3
'The fish has been bought by him.'

Despite the claim that the patient voice is a form of passive, as in Karim et al. (2014), it is syntactically a transitive clause with both the agent and the patient holding argument status. Unlike the optional *oleh*-phrase in the passive voice in Malay, the agent in the patient voice is obligatory.

[16] It must be noted that it is more common to find 1st and 2nd person pronominal and clitic agents in the patient voice, as opposed to 3rd person agents in general. However, use of the 3rd person does not cause ungrammaticality, at least in the variety of Malay studied in this thesis.

In the absence of any auxiliaries or negators, a patient voice clause seems similar to a bare active voice clause in spoken Malay with object topicalisation, e.g.:

(60) a. *Ikan itu*$_1$ *dia beli* t$_1$. (Bare Active Voice + Topicalisation)
fish DIST 3.SG buy
'The fish, <u>he</u> bought.'

b. *Ikan itu dia beli.* (Patient Voice)
fish DIST 3.SG buy
'He bought <u>the fish</u>.'

Therefore, one might suspect that the patient voice is indeed the bare active voice with topicalisation. However, the addition of an auxiliary or a negator dispels this hypothesis as both voices have different word orders. They occur between the agent and the verb in the active voice (bare or not) and preceding the agent in the patient voice. Topicalisation does not trigger subject-auxiliary inversion in Malay.

(61) a. *Ikan itu*$_1$ *dia akan beli* t$_1$. (Bare Active Voice + Topicalisation)
fish DIST 3.SG PROS buy
'The fish, <u>he</u> will buy.'

b. *Ikan itu akan dia beli.* (Patient Voice)
fish DIST PROS 3.SG buy
'He will buy <u>the fish</u>.'

Another way to verify that the patient voice does not involve topicalisation is by embedding it in a relative clause, which bans topicalisation. As in (24), the patient voice can be used in a relative clause, which entails that its derivation does not involve topicalisation. To repeat, the contrast below demonstrates that the patient voice is possible in a relative clause, whilst the bare active voice with topicalisation is not:

(62) a. **Tempat di mana ikan itu aku akan beli akan tutup.* (Active Voice)
place LOCwhere fish DIST 1.SG PROS buy PROS close
(The place where, that fish, <u>I</u> will buy will close.)

b. *Tempat di mana ikan itu akan aku beli akan tutup.* (Patient Voice)
place LOCwhere fish DIST PROS 1.SG buy PROS close
'The place where I will buy <u>that fish</u> will close.'

Moreover, this finding indicates that, whilst the patient of the patient voice occupies SpecTP, the agent occupies a position between auxiliaries and the verb, presumably Spec*v*P, where agents are merged.

1.5.9 Demonstratives

Although there is no fixed class of determiners in the language, Malay does have demonstratives, which close off the maximal projection of the nominal. They occur at the right edge of the DP and c-command the NP, including any restrictive relative clauses postmodifying the NP. It is ungrammatical for a restrictive relative clause to occur outside the DP layer, as seen in the contrast below:

(63) a. *Anak yang sedang ber-cakap ini boleh ber-kira.*
child REL PROG INTR-speak PROX can INTR-count
'The child that is speaking here can count.'

b. **Anak ini yang sedang ber-cakap boleh ber-kira.*
child PROX REL PROG INTR-speak can INTR-count
(The child here, that is speaking, can count.)

However, non-restrictive relative clauses occur outside the constituent marked by the demonstrative:

(64) *Anak ini, yang saya sayang sangat-sangat, boleh ber-kira.*
child PROX REL 1.SG love very-RED can INTR-count
'This child, whom I love very much, can count.'

1.5.10 NP Modifiers

Various pre-nominal classifiers are used alongside cardinal numerals to quantify the noun. The classifier must be appropriate to the noun that is quantified, as shown below:[17]

(65) a. *Tiga biji telur* (Small, round items)
three seed.CLF egg
'Three eggs'

b. *Tiga ekor kucing* (Animals)
three tail.CLF cat
'Three cats'

c. *Tiga orang guru* (People)
three person.CLF teacher
'Three teachers'

[17] Although the gloss in these examples includes the word on which the classifiers are based, classifiers in examples henceforth will be glossed only as CLF for simplicity.

On the other hand, other modifiers, which can come in the form of APs, NPs, PPs, and CPs, occur in post-nominal position, as shown below:

(66) a. *Telur bujur* (AP)
egg oval
'An oval egg'

b. *Kucing peŋ-musnah* (NP)
cat AG-destroy
'A destroyer of a cat'

c. *Guru di sekolah* (PP)[18]
teacher LOC school
'A teacher at school'

d. *Kereta yang telah ter-bengkalai* (CP)
car REL PRF NVOL-abandon
'A car that has been abandoned'

The non-clausal modifiers and the modificand may also be intervened by what scholars such as Simin (1988) call ligature *yang*, which is used to "ligate an attribute to its head noun so as to distinguish it from other nouns that are not so ligated" (p. 193). For instance, the NPs in the examples below differ from the ones in (66) in that the modifier following *yang* is not a property of the NPs in (66), hence is contrastive. The ligature is said to be "individualising"; it possesses contrastive properties.

(67) a. *Telur bujur yang hitam* (AP)
egg oval LIG black
'A <u>black</u> oval egg'

b. *Kucing peŋ-musnah yang peŋ-sayang* (NP)
cat AG-destroy LIG AG-love
'A <u>loving</u> destroyer of a cat'

c. *Guru di sekolah yang di sana* (PP)
teacher LOC school LIG LOC there
'A teacher at the school <u>over there</u>'

[18] Due to the varying degrees of locative specification by locative preposition *di*, it is glossed here as LOC. It could correspond to either English *at* or *in*.

1.5.11 Pronouns

Based on several facts that are true of what Noguchi (1997) calls *N-pronouns* in Japanese, personal pronouns in Malay do not comprise a functional determiner category. Most importantly, personal pronouns in Malay can be modified by relative clauses, as well as demonstratives themselves. Although pronouns in English can be modified by relative clauses and determiners in exceptional cases, it is argued by Noguchi (1997) that such a use of pronouns demonstrates category conversion, which is not "a core characteristic of English personal pronouns" (p. 779). The use of demonstratives with personal pronouns in Malay and Japanese does not change the meaning of the phrase, which makes it an unmarked characteristic. For example, the relative clause and demonstrative in (68b) do not specify a single proximal entity that has the property of being old from a set of possible entities how such modifiers in English do. They only serve to modify the pronoun in a non-restrictive manner.

(68) a. *Saya ini sudah letih ber-lari.*
1.SG PROX already tired INTR-run
'I am already tired of running.'

b. *Saya yang telah tua ini sudah letih ber-lari.*
1.SG REL PRF old PROX already tired INTR-run
'This little old me is already tired of running.'

If pronouns in Malay were of a functional D category, they should not be able to co-occur with another D^0 head, as in (68a). However, since they are lexical in Malay, the combination is possible, as also shown below with a possessive pronoun:

(69) *Kucing saya ini di-lenyek oleh lori.*
cat 1.SG PROX PASS-squash by lorry
'This cat of mine was squashed by a lorry.'

In the following example, the pronoun co-occurs with various other nominal and nominal-modifying projections such as the classifier, relative clause, and demonstrative, with the demonstrative closing off the projection.

(70) *Dua ekor kucing hitam saya yang mati itu telah di-tanam.*
two CLF cat black 1.SG REL die DIST PRF PASS-bury
'Those two black cats of mine that died have been buried.'

1.5.12 Ā-Movement

Focalisation is marked by *lah*, which encodes contrastive focus, especially when the focus is a phrasal constituent. On the other hand, topicalisation is marked by *kan*, which introduces a new topic.[19] As shown below, both elements are optional:

(71) a. *Semalam* (*lah*) *Ali ber-lari balik ke rumah.* (Focalisation)
yesterday FOC A. INTR-run return to house
'YESTERDAY Ali ran back home.'

b. *Semalam* (*kan*) *Ali ber-lari balik ke rumah.* (Topicalisation)
yesterday TOP A. INTR-run return to house
'Yesterday, Ali ran back home.'

c. *Semalam* (*kan*) *ke rumah* (*lah*) *Ali ber-lari balik.* (Top + Foc)
yesterday TOP to house FOC A. INTR-runreturn
'Yesterday, TO HOME Ali ran back.'

Both *lah* and *kan* also have functions other than focus and topic marking. When it is a head that is modified by *lah*, it does not necessarily carry contrastive focus. Ajamiseba (1983) states that "-*lah* constructions are sentences that contain new information on the content level and on the metalevel signalled by the comment marking particle -*lah*" (p. 78), and this characterisation of *lah* is similar to the one by Müller-Gotama (1995). In example (72), *lah* marks new information but does not carry a contrastive focal reading.

Context: *Ali dikejar anjing.* (Ali was chased by dogs.)

(72) *Ber-lari-lah dia balik ke rumah.*
INTR-run-COM 3.SG return to house
'(So) he ran back home.'

Conversely, *lah* marks verum focus in the example below:

Context: *Ali berjalankah berlari balik rumah?* (Did Ali walk or run back home?)

(73) *Ber-lari-lah dia balik ke rumah.*
INTR-run-FOC 3.SG return to house
'He RAN back home.'

[19] *Kan* is only used in spoken Malay. Perhaps an appropriate translation for *kan* in English would be "*y'know* X, [IP …]" where X is the topic modified by *kan*.

As for the other form of *kan*, it is used as a question tag. It is the contracted form of contrastive marker *bukan*.

(74) *Semalam ~~bu~~kan Ali ber-lari balik ke rumah?*
yesterday CNTR A. INTR-run return to house
'It was yesterday, wasn't it, that Ali ran back home?'

1.5.13 Extraction

The extraction of arguments in Malay is governed by the voice morphology on the verb. Only the trigger may be extracted, meaning that the argument to be extracted must tally with the appropriate voice marker on the verb. A mismatch between the correct voice marker and the argument extracted would cause ungrammaticality, as illustrated in (75b). Therefore, only the agent may be extracted from a clause in the active voice and only the patient in the passive or patient voice. It turns out that the trigger always corresponds to the grammatical subject in Malay, e.g. the trigger of a clause may only correspond to the agent in the active voice or the patient in the passive and patient voices, all of which occupy the preverbal subject position, as illustrated in (16a-c).

(75) a. *Siapa-kah yang meŋ-beli nasi itu?* (Active – Agent)
who-Q COMP ACT-buy rice DIST
'Who bought that rice?'

b. **Apa-kah yang dia meŋ-beli?* (Active – Patient)
what-Q COMP 3.SG ACT-buy
(What did he buy?)

c. *Apa-kah yang di-beli oleh=nya?* (Passive – Patient)
what-Q COMP PASS-buy by=3
'What was bought by him?'

d. *Apa-kah yang telah dia beli?* (Patient – Patient)[20]
what-Q COMP PRF 3.SG buy
'What has he bought?'

[20] The aspectual marker is used in examples of the patient voice to show the clause-initial position of the auxiliary that is characteristic of the patient voice in Malay.

Notice that when the trigger is extracted, the resulting construction is a cleft whose focus is separated from the rest of the clause by *yang*.[21] In fact, relativisation also follows the same principle and looks similar. However, unlike clefts, relative clauses in Malay observe the doubly-filled COMP filter, which prevents both the specifier and head positions of CP to be filled, as shown below with relativised triggers:[22]

(76) a. *Orang* (**siapa*) *yang meŋ-beli nasi itu…* (Active – Agent)
person who REL ACT-buy rice DIST
'The person that bought that rice…'

b. **Benda* (*apa*) *yang dia meŋ-beli…* (Active – Patient)
thing what REL 3.SG ACT-buy
(The thing that he bought…)

c. *Benda* (**apa*) *yang di-beli oleh=nya…* (Passive – Patient)
thing what REL PASS-buy by=3
'The thing that was bought by him…'

d. *Benda* (**apa*) *yang telah dia beli…* (Patient – Patient)
thing what REL PRF 3.SG buy
'The thing that he has he bought…'

Conversely, when an adjunct is extracted, *yang* is not used, regardless of the voice of the clause:

(77) a. *Bila-kah* (**yang*) *dia meŋ-beli nasi itu?* (Active – Adjunct)
when-Q COMP 3.SG ACT-buy rice DIST
'When did he buy that rice?'

[21] I have glossed the *yang* of cleft constructions as COMP as a general complementiser, and the *yang* of relative clauses as REL based on its specific function in relative clauses.

[22] However, it is possible for the *wh*-phrase and *yang* to co-occur in free relative clauses, but only with relativised triggers, as shown below:

(v) a. *Saya suka apa yang di-beli oleh=nya itu.*
1.SG like what REL PASS-buy by=3 DIST
'I like what was bought by him.'

b. *Saya suka bila* (**yang*) *nasi di-beli oleh=nya.*
1.SG like when REL rice PASS-buy by=3
'I like when rice is bought by him.'

b. *Mengapa-kah* (**yang*) *nasi itu di-beli?* (Passive – Adjunct)
why-Q COMP rice DIST PASS-buy
'Why was <u>that rice</u> bought?'

c. *Di mana-kah* (**yang*) *nasi itu telah dia beli?* (Patient – Adjunct)
LOC where-Q COMP rice DIST PRF 3.SG buy
'Where did he buy <u>that rice</u>?'

Similarly, relative clauses with relativised adjuncts do not allow *yang*:

(78) a. *Masa bila* (**yang*) *dia meŋ-beli nasi itu…* (Active – Adjunct)
time when COMP 3.SG ACT-buy rice DIST
'The time when <u>he</u> bought that rice…'

b. *Sebab mengapa* (**yang*) *nasi itu di-beli…* (Passive – Adjunct)
reason why COMP rice DIST PASS-buy
'The reason why <u>that rice</u> was bought…'

c. *Tempat di mana* (**yang*) *nasi itu dia beli…* (Patient – Adjunct)
place LOC where COMP rice DIST 3.SG buy
'Where he bought <u>that rice</u>…'

This restriction reflects Austronesian alignment and is similar in languages such as Tagalog. As shown in the following examples by Aldridge (2007), extraction is only possible of the trigger, and the resulting construction is a cleft:

<u>Tagalog (Austronesian – Philippine)</u>

(79) a. *Sino ang b‹um›ili ng isda?* (Actor Voice – Agent)
who ANG ‹INTR.PRF›buy OBL fish
'<u>Who</u> bought a fish?'

b. **Sino ang b‹in›ili ang isda?* (Actor Voice – Patient)
who ANG ‹TR.PRF›buy ANG fish
(Who bought <u>the fish</u>?)

c. *Ano ang b‹in›ili ng babae?* (Patient Voice – Patient)
what ANG ‹TR.PRF›buy ERG woman
'<u>What</u> did the woman buy?'

Recall from Section 1.4.1 that spoken Malay is more grammatically economical and that various grammatical elements can be omitted. In spoken Malay, extraction is unrestricted in terms of the choice of argument, especially owing to the omission of the voice markers, which are believed to possess features that block movement of certain arguments (Aldridge, 2007; Cole et al., 2008; Soh & Nomoto,

2011). To illustrate, extraction of both the agent and the patient is possible in the following examples when the verb is bare, i.e. not affixed by either active *meŋ-* or passive *di-*. Note that the clause in example (80b) is not in the patient voice, as the agent does not occur below the auxiliary verb.

(80) a. *Siapa-kah yang telah beli nasi itu?* (Subject)
who-Q COMP PRF buy rice DIST
'Who has bought that rice?'

b. *Apa-kah yang Ali telah beli?* (Object)
what-Q COMP A. PRF ACT-buy
'What did Ali buy?'

Extraction of either the agent or the patient being possible when the verb is bare can therefore be taken as evidence that the affixes are responsible for the restrictions on the subject and extraction in Malay. As extraction of both the subject and object in simplified verbal clauses is possible in spoken Malay, the trigger vs. non-trigger distinction can be said to have broken down.

1.5.14 *Wh*-Interrogatives

*W*h-movement in Malay is optional, as *wh*-in-situ is possible in the language. In fact, partial *wh*-movement is also possible, as argued by Cole and Hermon (2000), as shown in the following examples:

(81) a. *Awak kata siapa suka kek?* (*Wh*-In-Situ)
2.SG say who like cake
'Who did you say likes cake?'

b. *Awak kata dia suka apa?*
2.SG say 3.SG like what
'What did you say he likes?'

c. *Awak kata siapa*$_1$ *yang* t$_1$ *suka kek?* (Partial *Wh*-Ex-Situ)
2.SG say who COMP like cake
'Who did you say likes cake?'

d. *Awak kata apa*$_1$ *yang dia suka* t$_1$?
2.SG say what COMP 3.SG like
'What did you say he likes?'

e. *Siapa*$_1$ *yang awak kata* t$_1$ *suka kek?* (*Wh*-Ex-Situ)
who COMP 2.SG say like cake
'Who did you say likes cake?'

f. *Apa*$_1$ *yang awak kata dia suka* t$_1$?
what COMP 2.SG say 3.SG like
'What did you say he likes?'

Even though it is possible for a *wh*-phrase to remain in-situ, the use of interrogative particle *kah* makes it obligatory for the *wh*-phrase to move to the left periphery. As illustrated below, *kah* cannot attach to *wh*-phrases in-situ:

(82) a. *Dia bagi apa* (**-kah*) *kepada Ali semalam?*
3.SG give what -Q to A. yesterday
'What did he give to Ali yesterday?'

b. *Dia bagi surat yang mana* (**-kah*) *kepada Ali semalam?*
3.SG give letter LIG which -Q to A. yesterday
'Which letter did he give Ali yesterday?'

c. *Dia bagi surat ini kepada siapa* (**-kah*) *semalam?*
3.SG give letter PROX to who -Q yesterday
'To whom did he give this letter yesterday?'

d. *Dia bagi surat ini kepada Ali bila* (**-kah*)?
3.SG give letter PROX to A. when -Q
'When did he give this letter to Ali?'

As shown below, movement must occur:

(83) a. *Apa-kah yang dia bagi kepada Ali semalam?*
what-Q COMP 3.SG give to A. yesterday
'What did he give to Ali yesterday?'

b. *Surat yang mana-kah yang dia bagi kepada Ali semalam?*
letter LIG which-Q COMP 3.SG give to A. yesterday
'Which letter did he give Ali yesterday?'

c. *Kepada siapa-kah dia bagi surat ini semalam?*
to who-Q 3.SG give letter PROX yesterday
'To whom did he give this letter yesterday?'

d. *Bila-kah dia bagi surat ini kepada Ali?*
when-Q 3.SG give letter PROX to A.
'When did he give this letter to Ali?'

Therefore, *kah* can be analysed as an affix with a strong interrogative feature that attracts a constituent to its specifier to function as its host. Failure for a *wh*-phrase to move to its specifier would result in the feature not getting checked, as well as a violation of the stray affix filter, hence render the derivation of the clause ungrammatical.

With regard to its syntactic position, *kah* heads FocP. It corresponds to focus marker *lah* in declarative sentences according to Kader (1981) and it occurs above *yang*, which heads a lower CP projection following the movement of an argument *wh*-phrase, as shown in example (84).

(84) [$_{FocP}$ *Apa -kah* [$_{FinP}$ *yang* [$_{TP}$ *dia sedang tulis?*]]]
what -Q COMP 3.SG PROG write
'What is he writing?'

1.5.15 Polar Interrogatives

When polar interrogatives are concerned, the requirement that the focus move to the left periphery seemingly becomes more relaxed, as it is permitted for *kah* to attach to a focus in-situ, as shown below:

(85) a. *Dia bagi-kah surat kepada Ali semalam?*
3.SG give-Q letter to A. yesterday
'Did he give a letter to Ali yesterday?'

b. *Dia bagi surat-kah kepada Ali semalam?*
3.SG give letter-Q to A. yesterday
'Did he give a letter to Ali yesterday?'

c. *Dia bagi surat ini-kah kepada Ali semalam?*
3.SG give letter PROX-Q to A. yesterday
'Did he give this letter to Ali yesterday?'

d. *Dia bagi surat ini kepada Ali-kah semalam?*
3.SG give letter PROX to A.-Q yesterday
'Did he give this letter to Ali yesterday?'

e. *Dia bagi surat ini kepada Ali semalam-kah?*
3.SG give letter PROX to A. yesterday-Q
'Did he give this letter to Ali yesterday?'

However, there is reason to believe that polar interrogatives are different from *wh*-interrogatives with respect to the interrogative *kah* particle.[23] The *kah* particle in polar interrogatives (henceforth treated as a disjunctive marker instead of the standard interrogative particle and glossed as DISJ.Q) functions as a disjunctive marker. Presumably, the constituent following *kah* is elided, but it may be phonologically realised optionally, as shown below:

(86) a. *Dia bagi surat kepada Ali semalam kah* (*tak*)?
3.SG give letter to A. yesterday DISJ.Q NEG
'Did he give a letter to Ali yesterday (or not)?'

b. *Dia bagi surat kah* (*buku*) *kepada Ali semalam?*
3.SG give letter DISJ.Q book to A. yesterday
'Did he give a letter (or a book) to Ali yesterday?'

c. *Dia bagi surat ini kah* (*yang itu*) *kepada Ali semalam?*
3.SG give letter PROX DISJ.Q LIG DIST to A. yesterday
'Did he give this letter (or that one) to Ali yesterday?'

d. *Dia bagi surat ini kepada Ali kah* (*Zul*) *semalam?*
3.SG give letter PROX to A. DISJ.Q Z. yesterday
'Did he give this letter to Ali (or Zul) yesterday?'

e. *Dia bagi surat ini kepada Ali semalam kah* (*kelmarin*)?
3.SG give letter PROX to A. yesterday DISJ.Q day.before
'Did he give this letter to Ali yesterday (or the day before)?'

It can be assumed that disjunctive *kah* is bound by a c-commanding interrogative operator in the left periphery. This is made apparent in cases in which the interrogative form of the disjunctive marker cannot be used when *kah* already occupies CP. In such cases, the declarative variant *atau* must be used.

[23] The *kah* particle in polar interrogatives is pronounced differently from the one in *wh*-interrogatives. Unlike the *kah* particle in *wh*-interrogatives (pronounced as [kah]), the one in polar interrogatives is pronounced as [kə]. It is not possible to pronounce it the same way in content interrogatives. In fact, it is spelt as 'ke' in informal writing and speech transcription in Malay, as shown below:

(vi) *Mereka nak terima ke?*
3.PL want accept DISJ.Q
'Do they want to accept?'
(Amran, 2018)

(87) *Ada-kah dia bagi surat kepada Ali* { *atau* /* *kah* } *Zul semalam?*
do-Q 3.SG give letter to A. DISJ DISJ.Q Z. yesterday
'Did he give a letter to Ali or Zul yesterday?'

Since it is a disjunctive marker, there can be multiple instantiations of the head in a single phrase, as in the compound DP in (88a). In example (88b), *kah* is seen to conjoin different levels of syntactic structure, namely DP and CP. Otherwise, another possible analysis for (88b) would be for *ikan* to be another clause conjoined to the &P that has undergone conjunction reduction.

(88) a. *Dia suka ayam kah ikan kah itik?*
3.SG like chicken DISJ.Q fish DISJ.Q duck
'Does he like chicken, fish, or duck?'

b. *Dia makan ayam kah ikan kah dia tak makan langsung?*
3.SG eat chicken DISJ.Q fish DISJ.Q 3.SG NEG eat at.all
'Did he eat chicken or fish, or did he not eat at all?'

The structure of example (88b) is provided below:

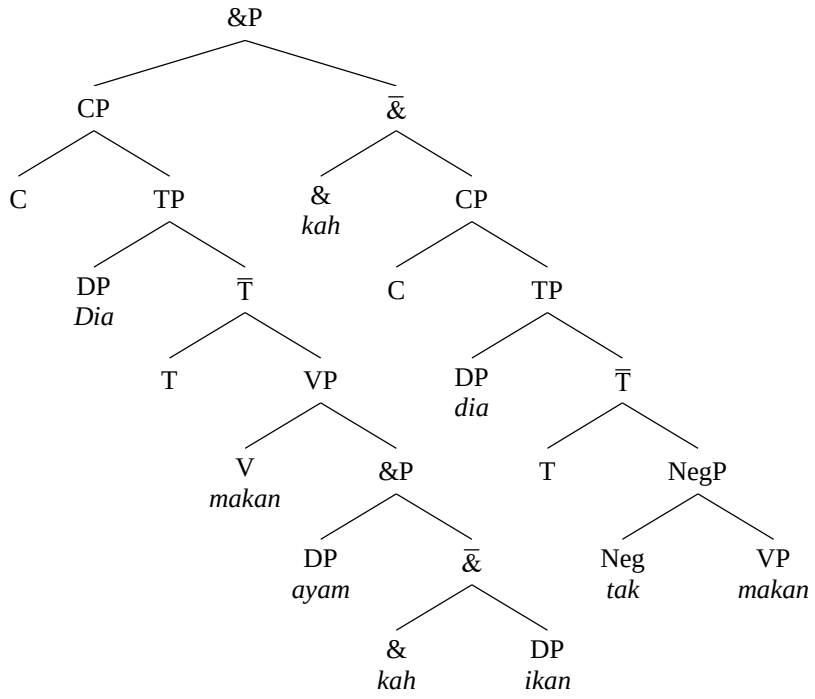


Figure 11: The Structure of a Polar Interrogative in Malay

Forming polar interrogatives using this disjunctive strategy with a syntactic head that shares the same form with some interrogative particle is apparently quite productive cross-linguistically, as argued in Bailey (2013). Even the sentence-final particle used in polar interrogatives in Mandarin Chinese *ma* (嗎) is argued by Aldridge (2011b) to have grammaticalised from a marker of negation in a disjunctive question of the likes of (88). Therefore, such an analysis for the *kah* particle in polar interrogatives in Malay is not something new or unprecedented.

1.6 Delimiting the Notion of Copula

Having provided a précis of the relevant parts of the grammar of Malay, let us return to determining whether Malay *ialah* and *adalah* are copulas using the criteria and flowchart in Section 1.1.2.

1.6.1 The Copular Status of *Ialah* and *Adalah*

First and foremost, a copula must accompany a nonverbal predicate. This might seem an elementary characterisation for languages that make a clear distinction between verbal and nonverbal categories, such as English, but it happens that such a distinction cannot be made in a straightforward fashion in Malay, as described in Section 1.5.3. To illustrate, *lari* (run) can be what appears to be a noun or an intransitive verb, and an argument or a predicate without there being any change in its form:

(89) a. *Lari=nya se-ekor harimau tidak ter-tanding.* (Noun/Argument)
run=3 one-CLF tiger NEG NVOL-contest
'The running (speed) of a tiger is uncontested.'

b. *Dia lari.* (Intransitive Verb/Predicate)
3.SG run
'He ran.'

Nonetheless, it is possible to demonstrate a distinction between verbs and non-verbs in Malay using procliticisation, as argued in Section 1.5.3. As illustrated below, the 1st person proclitic may attach to the dynamic verb *lari* (run) and the stative verb *tahu* (know) but not to *guru* (teacher), indicating that it is not a verb. To remediate the ungrammaticality in (90c), a free-standing pronominal subject such as *aku* (1.SG) must be used.

(90) a. *Ku=lari.*
1=run
'I run.'

b. *Ku=tahu.*
1=know
'I know.'

c. **Ku=guru=nya.*
1=teacher=3
(I am the teacher.)

Seeing that *gurunya* is a nonverbal constituent, it is predicted to be able to be used as a nonverbal predicate and for the putative Malay copulas *ialah* and *adalah* to combine with it. Indeed, *ialah* is only compatible with the nonverbal predicate, as opposed to the verbal predicates, as illustrated below. Therefore, according to the flowchart, *ialah* accompanies a nonverbal predicate but does not accompany a verbal predicate, warranting its status truly as a copula.

(91) a. **Zul ialah lari.*
Z. COP run
(Zul runs.)

b. **Zul ialah tahu.*
Z. COP know
(Zul knows.)

c. *Zul ialah guru=nya.*
Z. COP teacher=3
'Zul is the teacher.'

As for *adalah*, it may combine with both verbal and nonverbal predicates. It happens that *adalah* carries perfective aspect and verum focus with a verbal predicate but not with a nonverbal predicate, as shown below (see Sections 2.2.1 and 2.2.2 for further discussion on the semantics and information structure of *adalah*). This semantic difference is taken to mean that items that appear identical are not necessarily one and the same, satisfying the second criterion. Therefore, *adalah* is a copula with nonverbal predicates and a focal aspectual auxiliary with verbal predicates.[24]

[24] The same can be said of English *be*, which lacks meaning when used with a nonverbal predicate (e.g. "she is a girl) but carries aspectual information when used with a verbal predicate (e.g. "she is walking").

(92) a. *Zul adalah lari.*
Z. ADALAH run
i. #'Zul runs.'
ii. 'Zul HAS run.'

b. *Zul adalah tahu.*
Z. ADALAH know
i. #'Zul knows.'
ii. 'Zul HAS known.'

c. *Zul adalah guru=nya.*
Z. ADALAH teacher=3
i. 'Zul is the teacher.'
ii. #Zul HAS been the teacher.

Concerning their form, *ialah* and *adalah* are not themselves verbs as they cannot serve as the host to the proclitic:

(93) a. **Ku=ialah guru=nya.*
1=IALAH teacher=3
(I am the teacher.)

b. **Ku=adalah guru=nya.*
1=ADALAH teacher=3
(I am the teacher.)

They behave likewise as auxiliaries, which also resist procliticisation. Based on this finding, the copulas can be analysed to be auxiliaries.

(94) a. **Ku=akan lari.* (Aspectual Marker)
1.SG=PROS run
(I will run.)

b. **Ku=mungkin lari.* (Epistemic Modal)
1.SG=might run
(I might run.)

c. **Ku=boleh lari.* (Deontic Modal)
1.SG=can run
(I can run.)

d. **Ku=ada lari.* (Eventive *Ada*)
1.SG=AUX run
(I did run.)

Given that *ialah* and *adalah* are auxiliaries, we can arrive at the conclusion that the nonverbal constituent *gurunya* serves as the predicate of (91c) and (92c) respectively. Most importantly, their combination with the nonverbal predicate makes it possible to consolidate the finding that the two items are copulas.

1.6.2 Linking Verbs

The first criterion leaves most linking verbs, also known as semi-copulas and pseudo-copulas in Hengeveld (1992), out of the question. Although certain linking verbs such as *jadi* (become) only minimally differ from the copula by the addition of aspectual information, they are verbal predicates.[25] For example, linking verbs may combine with the nonverbal constituent in a small clause, which is an environment that is often used to test for predicatehood, whereas the copula may not, as shown below. Thus, linking verbs are predicates, unlike copulas.

(95) a. *Nasi itu meŋ-buat* [SC *dia lapar*.] (Nonverbal Constituent)
rice DIST ACT-make 3.SG hungry
'That rice made him hungry.'

b. *Nasi itu meŋ-buat* [SC *dia jadi lapar*.] (Linking Verb)
rice DIST ACT-make 3.SG become hungry
'That rice made him become hungry.'

c. **Nasi itu meŋ-buat* [SC *dia adalah lapar*.] (Copula)
rice DIST ACT-make 3.SG COP hungry
(That rice made him be hungry.)

One exceptional linking verb that seems like a copula is *merupakan* (to form), which may also be used transitively. It is ambiguous as it may be a meaningful transitive verb that selects a nominal constituent as its direct object, as well as a meaningless copular verb that selects a nominal predicate. *Merupakan* itself does not contribute to the meaning of the sentence when used as a copular verb. In fact, it may be omitted altogether without any change in the semantics, as illustrated in (96b).

[25] Hengeveld (1992) argues that semi-copulas such as *become* and *remain* in English allow the nonverbal constituent to play the role as predicate based on the claim that the nonverbal constituent imposes certain selectional restrictions on the subject. For example, it is ungrammatical to say "the table became happy" because the adjective *happy* can only describe an animate entity. However, the same is true of pseudo-copulas such as *seem* and *look*, which are characterised as lexical predicates: "*the table seems/looks happy". Therefore, such selectional properties cannot be taken as a diagnostic for the predicatehood of the nonverbal constituent, as opposed to those of the verb itself.

(96) a. *Thailand meŋ-rupa-kan sempadan=nya.* (Transitive)
T. ACT-form-APPL border=3
'Thailand forms its borders.'

b. *Thailand* (*meŋ-rupa-kan*) *matlamat=nya.* (Copular)
T. ACT-form-APPL goal=3
'Thailand is his goal.'

The small clause test confirms the predicative status of transitive *merupakan*, but not copular *merupakan*, as illustrated in (97). According to the flowchart, the transitive variant fails to pass through the first stage – it does not accompany a nonverbal predicate since it is itself a predicate, so it is not identified as a copula. On the other hand, the other variant passes the first stage as it is non-predicative, and it accompanies a nonverbal predicate. A negative value for the second stage in the flowchart, which stipulates a semantic difference in its use in verbal and nonverbal predication, warrants its status as a copula. Therefore, the variant of *merupakan* that combines with nonverbal predicates can be characterised as a copula, as agreed by scholars such as Hassan and Mohd. (1994) and Alwi, Dardjowidjojo, Lapoliwa, and Moeliono (2003).

(97) a. *Zul meŋ-buat* [SC *Thailand matlamat=nya.*]
Z. ACT-make T. goal=3
'Zul made Thailand his goal.'

b. *Zul meŋ-buat* [SC *Thailand meŋ-rupa-kan sempadan=nya.*] (Transitive)
Z. ACT-make T. ACT-form-APPL border=3
'Zul made Thailand form its borders.'

c. **Zul meŋ-buat* [SC *Thailand meŋ-rupa-kan matlamat=nya.*] (Copular)
Z. ACT-make T. ACT-form-APPL goal=3
(Zul made Thailand his goal)

As a result of the first criterion, locative and existential clauses in Malay also lie beyond the scope of this thesis because the verb *ada* is a verbal predicate. As shown below, it can occur in a small clause:

(98) *Zul meŋ-buat* [SC *ada=nya per-se-faham-an antara negara.*]
Z. ACT-make EXIST=3 NMZ-one-understand-NMZ between country
'Zul let there be mutual understanding between countries.'

One complication that might arise from the first criterion is the fact that the copula *adalah* may combine with certain linking verbs, such as *menjadi* (become) and *merupakan* (form), as shown below. It might seem that the copula *adalah* (rather than the focal aspectual auxiliary) accompanies a verbal predicate, apparently contradicting the first criterion. On the contrary, there are reasons to believe otherwise.

(99) *Ke-jaya-an adalah* { *meŋ-rupa-kan / meŋ-jadi*} *matlamat=nya.*
NMZ-succeed-NMZ COP ACT-form-APPL ACT-become goal=3
'Success is his goal.'

First, notice that the first criterion implies that the item in question cannot itself be a predicate. On their own, *merupakan* has been shown to be either a transitive verbal predicate in (97b) or a copula in (97c), whilst *menjadi* has been shown to be an inchoative verbal predicate in (95b). However, the interpretation of (99) does not include any meanings of forming or becoming, entailing that both linking verbs are meaningless in combination with copula *adalah*. In fact, the linking verb or its combination with the copula may be omitted, leaving the subject and nonverbal predicate behind, as shown in (100). Their omission without changing the meaning of the sentence can be taken as evidence that *menjadi* is not the predicate. Therefore, it is the nonverbal constituent that is the predicate. Thus, the first criterion is not violated.

(100) *Ke-jaya-an* (*adalah* (*meŋ-jadi*)) *matlamat=nya.*
NMZ-succeed-NMZ COP ACT-become goal=3
'Success is his goal.'

Unsurprisingly, if only *adalah* was omitted in example (100), an inchoative interpretation would be forced, given that *menjadi* is a verbal predicate on its own. This finding entails that the use of *menjadi* as a meaningless verb, as opposed to its inchoative variant, is dependent on the occurrence of the copula *adalah*. Sentences like (99) then appear to be double-copula constructions in which there is a meaningless copular verb and a meaningless copular auxiliary occurring contiguously in the same clause (see Section 3.1.2 for arguments against a bi-clausal construction).

(101) *Ke-jaya-an* ~~*adalah*~~ *meŋ-jadi matlamat=nya.*
NMZ-succeed-NMZ COP ACT-become goal=3
i. #'Success is his goal.'
ii. 'Success becomes his goal.'

The key concept to the first criterion is the *predicate* status of the nonverbal constituent with which the copula (or the combination of copula and linking verb) combines. Although *jadi* may correspond to a verbal predicate on its own, it is not itself a predicate or does not form part of it when it combines with the copula *adalah*. With the first criterion in mind, the copula does not modify *menjadi* as it neither is optional – as the non-predicative use of *menjadi* relies on the occurrence of the copula – nor provides any meaning. Therefore, its meaningless variant (as opposed to the inchoative variant, which is ruled out by the first criterion) passes the first criterion and the first stage of the flowchart. Then, the difference in semantics in verbal (inchoative) and nonverbal predication (meaningless), as stipulated in the second criterion, then warrants its characterisation as a copular verb in nonverbal predication.

To further support a distinction between their copular and non-copular uses, one unique property of the copular variants of *merupakan* and *menjadi* – as well as the copula *adalah* – is their inability to be separated into the morphemes they appear to be comprised of. For example, the copula *adalah* may not be decomposed into *ada* and *lah*, whilst its focal aspectual variant may be decomposed into the two morphemes, given that *lah* is an optional focus morpheme (more on this in Section 2.2.3).

(102) a. *Zul ada* (*-lah*) *lari.*
Z. AUX -FOC run
i. #'Zul runs.'
ii. 'Zul HAS run.'

b. *Zul ada* (*-lah*) *tahu.*
Z. AUX -FOC know
i. #'Zul knows.'
ii. 'Zul HAS known.'

c. *Zul ada* *(*-lah*) *guru=nya.*
Z. AUX -FOC teacher=3
i. 'Zul is the teacher.'
ii. #'Zul HAS been the teacher.'

Similarly, none of the morphemes in *menjadi* and *merupakan*, i.e. the *meŋ-* prefix and the *-kan* suffix, may be separated from the verb in the double-copula construction. When *merupakan* or *menjadi* is used as a meaningless linking verb with copula *adalah* as its clausemate, the morphemes are obligatory.

(103) a. *Ke-jaya-an adalah* *(*meŋ-*)*rupa* *(*-kan*) *matlamat=nya.*
NMZ-succeed-NMZ COP ACT- form -APPL goal=3
'Success is his goal.'

b. *Ke-jaya-an adalah* *(*meŋ-*)*jadi* *matlamat=nya.*
NMZ-succeed-NMZ COP ACT- become goal=3
'Success is his goal.'

Conversely, when the two linking verbs are used as meaningful verbal predicates, the affixes may be omitted. Thus, in addition to their semantic vacuousness, this phenomenon demonstrates a morphological difference between their predicative and non-predicative uses.

(104) a. *Dia* (*meŋ-*)*rupa* (*-kan*) *arca dengan teknik papier-mâché.*
3.SG ACT- form -APPL sculpture with technique *papier-mâché.*
'He formed a sculpture with the papier-mâché technique.'

b. *Dia* (*meŋ-*)*jadi arca dengan teknik papier-mâché.*
3.SG ACT- become sculpture with technique *papier-mâché.*
'He became a sculpture with the papier-mâché technique.'

To summarise, the first criterion does not exclude linking verbs from being included in the characterisation of copulas in this thesis, as in the case of *merupakan* and *menjadi* in nonverbal predication in Malay. The syntactic category of the item in question plays no role in the characterisation of copulas. This is crucial so as to be inclusive of the copulas of other languages, which do not form a homogenous group and may comprise different syntactic categories.

1.6.3 Other Non-Copulas

The second criterion stipulates different semantics according to the type of predication and acts more like a filter to the first criterion than as a definition. The first criterion does not differentiate between copulas and many of the auxiliaries that carry TAM information, such as the modals and aspectual markers in Malay, especially since they can occur in verbless copular clauses, as shown below:

(105) *Dia mesti guru.*
3.SG must teacher
'He must be a teacher.'

Specifying that what is to be identified as a copula should have different semantic properties in verbal vs. nonverbal clauses, if its use in both types of predication is at all possible, makes sure that various auxiliaries, particles, or affixes that are likewise used in verbal clauses are not to be mistaken for copulas. Having different semantic properties in verbal predicates indicates that what we are dealing with is an item that is formally syncretic with the copula in nonverbal predication but is not essentially a copula. Of course, this criterion does not apply to languages that have no copula-like item in verbal clauses, e.g. *ialah* in Malay and *émkap* in Toba Batak, which exclusively occur in nonverbal clauses. For example, unlike *adalah*, *mesti* (must) carries the same modal information regardless of whether the clause is verbal. Therefore, although *mesti* passes the first criterion, it does not pass the second criterion and cannot be identified as a copula.

(106) a. *Dia mesti baca buku.*
3.SG must read book
'He must read books.'

b. *Dia mesti se-orang guru.*
3.SG must one-CLF teacher
'He must be a teacher.'

c. *Dia ada* (*-lah*) *baca buku.*
3.SG AUX -FOC read book
i. #'He read a book.'
ii. 'He HAS read a book.'

d. *Dia ada* *(*-lah*) *se-orang guru.*
3.SG ADA -LAH one-CLF teacher
i. 'He is a teacher.'
ii. #'He HAS been a teacher.'

This criterion also excludes items that might be a cause of confusion. For example, the particle *i* in Tok Pisin appears to be a copula at first, but it is used with verbal predicates in general as a predicate marker (Verhaar, 1991; Woolford, 1979):

Tok Pisin (English Creole – Melanesian Pidgin)

(107) a. *Em i kam.* (Verbal)
3.SG PRED come
'S/he comes.'

b. *Madang i bik.* (Nonverbal)
M. PRED big
'Madang is big.'
(Smith, 2008)

Besides that, markers of topic, focus, and other information-structural phenomena, e.g. Malagasy *dia* and Roviana *si*, present a difficult case in the study of copulas across languages as they are often analysed to be copulas despite their use with verbal predicates as well, due to their copula-like appearance, especially since they do not carry lexical meaning or information regarding TAM. For instance, Tagalog *ay* is often referred to as a copula in reference grammars such as Totanes (1745). It is used to mark the topic of a clause without any difference whatsoever in verbal or nonverbal clauses.

Tagalog (Austronesian – Philippine)

(108) a. *Si Mary ay nag-turo ng Ingles.* (Verbal)
ABS M. TOP PRF-teach OBL English
'Mary taught English.'

b. *Si Mary ay guro.* (Nonverbal)
ABS M. TOP teacher
'Mary is a teacher.'

Although copulas often evolve from discourse markers, as observed by Stassen (2013) with regard to particle copulas, these precursors of the copulas do not exclusively occur with nonverbal predicates, but also with verbal ones without any difference in meaning between the two types. Since they seldom show differences in meaning between verbal or nonverbal clauses, the second criterion prevents their identification as copulas and disambiguates them simply as discourse markers.

To conclude, the first criterion that I have provided is quite broad to allow the identification of many words across languages as copulas regardless of their form or function. Meanwhile, the second criterion, which acts as a filter upon the first, is narrow enough to exclude various words that are more appropriately labelled as other things, such as modals, auxiliaries, or discourse markers, owing to their use in both verbal and nonverbal predication without any difference.

Chapter 2: The Malay Copulas

Following from the characterisation of *ialah* and *adalah* as copulas in Malay, it is then imperative to investigate their individual properties. Each copula has different functions and a different distribution in different types of copular clauses. Prior to this thesis, no explicit analyses had been put forth about the semantic, information-structural, and morphosyntactic properties of the two items and how they differ (whether at all) from the morphemes from which they are speculated to have evolved, namely *ia-lah* (3.SG-FOC) and *ada-lah* (EXIST-FOC) (Yap, 2007).

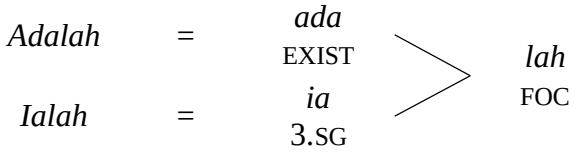


Figure 12: The Supposed Morphology of the Copulas

This chapter is an explication of the individual copulas. I argue that the two items are unlike their bimorphemic counterparts *ia-lah* (3.SG-FOC) and *ada-lah* (EXIST-FOC).[26] Although the copulas are believed to have evolved from the focal 3rd person pronoun and existential verb, they no longer associate with their former bimorphemic forms on all levels of linguistic analysis. In other words, they are monomorphemic; *lah* may no longer be separated from *ada* or *ia* as it has demorphologised such that "the earlier prefixes are now simply part of the phonology of the … stem" (Hopper & Traugott, 2003, p. 173). Other than the two copulas, I also examine in passing demonstrative *itu*. Although it is not recognised as a copula in traditional grammars, its use in copular clauses is akin to the other copulas, especially in that it is neither referential nor deictic, despite its guise as a demonstrative.

[26] The non-copular form will henceforth be written with a dash (-) to denote the separate morphemes of the item, whereas the copular form will be written without it:

ia-lah (3.SG-FOC)
ialah (COP)
ada-lah (AUX-FOC/EXIST-FOC/have-FOC)
adalah (COP)

2.1 *Ialah*

The copula *ialah* is commonly assumed to have evolved from 3rd person *ia* + focus marker *lah*. In spite of its bimorphemic appearance, it no longer associates with the morphemes from which it developed. In the present stage of the language, *ialah* is no longer a focused pronoun.

By showing that the syntax, morphology, semantics, and information structure of *ialah* does not transparently reflect those of the individual morphemes from which it developed, I argue that copula *ialah* has lost features associated with its homonymous pronominal counterpart and that it is currently monomorphemic. Simply put, *ialah* is disjoint from other occurrences of *ia* and its combination with *lah*.

2.1.1 Semantics

Ialah may function as a semantically vacuous copula. It does not contribute to the meaning of the sentence at all, as is apparent in example (109) in which the copula is shown to be optional. With or without the copula, the interpretation of the sentence remains the same.

(109) *Nama saya* (*ialah*) *Ali*.
name 1.SG COP A.
'My name is Ali.'

Copula *ialah* does not alter the interpretation of the clause because it is featurally impoverished. This is unlike other PRON-FOC forms, which cause changes in the interpretation of the clause due to their φ-features and their referentiality.

Concerning φ-features, it is important to remark that 3rd person *ia* is exclusively used with non-human referents in Modern Malay. As illustrated in (110), it is not possible for *ia* to be used with a human subject.[27]

[27] In previous stages of the Malay language, *ia* was used with both human and non-human referents, and the use of *ia-lah* as an argument was common. The following example in Classical Malay shows the use of *ia-lah* as an argument of a copular clause:

(vii) *Yang ber-gelar Amirulmukminin pun ia-lah.*
COMP INTR-name A. TOP 3.SG-FOC
'Who is named *Amirulmukminin* is him.'
(Bustan Al-Salatin – 17th century AD)

Context: *Ada nampak adik/buku/kucing saya tak?* (Have you seen my sibling/book/cat?)

(110) a. *Entah ke mana* { *dia* /**ia* } *pergi.* (*Dia* = Sibling)
not-know to where 3.SG.H 3.SG.NH go
'I don't know where he/she went.'

b. *Entah ke mana* { *ia* /* *dia* } *pergi.* (*Ia* = Book/Cat)
not-know to where 3.SG.NH 3.SG.H go
'I don't know where it went.'

The non-human feature of *ia* entails that a mismatch in humanness between the subject and *ialah* should result in some degree of ungrammaticality in copular clauses as well, should copula *ialah* transparently reflect the bimorphemic 3.SG.NH-FOC form. However, this prediction is not borne out as the apparent mismatch between the human subject and the presumably non-human *ialah* does not render example (111) ungrammatical. What this means is that *ialah* is unambiguously a copula when it combines with a human subject, as opposed to a PRON-FOC form.

(111) *Lelaki itu ialah Zul.*
man DIST COP Z.
'That man is Zul.'

On the other hand, when human *dia-lah* is used with an inanimate subject, the sentence is ill-formed, as in (112a). The contrast between (111) and (112) demonstrates that *ialah*, as a copula, is featurally impoverished. It has no φ-features encoding humanness. The opposite is true of *ia*, as it reveals its φ-features, as do all the other pronouns.

(112) a. **Benda itu dia-lah se-batang pokok.*
thing DIST 3.SG.H-FOC one-CLF tree
(That thing is a tree.)

b. *Benda itu ialah se-batang pokok.*
thing DIST COP one-CLF tree
'That thing is a tree.'

With respect to referentiality, copula *ialah* does not refer, despite its origins as a pronoun. The fact that the copula is semantically meaningless and the finding that it can combine with human subjects is sufficient evidence for this claim. To further consolidate the findings so far, consider the following examples:

(113) a. { *Saya* / *kamu* / *dia* / *ia* } *meŋ-cium Ali.* (Subject)
1.SG 2.SG 3.SG 3.SG.NH ACT-kiss A.
'I/you/she/it kissed Ali.'

b. { *Saya-lah* / *kamu-lah* / *dia-lah* /**ia-lah* } *yang meŋ-cium Ali.*[28]
1.SG-FOC 2.SG-FOC 3.SG.H-FOC 3.SG.NH-FOC COMP ACT-kiss A.
'I/YOU/SHE/IT kissed Ali.'

c. *Ali meŋ-cium* { *saya* / *kamu* / *dia* /**ia.* } (Object)
A. ACT-kiss 1.SG 2.SG 3.SG 3.SG.NH
'Ali kissed me/you/her/it.'

d. *Ali meŋ-cium* { *saya-lah* / *kamu-lah* / *dia-lah* /**ia-lah.* }
A. ACT-kiss 1.SG-FOC 2.SG-FOC 3.SG.H-FOC 3.SG.NH-FOC
'Ali kissed ME/YOU/HER/IT.'

Examples (113) reveal that *ia* is only acceptable as a non-focal subject. Instead of the pronoun, the full DP must be used in the ungrammatical contexts. Although *ia* may be used as a subject, it resists being focused. In other words, the combination *ia* + *lah* is illicit when used as an argument, which entails that *ialah* exists exclusively as a copula. Furthermore, when *ia* functions as the object, it needs to cliticise onto the verb as *-nya* (which is unspecified for humanness and number). Unlike 3rd person human *dia*, *ia* cannot occur as a free-standing object, as illustrated below:

(114) *Ali meŋ-cium* { *=nya* / *dia* /**ia.* }
A. ACT-kiss =3 3.SG.H 3.SG.NH
'Ali kissed her/it.'

The infrequent occurrence of *ia* is due to the fact that the pronoun has fallen out of fashion in Modern Malay and has been replaced by *dia*, which used to be the oblique 3rd person pronoun *di-ia* (PREP-3.SG). It is now only used in formal Malay and can be found to be used as an expletive, albeit very rarely, as shown below:

(115) *Ia penting untuk meŋ-tingkat-kan ke-faham-an pelajar…*
3.SG.NH important for ACT-level-APPL NMZ-understand-NMZ student
'It is important to heighten students' understanding…'
(Mohamad Yatim, 2021)

[28] Subject foci affixed by *lah* must be clefted, as indicated by separation of the subject from the rest of the clause by *yang*.

Meanwhile, none of the other PRON-FOC forms can function as a copula because they are referential, i.e. they are interpreted as arguments within a copular clause. As illustrated in example (116), the use of *dia-lah* as a copula is infelicitous as it is interpreted as a pronoun. The only interpretation available for example (116) is that in which *lelaki itu* is a dislocated topic and the antecedent of *dia-lah*, which functions as the referential subject of the copular clause with a null copula. It is only when *ialah* is used that the interpretation of a non-focal copular clause is obtained, given that it is non-referential.

(116) *Lelaki itu dia-lah Zul.*
man DIST 3.SG.H-FOC Z.
i. 'That man, HE is Zul.'
ii. #'That man is Zul.'

Because all the PRON-FOC forms are referential, they get interpreted as arguments of copular clauses, unlike the meaningless copula *ialah*. This contrast shows how *ialah* and the other PRON-FOC forms are disjoint semantically. *Ialah* is featurally impoverished and semantically vacuous, whilst the other PRON-FOC forms carry meaning and can cause a change in the interpretation of a copular clause. Copula *ialah* no longer carries the features [+Referential] and [±Human].

2.1.2 Information Structure

Lah is a commonly used as a marker of focus. As such, it should be possible to see differences in the information-structural properties of a copular clause when a PRON-FOC when used in a copular clause. Indeed, when *dia-lah* is used in a copular clause, it is interpreted as a focal subject due to its referential and focal features, e.g. (116). Conversely, *ialah* does not inherit any information-structural role despite being an element with *lah*. The copula allows the constituents flanking it to carry the appropriate information-structural roles. To illustrate, example (117) expresses no change with *ialah*. Despite *lah*, the sentence does not obtain the dislocation + focalisation interpretation observed in (116) with the PRON-FOC form *dia-lah*.

(117) *Lelaki itu ialah Zul.*
man DIST COP Z.
i. #'That man, HE is Zul.'
ii. 'That man is Zul.'

(118a-b) show the difference in information-structural partitioning between copular clauses with copula *ialah* and PRON-FOC forms. When the PRON-FOC form is used, the information-structural partitioning of the copular clause is altered.

(118) a. [Subject/Topic *Lelaki itu*] *ialah* [Focus *Zul.*] (Copula)
man DIST COP Z.
'That man is Zul.'

b. [Topic *Lelaki itu*$_1$] [Subject/Focus *dia*$_1$*-lah*] *Zul.* (PRON-FOC)
man DIST 3.SG.H-FOC Z.
'That man, HE is Zul.'

This change in information structure is observed of all the PRON-FOC forms but not *ialah*, which demonstrates that *ialah* is the only one that is different from the other elements in that it may not host focus. In fact, it is not possible for *ialah* to host verum focus even by pitch accent, as argued by Mustaffa (2018). This is in accordance with the fact that 3rd person non-human *ia* resists focus and may be used as an expletive, as demonstrated in examples (113) and (115). In order to assert the truth value of a copular clause, some other element needs to be inserted as reinforcement, e.g. an adverb, as shown below:

(119) *Lelaki itu* *(*pasti*^{verum}) *ialah Zul.*
man DIST sure COP Z.
'That man SURELY is Zul.'

The inability of the copula to host focus also means that it cannot itself be affixed by the focus marker. In fact, several other auxiliaries also behave likewise. For example, whilst it is possible for modals to combine with *lah*, certain aspectual markers in Malay cannot be affixed by the focus marker, as shown below. This non-focal behaviour sets copula *ialah* apart from verbs and puts it in the same class as the aspectual marker, as a form of auxiliary.

(120) a. *Lelaki itu boleh* (*-lah*) *jumpa Zul.* (Modal)
man DIST can FOC meet Z.
'That man CAN meet Zul.'

b. *Lelaki itu akan* (*-*lah*) *jumpa Zul.* (Aspectual Marker)
man DIST PROS FOC meet Z.
'That man WILL meet Zul.'

c. *Lelaki itu ialah* (**-lah*) *Zul.* (Copula)
man DIST COP FOC Z.
'That man is Zul.'

Despite its inability to host focus, copula *ialah* is unable to co-occur with other elements affixed by the focus marker. This restriction is not out of the ordinary for verbal clauses, as there can only be one focus marker in a clause, given that a clause may only accommodate one FocP, as opposed to multiple TopPs, according to Rizzi (1997). As the head of FocP in the left periphery, a constituent to be focused must move to FocP to host *lah*. To illustrate, *lah* cannot attach to the verb in example (121a) as the first occurrence of the focus marker on the clefted constituent disables the merging of another occurrence of the same head in the same clause. Meanwhile, example (121b) demonstrates that one focus marker can occur in each clause.

(121) a. [*Lelaki itu-lah yang jumpa* (**-lah*) *Zul.*]
man DIST-FOC COMP meet FOC Z.
'It is THAT MAN who met Zul.'

b. [*Ali-lah yang ber-kata bahawa* [*lelaki itu-lah yang jumpa Zul.*]]
A.-FOC COMP INTR-say COMP man DIST-FOC COMP meet Z.
'It is ALI who said that it is THAT MAN who met Zul.'

What might be surprising is that *ialah* is also subject to this restriction, which suggests that *ialah* is somehow affected by focus marker *lah*, in spite of its non-focal property.

(122) *Lelaki itu-lah* (**ialah*) *Zul.*
man DIST-FOC COP Z.
'THAT MAN is Zul.'

However, *ialah* is actually governed by a different and independent restriction that disallows its occurrence in clauses whose constituents have undergone Ā-movement in general, not just focus-related phenomena. This restriction is contrasted with A-movement, which does not disallow the copula from surfacing, as argued in Section 3.2 in examples (185). To illustrate, the clauses in the examples below all relate to Ā-movement, in all of which the copula is banned from surfacing:[29]

[29] Copular clauses with *ialah* cannot be relativised due to a restriction on extraction, as explained in Section 3.3.4.

(123) a. *Siapa*$_1$ t$_1$ (**ialah*) *Zul?* (Wh-Movement)
who$_1$ COP Z.
'Who is Zul?'

b. *Siapa*$_1$ *lelaki itu* (**ialah*) t$_1$?
who$_1$ man DIST COP
'Who is that man?'

c. *Lelaki itu*$_1$ *kan,* t$_1$ (**ialah*) *Zul.* (Topicalisation)
man DIST$_1$ TOP COP Z.
'That man, is Zul.'

d. *Zul*$_1$ *kan, lelaki itu* (**ialah*) t$_1$.
Z. TOP man DIST COP
'Zul, that man is.'

That this restriction on copula *ialah* is different from the restriction that there may only be one instance of the focus marker per clause is supported by the following examples, which illustrate the grammaticality of the copula in a sentence in which the focus remains in-situ. Notice that the verbal clause remains ungrammatical due to the double occurrence of *lah*.

(124) a. *Lelaki itu ialah Zul-lah.*
man DIST COP Z.-FOC
'That man is ZUL.'

b. *Lelaki itu jumpa-lah Zul* (**-lah*).
man DIST meet-FOC Z. -FOC
'That man DID meet Zul.'

If the copula were interpreted as a PRON-FOC form, the restriction should have applied to the copular clause and rendered the second instance of the focus marker in clause-final position ungrammatical, as in the verbal example. Therefore, it holds that *ialah* is a non-focal copula and not a pronoun affixed by focus marker *lah*, revealing that it does not carry the feature [+Focus].

2.1.3 Morphology

Copula *ialah* is monomorphemic. Otherwise, separating *ia* and *lah* should be possible, given that *lah* is not an obligatory element – pitch accent alone on the constituent to be focused is sufficient to assign focus to it. On the contrary, doing so is impossible. As illustrated in (125), *lah* may be omitted from the verb, but not from the copula.

(125) a. *Lelaki itu benci* (*-lah*) *Ali*. (Verb)
man DIST hate -FOC) A.
'That man HATES Ali.'

b. *Lelaki itu ia* *(*-lah*) *Ali.* (Copula)[30]
man DIST IA -LAH A.
'That man is Ali.'

Ia and *lah* can be analysed to have fused into a single morpheme, preventing the copula from being decomposed. Due to this non-decomposability, it is not active in any morphological processes, e.g. it does not combine with other morphemes like other pronouns can. For example, when serving as an object, *ia* must cliticise onto verbs, unlike copula *ialah*, which remains separate.

(126) a. *Meŋ-baca ialah meŋ-timba ilmu.* (Copula)
ACT-read COP ACT-gain knowledge
i. 'Reading is gaining knowledge.'
ii. #'READING it is gaining knowledge.'

b. *Meŋ-baca=nya-lah meŋ-timba ilmu.* (PRON-FOC)
ACT-read=3-FOC ACT-gain knowledge
i. 'READING it is gaining knowledge.'
ii. #'Reading is gaining knowledge.'

The obligatory encliticisation of pronoun *ia* also applies to possessives. Cliticisation of 3rd person *ia* occurs on preceding nouns to denote possession, which is certainly not the function of the copula in Malay. In the following example, *ia-lah* is able to cliticise onto the preceding DP *ibu* (mother) to mark possession, unlike the common behaviour of a copula. The fact that *ialah* remains separate in the following examples provides further evidence of the non-pronominal status of copula *ialah*.

[30] Gloss not provided for *ialah* here and in other relevant examples to remain agnostic to the morphological status (w.r.t their being discrete morphemes or otherwise) and function (as copula or focused pronoun) of *ia*, *lah*, and *ialah*.

(127) a. *Nama ibu ialah Ani.* (Copula)
name mother COP A.
i. 'Mother's name is Ani.'
ii. #'His MOTHER'S name is Ani.'

b. *Nama ibu=nya-lah Ani.* (PRON-FOC)
name mother=3-FOC A.
i. #'Mother's name is Ani.'
ii. 'His MOTHER'S name is Ani.'

In the formation of polar interrogatives, an auxiliary gets affixed by interrogative marker *kah*, which is argued by Kader (1981) to be the interrogative equivalent of focus marker *lah*, e.g. *boleh-kah* (can-Q), *mesti-kah* (must-Q), etc. If copula *ialah* were morphologically decomposable, one would expect the form *ia-kah* (3.SG.NH-Q) to be possible.[31] However, a polar interrogative copular clause uses *ada-kah* (AUX-Q) to replace the copula altogether, which suggests that *lah* cannot be replaced by *kah*. Therefore, the dummy auxiliary is used instead.

(128) *Ada-kah lelaki itu Ali?*
AUX-Q man DIST A.
'Is that man Ali?'

The observations presented above reinforce the argument that the copula is not bimorphemic, i.e. not composed of 3rd person *ia* and focus marker *lah*. Owing to the fusing of the two morphemes into a single one, it is no longer decomposable. Therefore, *ia* and *lah* cannot be separated and *ialah* does not behave the way 3rd person *ia* does with regard to cliticisation in possessives and the object position.

[31] It must be noted that the form *ia-kah* does exist, but not as a copula. It is the interrogative form of affirmative marker *ya*. Below are examples of *ia-kah* in verbal and nonverbal questions:

(viii) a. *Ia-kah lelaki itu Ali?*
AFF-Q man DIST A.
'Is it true that man is Ali?'

b. *Ia-kah lelaki itu benci Ali?*
AFF-Q man DIST hate A.
'Is it true that man hates Ali?'

2.1.4 Distribution

A difference in distribution between *ialah* and the PRON-FOC forms is observed when the argument status of the item in question is taken into consideration. Since *ialah* is not pronominal, it may not function as an argument. Therefore, it never occurs in positions in which a nominal element typically occupies, such as subject and object positions. Conversely, the PRON-FOC forms are free to occur in any such positions.

	Clause-Initial (Subject)	Clause-Medial (Verb/Auxiliary)	Clause-Final (Object/Complement)
PRON-FOC	✓	-	✓
Copula	-	✓	-

Table 7: The Difference in Distribution between Copula 'Ialah' and PRON-FOC Form

As shown below, *ialah* may not occur in the clause-initial and -final positions of the copular clauses. On the other hand, a PRON-FOC form is perfectly grammatical in both positions. Interestingly, a PRON-FOC form may not occur between two DPs in what is known as a specificational copular clause, as in (129f). The copular clause does not get interpreted as a topical left-dislocation construction with a focal subject, as in (118b), due to the non-referentiality of the subject. As argued by Mikkelsen (2005a), this type of copular clause has a fixed topic-focus alignment, which does not allow the PRON-FOC form to be focal. The use of the PRON-FOC form violates the topic-focus alignment, rendering the examples ungrammatical.

(129) a. **Ialah salah se-jenis haiwan ter-ancam.* (Copula)
COP one.of one-CLF animal NVOL-threat
(IT is one of the types of threatened animals.)

b. **Salah se-jenis haiwan ter-ancam ialah.*
one.of one-CLF animal NVOL-threat COP
(One of the types of threatened animals is IT.)

c. *Salah se-jenis haiwan ter-ancam ialah orangutan.*
one.of one-CLF animal NVOL-threat COP orangutan
'One of the types of threatened animals is the orangutan.'

d. *Dia-lah salah se-orang rakan=nya.* (PRON-FOC)
3.SG.H-FOC one.of one-CLF friend=3
'HE is one of the friends.'

e. *Salah se-orang rakan=nya dia-lah.*
one.of one-CLF friend=3 3.SG.H-FOC
'One of the friends is HIM.'

f. **Salah se-orang rakan=nya dia-lah Zul.*
one.of one-CLF friend=3 3.SG.H-FOC Z.
(One of the friends, HE is Zul.)

Also related to information structure and the different distribution between the copula and the PRON-FOC forms is the fact that *ia-lah* does not exist as a PRON-FOC form, due to the impossibility of focalising 3rd person *ia*. This finding alone is sufficient to capture the exclusive copular status of *ialah*. As a non-focal element, it should not be possible for *ia* or *ia-lah* to serve as the focus of a cleft. As contrasted with *dia-lah* below, *ia-lah* truly cannot function as the focus of a cleft construction:

(130) a. *Dia-lah yang telah meŋ-sebab-kan rusuhan itu.*
3.SG.H-FOC COMP PRF ACT-cause-APPL riot DIST
'It is HE who has caused the riot.'

b. **Ialah yang telah meŋ-sebab-kan rusuhan itu.*
IALAH COMP PRF ACT-cause-APPL riot DIST
(It is IT that has caused the riot.)

If *ia-lah* were similar in distribution with *dia-lah*, the total number of occurrences of both forms should be about the same in a corpus. On the contrary, the use of *ia-lah* in non-cleft constructions far exceeds that of the other focused pronouns. It is apparent in the corpus of Utusan newspapers by Dewan Bahasa dan Pustaka (2013), that there are only two instances of clefts with *ia-lah* as the clefted constituent.

PRON-FOC Form	Total occurrences in corpus	Total occurrences as clefted constituent[32]	Percentage
Saya-lah (1.SG-FOC)	27	7	26%
Kami-lah (1.PL.INCL-FOC)	2	1	50%
Kita-lah (1.PL.EXCL-FOC)	24	10	42%
Kamu-lah (2.SG-FOC)	3	1	33%
Ia-lah (3.SG.NH-FOC)	12,834	2	0.02%
Dia-lah (3.SG.H-FOC)	41	17	41%
Mereka-lah (3.PL-FOC)	68	36	53%

Table 8: The PRON-FOC Forms as Clefted Constituent

[32] The clefted constituent occurs in the following pattern: (adjunct) [X-*lah*][*yang*-clause]. Examples in which the pronoun occurs possessively – [NP X]-*lah* – are not counted.

The sheer number of instances of the string "ialah" compared to the other forms in the table above is due to its occurrence as a copula in the pattern "[DP] *x* [DP]" in the corpus. Despite the very limited use and obsolescence of *ia* in Modern Malay (due to it having been replaced by pronoun *dia*), its *lah* form is by far the most frequently occurring focused pronoun-like form in the corpus. The other forms are not even remotely as numerous, as their distribution is limited to environments of focus such as clefts. The striking difference reflects the far higher frequency of copular clauses, whose copula happens to have the same form as the focused non-human 3rd person pronoun in Malay, as compared to clefts with actual PRON-FOC forms. The absence of the other forms in "[DP] *x* [DP]" position constitutes evidence that they are not copulas.

When epistemic modals are concerned, whilst they are free to occur with the PRON-FOC forms, since they occupy different positions in the syntax, their occurrence in copular clauses with *ialah* renders the sentence ungrammatical, due to complementary distribution, as shown below:

(131) a. *Lelaki itu dia-lah mesti jadi yang ter-pilih.*
man DIST 3.SG.H-FOC must become REL NVOL-choose
'That man, HE must be the chosen one.'

b. **Lelaki itu ialah mesti yang ter-pilih.*
man DIST IALAH must REL NVOL-choose
'That man must be the chosen one.'

Due to the non-referential and non-focal nature of copula *ialah,* it is found in positions that are totally different from the PRON-FOC forms. Unlike the pronouns that the PRON-FOC forms are, the copula may not occur in the subject and object positions of a clause, as well as any other position that is typically occupied by a nominal constituent. Additionally, the non-focal nature of *ia* and *ialah* prevents it from occurring in positions of focus, such as clefts.

2.2 *Adalah*

I argue here that copula *adalah* is also monomorphemic and disjunct from the other forms of *ada* and their combinations with focus marker *lah*, based on data pertaining to its semantics, information structure, morphology, and distribution. Particularly, copula *adalah* is not a focused form of *ada*, which may refer to one of the following items: an existential/locative verb, a possessive verb, or a (focal) aspectual auxiliary. The variants are shown in their respective constructions below:

(132) a. *Ada-lah se-buah kereta tadi.* (Existential)
EXIST-FOC one-CLF car just.now
'There WAS a car just now.'

b. *Saya ada-lah se-buah kereta.* (Possessive)
1.SG have-FOC one-CLF car
'I DO have a car.'

c. *Saya ada-lah pandu kereta itu.* (Aspectual)
1.SG AUX-FOC drive car DIST
'I DID drive that car.'

d. *Kereta ini adalah mesra alam.* (Copular)
car PROX COP friendly environment
'This car is environmentally friendly.'

2.2.1 Semantics

Examples (132a-d) are a clear indication of the different semantics of the variants. Each of them carries its own semantics, from existential to verum or aspectual meaning. Conversely, copula *adalah* does not contribute to the meaning of the clause.

The item that most closely resembles the copula is aspectual *ada*, since they both are auxiliaries that occur between a subject and a predicate. However, aspectual *ada* occurs in verbal clauses and adds a dimension of aspect, whilst the copula occurs in copular clauses and does not contribute any meaning to the semantics of the copular clause. Therefore, the deletion of the aspectual auxiliary in example (132c) should cause a change in meaning, unlike the deletion of the copula in (132d). As expected, deletion of the aspectual auxiliary is shown below to have caused the sentence to lose the perfective or past-tense reading. Conversely, deletion of the copula results in no change to the interpretation of the sentence.

(133) a. *Saya ~~ada-lah~~ pandu kereta itu.* (Aspectual)
1.SG AUX-FOC drive car DIST
'I drive that car.'
#'I DID drive that car.'

b. *Kereta ini ~~adalah~~ mesra alam.* (Copular)
car PROX COP friendly environment
'This car is environmentally friendly.'

As the main predicates of their sentences, existential and possessive *ada* cannot be deleted. Unlike aspectual *ada-lah* and copula *adalah*, deletion would render the sentences ungrammatical since the two variants express their own semantics:

(134) a. *~~*Ada-lah~~ *se-buah kereta tadi.* (Existential)
EXIST-FOC one-CLF car just.now
(There was a car just now.)

b. *Saya ~~ada-lah~~ *se-buah kereta.* (Possessive)
1.SG have-FOC one-CLF car
(I have a car.)

2.2.2 Information Structure

Due to the optional focus marker on the non-copular forms of *ada*, the truth value of the sentence in which each form occurs may be asserted. As shown in examples (135a-c), a verum focus interpretation is available when *lah* attaches to *ada* to form *ada-lah*. However, when it comes to the copula, this focus interpretation is not available. This unavailability of verum focus is due to the fact that the copula can neither be stressed nor host verum focus, a property identical to *ialah*.

(135) a. *Ada-lah*^{verum} *se-buah kereta tadi.* (Existential)
EXIST-FOC one-CLF car just.now
'There WAS a car just now.'

b. *Saya ada-lah*^{verum} *se-buah kereta.* (Possessive)
1.SG have-FOC one-CLF car
'I DO have a car.'

c. *Saya ada-lah*^{verum} *pandu kereta itu.* (Aspectual)
1.SG AUX-FOC drive car DIST
'I DID drive the car.'

d. *Kereta ini adalah mesra alam.* (Copular)
car PROX COP friendly environment
i. 'This car is environmentally friendly.'
ii. #'This car IS environmentally friendly.'

In order to assert the truth value of a copular clause, some other element needs to be inserted for reinforcement, e.g. an adverb, as shown below:

(136) *Kereta ini *(memang*[verum]*) adalah mesra alam.*
car PROX truly COP friendly environment
'This car truly IS environmentally friendly.'

Additionally, assigning verum focus to the copula is not possible in certain situations to avoid confusion between the copula and the other *ada*-FOC forms, as is made apparent when the copula is compared with the possessive verb. For instance, suppose that there is no context to tell whether *adalah* in example (137) is a possessive verb or a copula. If stressed, *adalah* can only have a possessive reading.

(137) *Saya adalah*[verum] *se-orang ibu.*
1.SG ADALAH one-CLF mother
i. 'I DO have a mother.' (Possessive)
ii. #'I am a mother.' (Copular)

2.2.3 Morphology

As is also true of copula *ialah*, copula *adalah* may not be decomposed into *ada* and *lah*. Thus, using *ada* as a copula without *lah* is ungrammatical. Conversely, the other 3 *ada*-FOC forms can function as usual regardless of whether they are affixed by *lah*, as shown in (138). This is because *lah* is merely used to convey new information or to supply an additional focus interpretation, such as contrastive focus. In other words, *lah* is an obligatory part of the copula but is optional on the aspectual auxiliary, existential and possessive verbs.

(138) a. *Ada se-buah kereta tadi.* (Existential)
EXIST one-CLF car just.now
'There was a car just now.'

b. *Saya ada se-buah kereta.* (Possessive)
1.SG have one-CLF car
'I have a car.'

c. *Saya ada pandu kereta itu.* (Aspectual)
1.SG AUX drive car DIST
'I did drive that car.'

d. **Kereta ini ada mesra alam.* (Copular)
car PROX ADA friendly environment
(This car is environmentally friendly.)

Furthermore, Mustaffa (2018) illustrates that copula *adalah* cannot be affixed by any verbal affix. It may not participate in any sort of morphological process, whether it be the addition or deletion of a morpheme. Contrastingly, it is possible for the other forms of *ada* to be affixed by both inflectional and derivational affixes, undergo reduplication, and host clitics. For example, they can be affixed by active voice marker *meŋ-* and applicative suffix *-kan*, reduplicated to yield a stative verb, derived into a nominal through affixation by nominalising affixes, and attached by clitics.

(139) a. *Mereka meŋ-ada-kan kenduri.* (Inflection)
3.PL ACT-EXIST-APPLfeast
'They are having a feast.'

b. *Ke-ada-an itu me-risau-kan saya.* (Derivation)
NMZ-EXIST-NMZ DIST ACT-worry-CAUS 1.SG
'That situation worries me.'

c. *Dia ada-ada.* (Reduplication)
3.SG have-RED
'He is well off.'

d. *Ada=nya kamu kerana ibu kamu.* (Cliticisation)
EXIST=3 2.SG because mother 2.SG
'Your being here is because of your mother.'

e. *Dia ada=nya nampak kamu.* (Emphatic Cliticisation)[33]
3.SG AUX=3 see 2.SG
'He DID see you.'

[33] This phenomenon refers to the cliticisation of the 3rd person clitic to mark emphasis. The clitic does not appear to be referential, as its person φ-feature does not match that of the 1st person subject, as shown below:

(ix) *Pandai=nya saya.*
smart=3 1.SG
'So smart, I am.'

None of these morphological processes are available to the copula, as the focus marker cannot be separated from the root to accommodate other affixes. Both addition and deletion of *lah* from the copula is impossible, making it clear that the copula is monomorphemic.

2.2.4 Distribution

The distribution of copula *adalah* and the other *ada*-FOC forms differs in terms of their type of predicate and syntax. The copula occurs in nonverbal clauses, whilst all the others occur in verbal clauses. As illustrated in Table 9, all four forms occur in different places in relation to the other constituents: existential *ada-lah* occurs at the beginning of the clause; possessive *ada-lah* intervenes between a possessor external argument and a possessee internal argument; aspectual auxiliary *ada-lah* precedes verbs; copula *adalah* intervenes between the subject and a nonverbal predicate.

	Subject	Auxiliary	Verb	Object/Complement
Existential	-	-	*Ada-lah* EXIST-FOC	*se-buah kereta* one-CLF car
Possessive	*Saya* 1.SG	-	*ada-lah* have-FOC	*se-buah kereta* one-CLF car
Aspectual	*Saya* 1.SG	*ada-lah* AUX-FOC	*pandu* drive	*se-buah kereta* one-CLF car
Copular	*McQueen* M.	*adalah* COP	-	*se-buah kereta* one-CLF car

Table 9: The Distributions of the Different Forms of 'Ada-lah'

In relation to negator *tidak*, copula *adalah* is found to be the odd one out of the other *ada*-FOC forms. *Tidak* follows the copula, but it precedes all the other *ada*-FOC forms.

It is also possible for pragmatic marker *lah* to co-occur with the 3rd person clitic on the same host, with a slight change in meaning. However, the 2 morphemes cannot switch places, so it is not possible for copula *adalah* to host *nya*, since *lah* cannot be removed.

(x) *Pandai=nya-lah saya.*
smart=3-FOC 1.SG
'So smart, I am.' (Narcissistic tone)

(140) a. *Tidak ada-lah se-buah kereta tadi.* (Existential)
NEG EXIST-FOC one-CLF car just.now
'There WASN'T a car just now.'

b. *Saya tidak ada-lah se-buah kereta.* (Possessive)
1.SG NEG have-FOC one-CLF car
'I DON'T have a car.'

c. *Saya tidak ada-lah pandu kereta itu.* (Aspectual)
1.SG NEG AUX-FOC drive car DIST
'I DIDN'T drive the car.'

d. *Kereta itu adalah tidak mesra alam.* (Copular)
car DIST COP NEG friendly environment
'That car is not environmentally friendly.'

Non-copular *ada* may co-occur with different auxiliaries, whereas the copula can never occur with auxiliary verbs.

(141) a. *Tidak akan boleh ada perlumbaan nanti.* (Existential)
NEG PROS can EXIST race later
'There never can be a race later.'

b. *Dia tidak akan boleh ada peluang untuk menang.* (Possessive)
3.SG NEG PROS can have opportunity for win
'He never can have an opportunity to win.'

c. *Dia tidak akan ada boleh menang.* (Aspectual)
3.SG NEG PROS AUX can win
'He never can win.'

d. **Perlumbaan itu boleh adalah malapetaka.* (Copular)
race DIST can COP disaster
(That race could be a disaster.)

In fact, *ada-kah* is able to co-occur with the existential and possessive verbs. However, it cannot co-occur with copula *adalah* because both forms are auxiliaries which are in complementary distribution.

(142) a. *Ada-kah ada kena-mengena dengan usia?* (Existential)
AUX-Q EXIST relation with age
'Is there a relation with age?'
(Abu Hassan, 2021)

b. *Ada-kah mereka ada dua.per.tiga majoriti pada masa kini?* (Possessive)
AUX-Q 3.PL have one.third majority at time now
'Do they have a one-third majority at present?'
(Yusoff Bakri, 2021)

c. **Ada-kah mereka adalah ahli politik?* (Copular)
AUX-Q 3.PL COP member politics
(Are they politicians?)

Polar interrogatives formed from existential and possessive clauses can be answered using the verb itself. This method of answering questions is similar to the way questions with other verbs are answered in Malay. Copular clauses are different as they cannot be answered using the copula. Instead, either the affirmative marker or the negator must be used.

(143) a. *Ada-kah kamu suka ber-lumba?* { *Suka.* / *Tak suka.* } (Verb)
AUX-Q 2.SG like INTR-race like NEG like
'Do you like to race?' 'Yes. / No.'

b. *Ada-kah perlumbaan nanti?* { *Ada.* / *Ti-ada.* } (Existential)
EXIST-Q race later EXIST NEG-EXIST
'Is there a race later?' 'Yes. / No.'

c. *Awak ada-kah kereta?* { *Ada.* / *Ti-ada.* } (Possessive)
2.SG have-Q car have NEG-have
'Do you have a car?' 'Yes. / No.'

d. *Ada-kah Ali kawan=nya*? {* *Adalah.* / *Ya.* / *Bukan.* } (Copular)[34]
AUX-Q A. friend=3 COP AFF CNTR
'Is Ali his friend?' 'Yes. / No.'

From the data in examples (143b-c), the observation that existential and possessive *ada* pattern with other verbs in that they can stand on their own as the answer to a question should additionally make it possible to find instances in which they are stranded in a clause as well.

(144) a. *Kalau tidak ada kereta, apa yang ada?* (Existential)
COND NEG EXIST car what COMP EXIST
'If there is no car, what is there?'

[34] Although *ada-kah* seems to be the copula in the question, the fact is that it is the dummy auxiliary. As argued in Chapter 2, the copula must be replaced by the dummy auxiliary altogether since *lah* cannot be separated from the copula to be replaced by *kah*.

b. *Kalau Ali tidak ada kereta, apa yang dia ada?* (Possessive)
COND A. NEG have car what COMP 3.SG have
'If Ali doesn't have a car, what does he have?'

c. *Kalau Ali tidak ada pandu kereta itu, Abu ada.* (Aspectual)
COND A. NEG AUX drive car DIST A. AUX
'If Ali did not drive the car, Abu did.'

However, similar to certain auxiliary verbs, e.g. aspectual marker *akan*, copula *adalah* cannot be stranded. It is obligatory for the copula to intervene between subject and predicate, unlike the other forms of *ada*, whose distribution varies.

(145) a. **Kalau saya adalah tidak penting siapa adalah?* (Copula)
COND 1.SG COP NEG important who COP
(If I am not important, who is?)

b. **Kalau awak tak-kan laku-kan=nya siapa akan?* (Aspect)
COND 2.SG NEG-PROS do-APPL=3 who PROS
(If you won't do it, who will?)

2.3 *Itu*

Malay seems to have a demonstrative that is used as a copula alongside the two other copulas. This demonstrative pronominal copula has not received any attention in the linguistics literature before, so it is not recognised as a copula. Its common use as a demonstrative makes judgements of its copular status in nonverbal clauses rather difficult. However, when used in nonverbal clauses, *itu* does appear to be a copula, especially because it does not carry the usual referential semantics and may invert with the subject.

2.3.1 Semantics

In terms of semantics, *itu* is vacuous, as are the other copulas in Malay. The lack of semantic content is surprising, given that the word normally supplies a DP with definiteness, or is referential and deictic on its own as a demonstrative pronoun in Modern Malay. To illustrate, the use of *itu* in the following example with a gerundive subject does not promote a definite interpretation of the DP subject. Moreover, it can be omitted, which entails that *itu* does not semantically or deictically refer to anything. Also, it can be substituted with an actual copula such as *adalah*.

(146) *Kalau meŋ-tangis* {*itu* / *adalah*} *perkara biasa…*
COND ACT-cry DIST COP thing normal
'If crying were something normal…'
(Che Noh, 2018)

A gerund in Malay does not warrant the use of a determiner the same way it does in English, as in the example "the crying of newly born babies". Besides, a gerund like *menangis* is a word that encodes an action, which is not referential, so it is obvious that *itu* does not function as a regular demonstrative in the example above. The deictic property of demonstratives does not pair with an abstract notion. The only way it is possible for *itu* to be used with *tangis* (cry) as a meaningful demonstrative is for the verb to be nominalised using nominalising suffix *-an* to form *tangisan*, as in the example below:

(147) *Kalau tangis-an itu perkara biasa…*
COND cry-NMZ DIST thing normal
'If that cry were something normal…'

It can therefore be said that *itu* in copular clauses does not have the features associated with its common demonstrative pronoun form. In the examples above, *itu* does not alter or add to the semantics of either constituent in the copular clause.[35]

2.3.2 Morphology

Demonstrative *itu*, as a determiner or pronoun, is morphologically invariant; however, it is observed to be able to combine with non-human 3rd person *ia* to form *iaitu*, which is used as an expression to provide clarification with respect to the preceding phrase or clause.[36] Its use this way is comparable to Latin "id est" or English "that/which is", which are copular clauses, as shown below:

---

[35] In Bahasa Indonesia, copular *itu* may be used with proximal demonstrative *ini* as the subject, which makes apparent the non-referential property of *itu*.

(xi) *Ini itu se-jenis buah.*
PROX DIST one-type fruit
'This is a type of fruit.'

[36] *Iaitu* most likely developed by analogy with *yakni*, which is likewise an expression for clarification. *Yakni* deceivingly appears to be a compound of *ia* (3.SG) and *ini* (PROX), but it is in fact a word borrowed from Arabic. The spelling (one beginning with <i> and the other with <y>) is irrelevant as both expressions were written in the Arabic script with ي /ja/ when *yakni* entered the language.

(148) a. *Bilirubin ia-itu pigmen ber-warna kuning…*
B. 3.SG-DIST pigment INTR-colour yellow
'Bilirubin, which is a yellow-coloured pigment…'
(Mahmud, 2021)

b. *Ter-dapat 299 kes ia-itu 222 kes di-sah-kan positif Covid-19…*
NVOL-get 299 case 3.SG-DIST 222 case PASS-true-APPL positive C.
'There were 299 cases, 222 cases of which are verified to be Covid-19 positive…
(N. A. Sulaiman, 2021)

2.3.3 Distribution

A strong indicator for the independence of *itu* from the subject or the predicate is its mobility. As shown in example (149) below, the subject of the interrogative clause is "murka" and *itu* originates in a position following the subject but has moved past it in the preceding interrogative clause. Essentially, the demonstrative has inverted with the subject, providing important evidence that it is a head that can undergo head movement, hence is neither a subject nor a resumptive pronoun.

(149) *Apa itu murka? ... Murka itu soalan anda.*
what DIST murka murka DIST question 2.SG
'What is *murka* (anger)?... *Murka* is your question.'
(R. Idris, 2018)

Furthermore, *itu* may lie outside the constituent marked by *kah* in a question. Given that demonstratives do not combine with *wh*-phrases in Malay, as illustrated in (150c), the position of *itu* following the interrogative DP in (150a) signals that it is not associated with the interrogative DP. If *itu* were part of the interrogative DP, it should precede *kah*, which should modify the whole DP, as in (150b).

(150) a. *Apa-kah itu melasma?*
what-Q DIST melasma
'What is melasma?'
(Abdul Rahim, 2016)

b. *Benda itu-kah melasma?*
thing DIST-Q melasma
'Is that thing melasma?'

c. **Benda-kah itu melasma?*
thing-Q DIST melasma
(Is that thing melasma?)

In the following example, the subject *namanya* has been extraposed, which demonstrates that *itu* forms a larger constituent with the predicate, and not the subject:

(151) t_1 *itu* *ber-bohong* *nama=nya*$_1$.
DIST INTR-lie name=3
'That is called lying.'
(Hussein, 2022)

The fact that it may co-occur with a gerundive subject in Malay is uncharacteristic of genuine demonstratives in terms of distribution. When a gerund functions as the object of a verb, the demonstrative becomes sharply ungrammatical, which makes it obvious that *itu* is not associated with either constituent.

(152) *Saya* *benci* *meŋ-tipu* (**itu*).
1.SG hate ACT-lie DIST
'I hate lying.'

In relation to the other copulas, the distribution of *itu* is wider than that of *ialah* or *adalah*. For example, it is observed to be able to occur in environments that are not accommodating to them, such as in interrogatives:

(153) *Namun* *tahu-kah* *kita* *apa* *makna* *tauhid* *itu?*
however know-Q 1.PL.INCL what meaning tauhid DIST
'However, do we know what the meaning of *tauhid* is?'
(Mohd Yusoff, 2016)

Further, example (153) shows the possibility of stranding *itu*, which is not possible with the other copulas. As shown in Section 2.2.4, stranding of the other copulas causes ungrammaticality.

(154) *Namun* *tahu-kah* *kita* *apa* *makna* *tauhid* { Ø / *itu* /* *ialah* /* *adalah*}?
however know-Q 2.PL.INCL what meaning tauhid COP DIST COP COP
'However, do we know what the meaning of *tauhid* is?'

Suppose that the subject of a copular clause with *itu* is a left-dislocated topic and *itu* is its resumptive pronoun of the likes below:

[$_{TopP}$ SUBJECT$_1$ [$_{FocP}$ AUX-*kah* [$_{TP}$ *itu*$_1$ [$_{PredP}$ PREDICATE]]]]

If the subject were left-dislocated and a resumptive pronoun were left in its place, it should be possible to see *adakah* intervening between the topicalised DP and *itu* in a polar question, given that topics and *adakah* occur in the CP above the TP whose specifier should be occupied by the resumptive pronoun. However, the occurrence of *adakah* preceding the subject of the copular clause in (155) indicates that the subject has not been left-dislocated or topicalised. Both the subject and *itu* are therefore analysed to be contained within the root TP.

(155) *Ada-kah sembelit itu normal?*
AUX-Q constipation DIST normal
'Is constipation normal?'
(Fadzlyana, 2017)

2.4 Coexistence of the Monomorphemic and Bimorphemic Forms

In the current stage of the language, the monomorphemic copula *adalah* coexists with the bimorphemic *ada*-FOC forms. According to Hopper (1991), this phenomenon in which older and newer meanings of the same item is used within the same time period is called *layering*.

Speakers seem to be well aware of the differences between the copular and non-copular uses of the string or representation "adalah". Formally, it can be said that there are different lexical entries for "adalah" in the mental grammars of speakers, similar to how "ada" exists as different entries with meanings that encode possession, existence, perfective aspect, or nothing at all (as in the case of the dummy auxiliary in *ada*-support).

Although speakers recognise the morpheme boundary between *ada* and *lah* in the verbal domain, they do not construe the copula *adalah* to be comprised of the two separate morphemes. Evidently, this knowledge of the bimorphemic nature of *ada-lah* in non-copular clauses does not cause interference in the production of the monomorphemic copula in copular clauses by diglossic speakers who mainly speak varieties with the bimorphemic variant but also have access to varieties that have the monomorphemic copula. As discussed with respect to examples (138), repeated below, the *lah* morpheme may be omitted in verbal clauses, but strictly not in copular clauses.

(156) a. *Ada se-buah kereta tadi.* (Existential)
EXIST one-CLF car just.now
'There was a car just now.'

b. *Saya ada se-buah kereta.* (Possessive)
1.SG have one-CLF car
'I have a car.'

c. *Saya ada pandu kereta itu.* (Aspectual)
1.SG AUX drive car DIST
'I did drive that car.'

d. **Kereta ini ada mesra alam.* (Copular)
car PROX ADA friendly environment
(This car is environmentally friendly.)

As for the string "ialah", no such layering is present in Modern Malay since *ia* is only acceptable as a non-focal subject, as described in examples (113), repeated below:

(157) a. { *Saya* / *kamu* / *dia* / *ia* } *meŋ-cium Ali.* (Subject)
1.SG 2.SG 3.SG 3.SG.NH ACT-kiss A.
'I/you/she/it kissed Ali.'

b. { *Saya-lah* / *kamu-lah* / *dia-lah* /**ia-lah* } *yang meŋ-cium Ali.*
1.SG-FOC 2.SG-FOC 3.SG.H-FOC 3.SG.NH-FOC COMP ACT-kiss A.
'I/YOU/SHE/IT kissed Ali.'

c. *Ali meŋ-cium* { *saya* / *kamu* / *dia* /**ia.* } (Object)
A. ACT-kiss 1.SG 2.SG 3.SG 3.SG.NH
'Ali kissed me/you/her/it.'

d. *Ali meŋ-cium* { *saya-lah* / *kamu-lah* / *dia-lah* /**ia-lah.* }
A. ACT-kiss 1.SG-FOC 2.SG-FOC 3.SG.H-FOC 3.SG.NH-FOC
'Ali kissed ME/YOU/HER/IT.'

Despite the knowledge of 3rd person *ia* and focus marker *lah*, the two morphemes cannot combine to either form focal 3rd person pronoun *ia-lah* or copula *ialah*. As a result, in no case or instance can the string "ialah" be split into *ia* and *lah*, which is concrete evidence for the monomorphemic nature of the copula *ialah* in the current stage of the language. Therefore, the string "ialah" only exists as a monomorphemic copula in Modern Malay.

2.5 Summary

In sum, copula *ialah* is different from all the PRON-FOC forms in terms of semantics, information structure, morphology, and distribution. Evidently, the surface string *ialah* is exclusively a copula in the current stage of the language, which shows absolutely no signs of being 3rd person non-human pronoun *ia* and its combination with focus marker *lah*. It carries no semantics, unlike the PRON-FOC forms, which are referential. Also due to the semantic vacuousness of the copula, it cannot carry focus despite the apparent focus marker on it. It is a monomorphemic element within which *ia* and *lah* cannot be separated. Lastly, due to its non-referential and non-focal nature, it may not occur in positions one would expect to find a nominal, such as a PRON-FOC form.

	PRON-FOC form	Copula *ialah*
Semantics	Meaningful [+Referential]	Meaningless [–Referential]
Information Structure	Focus possible	Focus impossible
Morphology	Decomposable Morphological processes possible	Non-decomposable Morphological processes impossible
Distribution	Argument/possessive position	Copular clauses: DP COP DP

Table 10: Summary of Comparison between Copula 'Ialah' and the PRON-FOC Forms

On the other hand, *adalah* is different from the other forms of *ada*, namely the existential verb, the possessive verb, and the aspectual auxiliary. The non-copular forms of *ada* pattern with each other in a number of ways to the exclusion of the copula, which shows that copula *adalah* is somewhat special. Copula *adalah* is semantically vacuous, unlike the existential and possessive verbs, which function as the main predicates of their sentences, and even aspectual *ada*, which carries perfective aspect. It cannot host verum focus, which is a property it shares with copula *ialah*. Despite its form with what seems to be focus marker *lah*, *ada* and *lah* cannot be separated from the copula, which prevents it from participating in any sort of morphological process, entailing its monomorphemic form. Finally, copula *adalah* is the only item that occurs in copular clauses between a subject and a nonverbal predicate and preceding negator *tidak*, unlike the other variants of *ada*, which occur in verbal clauses following *tidak*.

	Ada-FOC form	Copula *adalah*
Semantics	Meaningful	Meaningless
Information Structure	Focus possible	Focus impossible
Morphology	Decomposable Morphological processes possible	Non-decomposable Morphological processes impossible
Distribution	Verbal clauses Following NEG Stranding possible	Nonverbal clauses Preceding NEG Stranding impossible

Table 11: Summary of Comparison between Copula 'Adalah' and the Ada-FOC Forms

Ample empirical evidence has been provided to support the different properties between the copulas and the roots from which they originated, namely 3rd person *ia* and existential verb *ada*. The copulas do not reflect the semantics, information structure, morphology, and distribution of their homonymous counterparts. Therefore, these differences reveal that *ialah* and *adalah* are two items that have developed into copulas that are disjoint from other occurrences of 3rd person *ia*, apsectual/existential/possessive *ada* and their combinations with focus marker *lah*.

As for *itu*, it lies outside the maximal phrase of both the subject and the predicate of a copular clause, thus is a lexeme that is separate from either constituent. It does not occupy a position in the left periphery of the clause and has the same syntactic distribution and behaviour as auxiliaries as it is possible for it to undergo subject-auxiliary inversion in *wh*-interrogatives. Therefore, *itu* is a head, not a phrasal constituent.

	Demonstrative *itu*	Copula *itu*
Semantics	Meaningful	Meaningless
Morphology	Morphological processes impossible	Morphological processes possible
Distribution	Combines with subject Impossible with gerunds Preceding *kah*	Inverts with subject Possible with gerunds Following *kah*

Table 12: Summary of Comparison between Copula and Demonstrative 'Itu'

Chapter 3: The Syntax of Copular Clauses

This chapter presents the clausal structure of copular clauses in Malay. It begins with addressing some misconceptions about the copulas and sheds light on several issues in previous analyses of copular clauses in Malay. Several misleading claims about copular clauses in Malay, such as their syntactic structure and the elements with which they may combine, are reviewed and set straight by presenting data that more accurately reflect the reality of copular clauses in real-world use.

The actual position of the copulas in the structure of a copular clause is identified in this chapter, as previous studies have not made explicit claims about it but have merely assumed that they are verbal heads. For example, copula *adalah* is assumed to head VP in Arka (2013). Although it is argued to be merged in *v*P and moved to T^0 in Mustaffa (2018), there is evidence pertaining to the structure of the extended VP that the copulas are directly merged in the inflectional layer.

Other than that, several syntactic phenomena that hint at the privileged status of the subject in copular clauses in Malay are analysed, such as the selection of 3rd person subjects by *ialah*, the topical subject in specificational copular clauses, and the possibility of clefting only the subject of predicational copular clauses. One discovery that this chapter presents is the selectional property of copula *ialah* in relation to the subject, such that it may only combine with 3rd person subjects, due to the copula having grammaticalised from 3rd person pronoun *ia*. Further, the choice of copula in specificational copular clauses invariably being *ialah* is indicative of a specific feature on the copula that dictates the use of *ialah*, as opposed to *adalah*. This feature is the [*u*Top] feature by Mikkelsen (2005b) that requires the topical predicate to occupy SpecTP, yielding a specificational copular clause. The lack of this feature on *adalah* in inverse copular clauses confirms this hypothesis. With regard to clefting, only the subject of predicational copular clauses may undergo clefting, as it is the only constituent identifiable as the trigger. The other constituents in predicational and other kinds of copular clauses cannot be identified as triggers as both DPs in equative copular clauses have equal status and, in specificational copular clauses, the underlying subject does not correspond to the grammatical subject.

3.1 Misconceptions about Copular Clauses in Malay

There may be studies on different aspects of the Malay copulas scattered in the field by different scholars, but an ongoing tradition of research that builds upon each study to consolidate previous claims and findings seems to be missing. Therefore, this section attempts to dispel the misconceptions by invoking empirical data which might hopefully support existing research on copular clauses in Malay.

3.1.1 Categorical Selection

Mustaffa (2018) argues that the choice of copula is governed by the relationship between the constituents flanking the copula being predicational, specificational, or equative. Constituents that function as predicates, which ascribe a property to the subject (e.g. DPs, APs, or PPs) or implies set inclusion (e.g. indefinite DPs), license *adalah*, whereas those that specify a value for a variable introduced by the subject (see Higgins, 1979) or that can be equated with the subject, license *ialah*.

(158) a. *Hadiah itu adalah* {*rezeki* / *halal* / *untuk kamu.*} (Predicational)
gift DIST COP blessing halal for 2.SG
'That gift is a blessing/*halal*/for you.'

b. *Pemenang pertandingan itu ialah Ali.* (Specificational)
winner contest DIST COP A.
'The winner of that contest is Ali.'

c. *Hannah Montana ialah Miley Cyrus.* (Equative)
H. COP M.
'Hannah Montana is Miley Cyrus.'

Such an analysis is able to capture the data that only DPs are compatible with *ialah*, whereas the full array of nonverbal predicates may occur with *adalah*. This is because only DPs normally participate in a specificational or equative relation with the subject, seeing as DPs can function as referential arguments (although constituents from other syntactic categories may also occur in specificational and equative copular clauses). Also, there appears to be a human vs. non-human distinction in the choice of copula in equative copular clauses in Malay, as described further in Section 3.3.2. On the other hand, predicative DPs, APs, and PPs are not referential, as referentiality is a property of arguments, and a constituent cannot function as both things

simultaneously. Therefore, the three relations reflect the referentiality of the constituents flanking the copula, as summarised below:[37]

Type	Pre-Copular XP	Copula	Post-Copular XP
Predicational	Referential	*adalah*	Non-referential (DP, AP, PP)
Specificational	Non-referential	*ialah*	Referential (DP)
Equative	Referential	*ialah*/*adalah*	Referential ($DP_{H/NH}$)

Table 13: Referentiality According to Type of Copular Clause

This way of looking at it accounts for cases in which the interpretation of the copular clause is ambiguous among specification, predication, and equation, which leads to the possibility of using either *ialah* or *adalah* – a result that is not predicted by rules based on the syntactic category of the post-copular constituent. To illustrate, when decontextualised, the example below is three-way ambiguous: on a predicational reading, the subject is ascribed the property of being a prime minister by the post-copular DP predicate; on a specificational reading, the subject introduces a variable to which the post-copular DP referent specifies a value; on an equative reading, the subject and the referent denoted by the post-copular DP are one and the same entity.

(159) *Pe-laku=nya* {*ialah* / *adalah*} *Perdana Menteri.*
AG-do=3 COP COP prime minister
'The perpetrator is the Prime Minister.'

This view is in opposition to the common misconception among Malay grammarians that the choice of copula is conditioned by the syntactic category of the nonverbal predicate. This is made even more misleading to the general public as the official grammar by Dewan Bahasa dan Pustaka prescribes the rule that *ialah* is only compatible with nominal predicates, whereas *adalah* is compatible with adjectival and prepositional predicates, to the exclusion of nominal predicates (Karim et al., 2014).

[37] The post-copular constituent determines the choice of copula in predicational and equative copular clauses, but it seems that it is the pre-copular constituent that does so in specificational copular clauses, as described further in Section 3.3.2.

Copula	NP	AP	PP
Ialah	✓	✗	✗
Adalah	✗	✓	✓

Table 14: The Malay Copulas and Their Predicates According to Karim et al. (2014)

The rule by Karim et al. (2014) does not reflect natural use of the Malay copulas and is not based on either synchronic or diachronic evidence, which is in and of itself fine, as the reference pertains to the standard variety. However, the fact that different scholars have different opinions about which copula goes with which category of nonverbal predicate in the spoken variety is a matter for concern, as what is envisioned to be a clear grammatical rule has not received a general consensus among scholars. The table below shows the judgements of different scholars with regard to the nonverbal predicates compatible with copula *adalah*.

Scholar	Nominal	Adjectival	Prepositional
Omar (2014)	✓	✓	✓
Omar and Rama (1968)	✓	✓	✓
Arbak Othman (1987)	✓	✓	✓
Sneddon (1996)	✓	✓	✓
Karim et al. (2014)	✗	✓	✓
Awang Sariyan (1984)	✗	✓	✓
Harahap (1991)	✓	✗	✓
Kader (1986)	✓	✗	✓
Arka (2013)	✓	✗	✗

Table 15: The Nonverbal Predicates Compatible with 'Adalah' According to Scholar

It has been statistically shown that native speakers of Malay do not adhere to the prescribed rule (Khairul Taufiq & Nor Hashimah, 2017; Mustaffa, 2018). The following examples of copular clauses in which *adalah* combines with nominal predicates are grammatical. Conversely, the grammaticality of the use of the copula *ialah* is marginal despite the nominal predicate.

(160) a. *Saya* {*adalah*/? *ialah*} *ahli Majlis Ter-tinggi UMNO Malaysia.*
1.SG COP COP member council SUP-high U. M.
'I am a member of the highest council of UMNO Malaysia.'
(Nasir & Abu Bakar, 2021)

b. *Kamu* {*adalah*/? *ialah*} *se-orang pemuda.*
2.SG COP COP one-CLF youngster
'You are a youngster.'
(Mohamada, 2020)

Other than that, there have also been scholars who have claimed that *adalah* is not possible with adjectival predicates. This is the position taken by Kader (1986) and Harahap (1991). More recently, Arka (2013) claims that *adalah* cannot be used with adjectival predicates, based on the ungrammaticality of the copula with *sakit* (ill).

(161) **John adalah sakit.*
John COP ill
(John is ill.)
(Arka, 2013, p. 31)

This misconception seems to be a product of a hasty conclusion as, although it is true that *adalah* is impossible with certain adjectives, specifically temporary states such as *sakit* (ill), *lapar* (hungry), *pening* (dizzy), etc., it may occur with other types of adjectives without a hint of ungrammaticality. Example (162) illustrates that the copula is not just possible with an AP, but is obligatory, provided the right context and syntactic environment. There are rules that govern overt encoding of the copula, as is explicated in Chapter 4.

Context: *Cubalah bagaimanapun, tiada alasan untuk mengizinkan zina…*
(Try as you may, there is no justification for adultery…)

(162) … *kerana haram* *(*adalah*) *haram.*
because haram COP haram
'… because *haram* is *haram*.'

Corpus studies by Khairul Taufiq and Nor Hashimah (2017) and Mustaffa (2018) have made it clear that, although it is true that *ialah* is exclusively used with nominals, *adalah* is compatible with all three categories of nonverbal predicates, and more. In fact, the use of *adalah* with nominal predicates outnumbers its use with adjectival and prepositional predicates combined.

	Post-Copular Constituent						
	DP	AP	PP	VP	CP/TP	Others	Total
N	9458	3873	2687	1909	71	28	18026
%	52.47	21.49	14.91	10.59	0.39	0.15	100

Table 16: Predicates Following 'Adalah' in DBP Corpus of Utusan Newspapers (Mustaffa, 2018)

3.1.2 Co-occurrence of Copulas and Verbs

Further on the type of predicate possible with each copula, it is believed that the copulas must only combine with nonverbal constituents. This is again not an accurate reflection of the authentic use of the copulas. Karim et al. (2014) prescribe that *adalah* cannot be used preceding verbs, but post-copular constituents that are verbal are very common, regardless of the standard or vernacular status of the language. The following example is one in which the copula combines with a separate clause, presumably underlyingly a CP, but its appearance as a verbal constituent on the surface is enough to refute the prescriptive claim that the copula does not combine with a verb.

(163) *Sasaran saya ialah meŋ-menang-i kejohanan ini.*
goal 1.SG COP ACT-win-APPL championship PROX
'My goal is to win this championship.'
(AFP, 2019)

If verbs and copulas truly cannot co-occur, even more surprising would be the high productivity of *adalah* with certain linking verbs, as in the double-copula construction.[38] Given that high productivity is a sign that native speakers find such constructions grammatical, the rule prescribed does not reflect the grammaticality judgements of native speakers

(164) a. *KBAT adalah meŋ-rupa-kan kemahiran yang perlu di-kuasa-i oleh murid.*
K. COP ACT-form-APPL skill REL need PASS-power-APPL by pupil
'*KBAT* is a skill that needs to be mastered by pupils.'
(BERNAMA, 2015)

b. *Segala sesuatu adalah ber-ada di bawah ke-tentu-an Allah.*
every something COP INTR-EXIST LOC under NMZ-certain-NMZ A.
'Everything is under the certainty of Allah.'
(M. Sulaiman, 2016)

[38] Unfortunately, *ialah* cannot be used in the same environment.

Some of these linking verbs are semantically contentful on their own, whilst others semantically vacuous. To illustrate, both the copula and the verb may be omitted in the example below without any change in the meaning of the sentence, which suggests that they are both semantically vacuous. In fact, the sentence reads more easily by omitting *menjadi* and retaining the copula.

(165) *Mampu ber-aksi dengan baik*… (*adalah* (*meŋ-jadi*)) *impian saya.*
able INTR-perform with well… COP ACT-become dream 1.SG
'To be able to perform well is my dream.'
(Muhammad Malik, 2019)

One might construe *adalah*, in this case, as the aspectual auxiliary *ada-lah*. If that were true, the omission of the optional *lah* morpheme would disambiguate the sentence and clarify the status of the word *adalah* as the focused aspectual auxiliary, as opposed to the copula. Focused or not, the aspectual auxiliary promotes verum focus and forces an inchoative interpretation of the verb, which causes an obvious change in the interpretation of the whole sentence. As a result, *menjadi* is interpreted as a dynamic verb instead of some semantically vacuous linking verb, as illustrated in (166). The change in interpretation indicates that *adalah* is not the focused aspectual auxiliary.

(166) *Mampu ber-aksi dengan baik ada meŋ-jadi impian saya.*
able INTR-perform with well AUX ACT-become dream 1.SG
i. #'To be able to perform well is my dream.'
ii. 'To be able to perform well DID become my dream.'

Furthermore, *menjadi* is not interpreted as heading a separate subordinate clause in which it forms a clausal constituent that occurs in post-copular position. In fact, such an interpretation is semantically anomalous, as shown below:

(167) *Mampu ber-aksi dengan baik adalah meŋ-jadi impian saya.*
able INTR-perform with well COP ACT-become dream 1.SG
i. 'To be able to perform well is my dream.'
ii. #'To be able to perform well is to become my dream.'

Syntactically speaking, if the linking verb were to head a clause of its own, it should be possible to insert *untuk* (for), which marks certain non-finite clauses, as illustrated below with both verbal and nonverbal clauses:

(168) a. *Hasrat=nya adalah* (*untuk*) *meŋ-jadi isteri impian saya.*
intention=3 COP for ACT-become wife dream 1.SG
'Her intention is to become the wife of my dreams.'

b. *Dia ber-hasrat* (*untuk*) *meŋ-jadi isteri impian saya.*
3.SG INTR-intention for ACT-become wife dream 1.SG
'She intends to become the wife of my dreams.'

Doing so with example (165) is illicit, entailing that *menjadi* does not head a separate, subordinate clause:

(169) *Mampu ber-aksi dengan baik adalah* (**untuk*) *meŋ-jadi impian saya.*
able INTR-perform with well COP for ACT-become dream 1.SG
'To be able to perform well is my dream.'

In addition to that, the constituents flanking the copula are free to invert, given the right circumstances (see Section 3.3.3 for an analysis of copular inversion in Malay). For instance, it is possible to invert the pre- and post-copular constituents in (168a) since the constituent headed by *menjadi* participates in a specificational relationship with the other constituent. On the other hand, the same inversion is not possible in example (165), as illustrated in (170b), due to the fact that *menjadi* does not form part of the predicative clause of the copular clause, i.e. "impian saya" is the true predicate to the exclusion of *menjadi*.

(170) a. *Meŋ-jadi isteri impian saya adalah hasrat=nya.*
ACT-become wife dream 1.SG COP intention=3
'To become the wife of my dreams is her intention.'

b. **Meŋ-jadi impian saya adalah mampu ber-aksi dengan baik.*
ACT-become dream 1.SG COP able INTR-perform with well
(My dream is to be able to perform well.)

In light of this, omitting *menjadi* remedies the ungrammaticality:

(171) *Impian saya adalah mampu ber-aksi dengan baik.*
dream 1.SG COP able INTR-perform with well
'My dream is to be able to perform well.'

In terms of morphology, there are different requirements with regard to voice marking on the verb. Verbs heading subordinate clauses are free to omit the voice marker. To illustrate, the verb introducing the subordinate clauses in the following examples may be bare:

(172) a. *Hasrat=nya adalah (meŋ-)jadi isteri impian saya.*
intention=3 COP ACT- become wife dream 1.SG
'Her intention is to become the wife of my dreams.'

b. *Dia ber-hasrat (meŋ-)jadi isteri impian saya.*
3.SG INTR-intention ACT- become wife dream 1.SG
'She intends to become the wife of my dreams.'

Contrastively, the non-subordinate linking verb that co-occurs with the copula in the root clause must be affixed by a voice marker: [39]

(173) *Mampu ber-aksi dengan baik adalah *(meŋ-)jadi impian saya.*
able INTR-perform with well COP ACT- become dream 1.SG
'To be able to perform well is my dream.'

Other than linking verbs, *adalah* is also used with verbs that have deverbalised into adjectives by circumfixion by the active voice marker and the applicative suffix, *meŋ–kan* (ACT–APPL). These deverbal adjectives, which are verbal in form but adjectival in function, are comparable to deverbal adjectives that have been derived via affixation of the participial *-en* or *-ing* suffixes in English to form adjectives such as *boring/bored, disappointing/disappointed, frightening/frightened, saddening/saddened*, etc. Other than their use with copula *adalah*, the way to discern the adjectival status of such lexemes is its use with the active form of *jadi* (become), which requires a nonverbal predicate.

(174) *Perkara itu* {*adalah* / *meŋ-jadi* }{ *meŋ-bosan-kan* / *meŋ-kecewa-kan* / etc.}
matter DIST COP ACT-become ACT-bore-APPL ACT-disappoint-APPL etc.
'That matter (is/becomes) (boring/disappointing/etc.)'

More generally, stative verbs that function as predicates that are descriptive or those that ascribe a property, akin to adjectival or prepositional predicates, can combine with *adalah*. The post-copular constituents in the examples below are formally and morphosyntactically verbs, but their function is similar to that of a prepositional or adjectival predicate. For example, both verbs are affixed by a verbal prefix and the verb in (175a) takes a DP complement, whilst the one in (175b) has an agent.

[39] Also, considering that the voice marker heads VoiceP, which is structurally higher than *v*P, as established in Section 1.5.7, the copulas could not have been merged within the verbal layer.

(175) a. *Doa-doa kami adalah ber-sama mereka.* (Prepositional)
prayer-RED 1.PL.EXCL COP INTR-same 3.PL
'Our prayers are with them.'
(BERNAMA, 2019)

b. *Tindak-an sedemikian adalah di-larang oleh agama.* (Adjectival)
act-NMZ such COP PASS-prohibit by religion
'Such an action is prohibited by religion.'
(Abd Jamil, 2021)

3.1.3 The Internal Structure of the Copular Clause

Copular clauses in Malay are not merely juxtapositions of subject and predicate, as assumed by Karim et al. (2014). In other words, there is more syntactic structure between the two constituents than meets the eye. The fact that a copula, albeit optional, could surface constitutes enough evidence for this claim. According to Mustaffa (2018), it is possible to find constituents associated with the inflectional and verbal layers of a clause in copular clauses as well. For example, temporal adverbials are able to occur in copular clauses without an overt copula, as in (176a), which signals that copular clauses comprise at least a VP, considering that such constituents require a verbal layer to which to adjoin. In addition to that, the possibility of there occurring auxiliaries in copular clauses indicates that a copular clause has an inflectional layer, as in (176b).

(176) a. *Dia sihat semalam.*
3.SG healthy yesterday
'She was healthy yesterday.'

b. *Kulit=ku akan cerah.*
skin=1 PROS fair
'My skin will be fair.'

What this means is that there is internal structure, albeit silent most of the time, that may host inflectional and verbal constituents. Therefore, a root copular clause with a nonverbal predicate should be like any other clause with a verbal predicate such that it is comprised of a VP and an IP/TP.

Moreover, following research by Stowell (1983) and Bowers (1993), the derivation of a copular clause in Malay as well should begin from a small clause in which the subject and the predicate are base generated. As commonly assumed, the constituents of a small clause (labelled as PredP) occur in a configuration in which the subject occupies the specifier position, whilst the predicate the complement position (Bowers, 1993; Den Dikken, 2006; Mikkelsen, 2005a; Moro, 1997). Example (177a) shows *jadi* (become) taking a small clause as its complement. The inversion of the subject and predicate in (177b) within the small clause causes ungrammaticality, which demonstrates the ordering of the pair.

(177) a. *Saya meŋ-jadi-kan* [$_{PredP}$ *dia tujuan saya untuk hidup.*]
1.SG ACT-become-APPL 3.SG purpose 1.SG for live
'I made her my purpose to live.'

b. **Saya meŋ-jadi-kan* [$_{PredP}$ *tujuan saya untuk hidup dia.*]
1.SG ACT-become-APPL purpose 1.SG for live 3.SG
(I made my purpose to live her.)

Typically, copular clauses are derived via raising of the subject from the small clause to the specifier of TP. To illustrate, without the applied subject in the matrix clause, the subject in the small clause can raise to SpecTP.

(178) *Dia*$_1$ *meŋ-jadi* [$_{PredP}$ t$_1$ *tujuan saya untuk hidup.*]
3.SG ACT-become purpose1.SG for live
'She became my purpose to live.'

However, it is difficult to identify small clauses in Malay because of the absence of morphological tense in the language and the impossibility of the copulas in certain non-finite clauses, which are unmarked morphologically. Verbs such as *anggap* (consider) and *kelihatan* (seem) are therefore difficult to analyse with respect to small clauses as they may also introduce a complement clause with an optionally null complementiser. So, the best candidates to analyse would be verbs that do not take clausal complements such as *namakan* (name), *gelar* (name), *panggil* (call), etc.

Nevertheless, it is possible to see the $Pred^0$ head surface, which validates the postulation of PredP. For instance, the matrix verb in example (179a) selects a small clause in which preposition *sebagai* (as) appears to be a copula-like element of the likes of those in Den Dikken (2006) that marks a predicational relation between the two DPs. When the preposition is present, the use of the complementiser renders the sentence ungrammatical, which indicates that there is no complement clause.

(179) a. *Saya anggap (*bahawa) dia sebagai abang saya.*
1.SG consider COMP 3.SG as brother 1.SG
'I consider him my brother.'

b. *Saya anggap bahawa dia (* sebagai) abang saya.*
1.SG consider COMP 3.SG as brother 1.SG
'I consider that he is my brother.'

The subject of the small clause may encliticise onto the matrix verb, which indicates that the verb and the subject are local. In other words, no intervening projection is present to prevent cliticisation of the subject onto the verb, constituting evidence that the verb selects a small clause, rather than a complement clause.

(180) a. *Saya anggap=nya sebagai abang saya.*
1.SG consider=3 as brother 1.SG
'I consider him my brother.'

In some cases, the preposition is optional, but in others it is obligatory. In the example below, the preposition is obligatory as omitting it would alter the interpretation of the constituent "Naib Canselor" to mean the agent of the passive sentence, i.e. "Zul was elected *by* the Vice Chancellor".

(181) a. *Zul di-lantik *(sebagai) Naib Canselor.*
Z. PASS-elect as Vice Chancellor
'Zul was elected Vice Chancellor.'

Therefore, a copular clause in Malay should have the following structure:

$$[_{TP}\ [_{VP}\ [_{PredP}\ [\text{SUBJECT}]\ [_{Pred'}\ Pred^0\ [\text{PREDICATE}]]]]]$$

A further description of the structure of copular clauses in Malay is presented in the next Section.

3.2 The Category and Position of the Copulas

Based on the findings in Section 3.1 regarding the co-occurrence of copulas with verbs and the presence of the small clause in the structure of a copular clause, as well as further evidence presented in this section, we can chart the specific position of the copula in a copular clause.

Copulas are commonly assumed to head VPs or TPs. If they are construed to be auxiliaries, it should be the case that they head a projection in the inflectional layer of the clause, likely TP. The strongest piece of evidence for the claim that copulas *ialah* and *adalah* head TP is their distribution relative to negators. When negation is concerned, *adalah* is unlike verbs such that it precedes the negator. Given that *tidak* precedes verbs and all auxiliaries except epistemic modals, the position of *adalah* preceding the negator confirms that it heads TP. As illustrated below, the copula precedes the negator, whereas the existential verb follows it:

(182) a. *Wanita itu adalah tidak waras.*
woman DIST COP NEG sane
'The woman is not sane.'

b. *Wanita itu tidak ada di sini.*
woman DIST NEG EXIST LOC here
'The woman is not here.'

Similarly, *ialah* precedes the negator *bukan*. Although the co-occurrence of *ialah* and *bukan* is judged to be marginal by speakers, examples of such are attested in newspapers, as illustrated below:

(183) a. *Yang penting ialah bukan sahaja bagaimana kita hendak meŋ-tuntut hak-hak…*
COMP important COP CNTR only how 1.PL want ACT-demand right-RED
'What is important is not only how we are to demand the rights…'
(Linch, 2022)

b. *Isu=nya ialah bukan parti mana yang punya banyak kerusi dalam blok peŋ-bangkang…*
issue=3 COP CNTR party which COMP possess many seat in bloc AG-oppose
'The issue is not which party possesses many seats in the opposition bloc…'
(Isa, 2022)

The status of the copulas as heads of TP is confirmed by their impossibility to co-occur with epistemic modals, which I have shown to be heads of TP as the highest projection in the inflectional layer, by their position preceding both negators *tidak* and *bukan*, in Sections 1.5.4 and 1.5.6 respectively. The impossibility for them to co-occur indicates that the copulas and epistemic modals stand in complementary distribution.[40] Therefore, the copulas are auxiliaries, as are the epistemic modals. As illustrated below, both *ialah* and *adalah* may not be used when epistemic *mesti* is present:

(184) a. *Wanita itu mesti* (**adalah*) *tidak waras.*
woman DIST must COP NEG sane
'The woman must not be sane.'

b. *Wanita itu* (**adalah*) *mesti tidak waras.*
woman DIST COP must NEG sane
'The woman must not be sane.'

c. *Wanita itu mesti* (**ialah*) *kawan saya.*
woman DIST must COP friend 1.SG
'The woman must be my friend.'

d. *Wanita itu* (**ialah*) *mesti kawan saya.*
woman DIST COP must friend 1.SG
'The woman must be my friend.'

To be certain that it is the TP that the copulas head, and not some clause-peripheral position, consider examples (185). The examples show that the subject of the subordinate copular clause has undergone raising into the matrix passive clause.

(185) a. *Senjata yang di-guna-kan di-percaya-i adalah parang milik suami=nya.*
weapon REL PASS-use-APPL PASS-believe-APPL COP machete own husband=3
'The weapon that was used is believed to be the husband's machete.'
(Abdul Rahman, 2018)

b. *Wanita itu juga di-faham-kan ialah rakan.kongsi kepada aktor terbabit.*
woman DIST also PASS-understand-APPL COP partner to actor involved
'The woman is also understood to be a partner to the actor involved.'
(Mohamad, 2016)

[40] The order (copula/epistemic modal) > *bukan* > *tidak* entails that the contrastive negator *bukan* is either selected by T or adjoins to the phrase that is selected by T, i.e. NegP or *v*/VP. Perhaps *bukan* is an adverbial phrasal constituent, like *never* in English.

According to the standard Minimalist analysis of raising, the complement clause of a raising predicate is defective such that it does not inherit the tense and φ-features that are inherent to the CP layer, on account of the absence of CP (Chomsky, 2008). Unlike control clauses, raising out of a defective clause is possible, owing to there being no CP layer to prevent movement. In relation to the position of the copulas, a clause-peripheral projection is ruled out as the position of the copulas as there is no CP layer in raising examples (185). Therefore, the copulas are confirmed to be auxiliaries that head TP.

Finally, recall from Section 3.1.2 that the copula *adalah* may co-occur with certain linking verbs such as *meŋ-jadi* (ACT-become), *meŋ-rupa-kan* (ACT-form-APPL), *ber-asal* (INTR-origin), *ber-ada* (INTR-EXIST), *ber-maksud* (INTR-meaning) etc. Combining the findings regarding the position of the copulas preceding negators and the co-occurrence of the copulas with linking verbs, we find conclusive evidence that the copulas are auxiliaries that head TP. In the following example, the copula co-occurs with a linking verb in a negative copular clause. There is no chance for the copulas to be merged anywhere lower than the inflectional layer.

(186) *Ber-main di tempat ber-kecuali adalah tidak meŋ-jadi halangan…*
INTR-play LOC place INTR-except COP NEG ACT-become obstacle…
'Playing in a neutral place is not an obstacle…'
(Agensi, 2021)

The following diagram illustrates the structure of a copular clause in Malay and clarifies the syntactic position of the copula. It is clear that there is a verbal layer in the structure, as is made apparent by the possibility of the copula to co-occur with certain linking verbs. The fact that the linking verbs themselves can be marked by voice morphology entails that the full extended VP is present, i.e. VoiceP-*v*P-VP. As for the structure below the VP, the presence of a small clause in which the subject and predicate are initially merged has been argued for in Section 3.1.3.

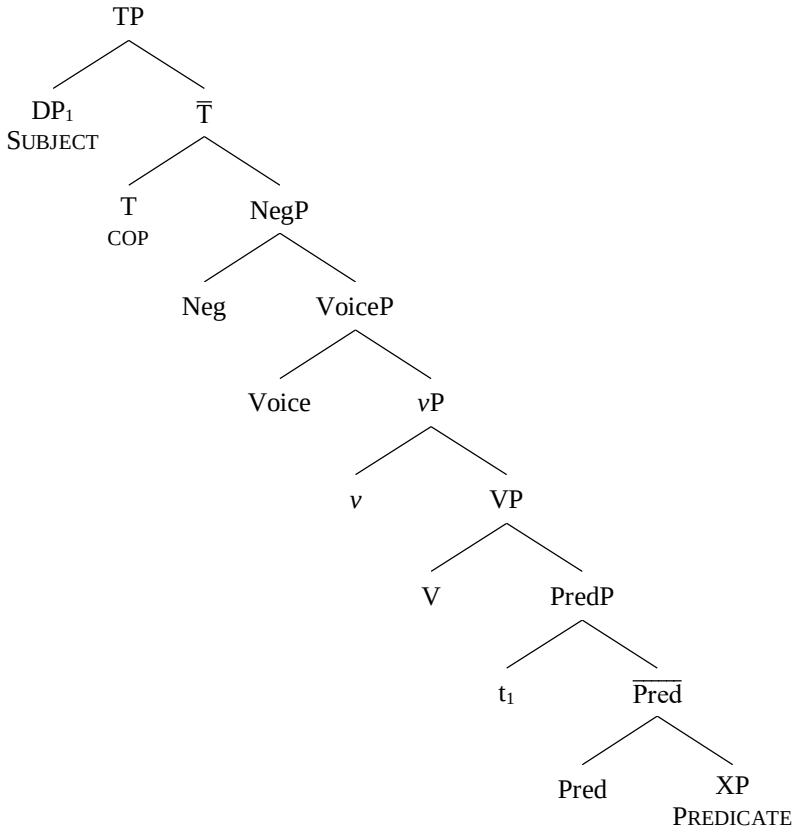


Figure 13: The Full Syntactic Structure of the Copular Clause

3.3 Subjects in Copular Clauses in Malay

Subjects seem to have privileged status in copular clauses in Malay. The following subsections are an assembly of several syntactic phenomena that demonstrate the privileged status of subjects in copular clauses in Malay. Although different types of copular clauses exhibit different properties, one constant among them is the influence that the subject has on copular clauses, such as the selection of 3rd person subjects exclusively in copular clauses with *ialah*, the obligatoriness of the grammatical subject to correspond to a topic in specificational copular clauses, and the possibility of clefting and relativising only the subject of predicational copular clauses.

3.3.1 Selection of 3rd Person Subjects by *Ialah*

The copula *ialah* exhibits a requirement for its subject to be 3rd person. Although certain authors claim that combining *ialah* with 1st or 2nd person subjects is possible, the reality is that such sentences are degraded. For example, Sneddon (1996) states "*ialah* only occurs after third person subjects", whilst Harahap (1991) provides examples of *ialah* combining with *saya* (1.SG), vouching for their grammaticality. Consider the equative copular clauses in the examples below, in which the use of *ialah* with a 1st or 2nd person subject is significantly degraded.[41] Also notice that the acceptability of *ialah* with a 3rd person pronominal subject is slightly less than with a full DP, which allows interchangeability of the copula.

Context: *Biarlah saya buat apa yang saya nak.* (Let me do as I please.)

(187) a. *Saya* {??*ialah* /*adalah*} *saya.* (1st Person)
1.SG COP COP 1.SG
'I am me.'

b. *Kamu* {??*ialah* /*adalah*} *kamu.* (2nd Person)
2.SG COP COP 2.SG
'You are you.'

c. *Dia* {?*ialah* /*adalah*} *dia.* (3rd Person)
3.SG COP COP 3.SG
'He is him.'

d. ... *dan Zul* {*ialah* /*adalah*} *Zul.* (Full DP)
and Z. COP COP Z.
'… and Zul is Zul.'

Such a restriction is not just a hunch but supported by statistical evidence. Consider Table 17 below, which draws data from the corpus of 329 modern novels by Dewan Bahasa dan Pustaka (2013). Almost all of the 8,125 sentences with *ialah* have full DP subjects, which are 3rd person by default. As for those with pronominal subjects, a total of only 8 examples with non-3rd person subjects are found, which attests to the constraint that copula *ialah* occurs with 3rd person subjects only.

[41] Presumably, the grammaticality judgements are marginal due to the prevalence of the prescribed rule that *ialah* must be used when the complement of the copula is nominal. Otherwise, the selectional property of the copula should render the sentence ungrammatical.

	Saya (1.SG)	*Kita* (1.PL.INCL)	*Kami* (1.PL.EXCL)	*Kamu* (2.SG/PL)	*Dia* (3.SG)	*Mereka*[42] (3.PL)
N	58,952	45,105	29,585	8,438	162,802	77,584
N [+*ialah*]	6	2	0	0	44	43
Frequency	1.02	0.44	0.00	0.00	2.70	5.54

Table 17: The Subjects of Copular Clauses with 'Ialah' in DBP Corpus of Modern Novels[43]

However, there appears to be a further human vs. non-human distinction:

Context: <u>*Cubalah sebanyak mana pun, sifatnya/resipinya takkan boleh diubah.*</u>
<u>(Try as you may, his nature/the recipe can't be changed.)</u>

(188) a. *Zul* {*ialah* /*adalah*} *Zul.* (Human)
Z. COP COP Z.
'Zul is Zul.'

b. *Kari* {??*ialah* /*adalah*} *kari.* (Non-Human)
curry COP COP curry
'Curry is curry.'

Therefore, the judgements indicate a tendency for *ialah* to occur with human DPs and *adalah* to occur elsewhere. Thus, it can be analysed that the use of *ialah* is more constrained than the use of *adalah*, which seems to be the elsewhere copula. This is even more apparent when non-DP constituents are equated. To illustrate, the use of *ialah* in the examples below are ungrammatical:

Context: <u>*Cubalah bagi alasan mana pun, hukumnya tetap sama.*</u>
<u>(Justify as you may, the rule remains the same.)</u>

(189) a. *Meŋ-curi* {**ialah* /*adalah*} *meŋ-curi.* (Verbal)
ACT-steal COP COP ACT-steal
'Stealing is stealing.'

b. *Haram* {**ialah* /*adalah*} *haram.* (Adjectival)
haram COP COP haram
'*Haram* is *haram*.'

[42] *Mereka*, recently borrowed from Old Javanese presumably after the Classical Malay period according to Adelaar (1992), was originally a noun meaning 'people'. Presumably, its high occurrence with the copula reflects its former non-pronominal status.

[43] % = (N [+ *ialah*] ÷ (N ÷ 10,000))

Understandably, because the copula *adalah* had no pronominal origins, there should not be such a constraint on the copula. This is apparent in the table below, which reveals a more distributed occurrence of subjects with different person features. A total of 10,814 examples of sentences with *adalah* is registered and the total of examples with pronominal subjects amounts to 568 (5.25%). However, there does appear to be a statistical tendency for the copula to combine with 3rd person pronominal subjects as well, though not too obvious.

	Saya (1.SG)	*Kita* (1.PL.INCL)	*Kami* (1.PL.EXCL)	*Kamu* (2.SG/PL)	*Dia* (3.SG)	*Mereka* (3.PL)
N	58,952	45,105	29,585	8,438	162,802	77,584
N [+*Adalah*]	38	53	21	24	277	155
Frequency	6.45	11.75	7.10	28.44	17.01	19.98

Table 18: The Subjects of Copular Clauses with 'Adalah' in DBP Corpus of Modern Novels

3.3.2 Specificational Copular Clauses

As described in Section 3.1.1, choosing which copula to use on a syntactic categorical basis is untenable as the use of *adalah* with nominal predicates is widely attested. The referentiality of the complement of the copula plays a part in the choice of copula. For example, when the post-copular constituent corresponds to a non-referential DP, the copular clause is interpreted to be predicational, warranting the use of *adalah*. To illustrate, the copula in the following example combines with a nominal predicate, despite it being a name. As such, a predicational reading is induced (i.e. the name denotes a property, rather than a person), hence the use of *adalah*.

(190) *Beg ini* {* *ialah* /*adalah*} Alexander McQueen.
bag PROX COP COP A.
'This bag is Alexander McQueen.'

Now consider specificational copular clauses, which are copular clauses in which the post-copular constituent is interpreted as though it supplies a value to a variable that is set up by the subject, as described by Akmajian (1970a). To illustrate, the post-copular DP in the following examples is the answer and value to the implied question and variable in "whom/what does he like?" posed by the subject.

(191) a. *Orang yang dia suka* {*ialah* /??*adalah*} *Zul.* (Human)
person REL 3.SG like COP COP Z.
'The person that he likes is Zul.'

b. *Benda yang dia suka* {*ialah* /??*adalah*} *kari.* (Non-Human)
thing REL 3.SG like COP COP curry
'The thing that he likes is curry.'

Regardless of the features of the post-copular DP, the copula used remains the same, *ialah*. It is only when the copular clause is specificational that the copula used is consistently *ialah*. This finding indicates that the features of the post-copular constituent have no bearing on the copula. Perhaps it is the features of the pre-copular constituent, i.e. the surface subject, that plays a role in determining the choice of copula in specificational copular clauses in Malay.

Partee (1998) and Mikkelsen (2005b) notice that specificational copular clauses have a unique information-structural alignment in that the pre-copular constituent always corresponds to a topic, whilst the post-copular constituent the focus. This Topic-Focus alignment apparently cannot be altered, as doing so would render the copular clause anomalous. The contrast between the examples below shows that, when the subject of a specificational copular clause corresponds to the *wh*-phrase of the question, the answer is infelicitous, which is due to the focus carried by the *wh*-phrase and its corresponding answer. It is apparent from this contrast that, whereas predicational copular clauses may have free information-structural order, specificational copular clauses have a fixed topic-focus order.

Context: Who is the designer?

(192) a. [$_{Top}$ *The designer*] *is* [$_{Foc}$ *Alexander McQueen.*] (Specificational)

b. [$_{Foc}$ *Alexander McQueen*] *is* [$_{Top}$ *the designer.*] (Predicational)

Context: Who is Alexander McQueen?

(193) a. #[$_{Foc}$ *The designer*] *is* [$_{Top}$ *Alexander McQueen.*] (Specificational)

b. [$_{Top}$ *Alexander McQueen*] *is* [$_{Foc}$ *the designer.*] (Predicational)

The same pattern is observed in Malay:

Context: *Siapakah perekanya?* (Who is the designer?)

(194) a. [$_{Top}$ *Pe-reka=nya*] *ialah* [$_{Foc}$ *Alexander McQueen*.] (Specificational)
AG-design=3 COP A.
'The designer is Alexander McQueen.'

b. [$_{Foc}$ *Alexander McQueen*] *adalah* [$_{Top}$ *pe-reka=nya.*] (Predicational)
A. COP AG-design=3
'Alexander McQueen is the designer.'

Context: *Siapakah Alexander McQueen?* (Who is Alexander McQueen?)

(195) a. #[$_{Foc}$ *Pe-reka=nya*] *ialah* [$_{Top}$ *Alexander McQueen*.] (Specificational)
AG-design=3 COP A.
'The designer is Alexander McQueen.'

b. [$_{Top}$ *Alexander McQueen*] *adalah* [$_{Foc}$ *pe-reka=nya.*] (Predicational)
A. COP AG-design=3
'Alexander McQueen is the designer.'

Based on this observation, Mikkelsen (2005b) analyses the derivation of a specificational copular clause to be such that a [*u*Top] feature on the copula triggers movement of the predicate within the small clause, which carries an interpretable topic feature, to the subject position due to "the desire to align the topic with the subject position" (Mikkelsen, 2005b; Partee, 1998), as illustrated below:

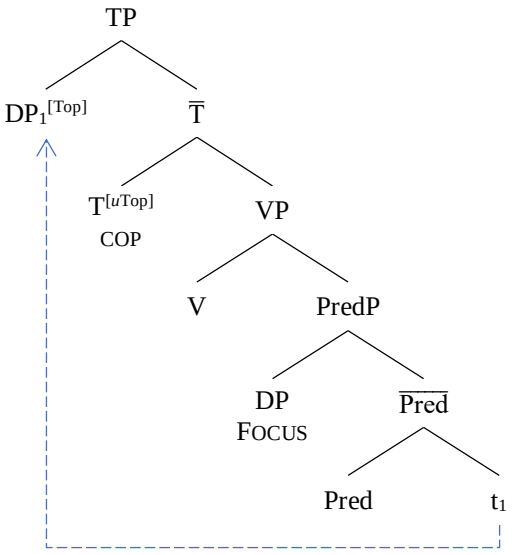


Figure 14: The Derivation of a Specificational Copular Clause

Associating this analysis to the data pertaining to specificational copular clauses in Malay, it seems plausible to hypothesise that the [*u*Top] feature is what makes *ialah* the invariable copula in specificational copular clauses in Malay. No such requirement is shared by the copula in predicational and equative copular clauses, as the pre-copular constituent could correspond to either the focus or the topic, as demonstrated in (194b) for predicational copular clauses and in the examples below for equative copular clauses. The choice of copula in equative copular clauses is not determined by information-structural factors.

Context: <u>*Siapakah Hannah Montana dalam cerita itu?*</u>
<u>(Who is Hannah Montana in the story?)</u>

(196) a. [$_{\text{Foc}}$ *Miley Cyrus*] *ialah* [$_{\text{Top}}$ *Hannah Montana.*] (Equative)
M. COP H.
'Miley Cyrus is Hannah Montana.'

b. [$_{\text{Top}}$ *Hannah Montana*] *ialah* [$_{\text{Foc}}$ *Miley Cyrus.*]
H. COP M.
'Hannah Montana is Miley Cyrus.'

3.3.3 Copular Inversion

Specificational copular clauses must not be equated with cases of copular inversion, which appear similar but are underlyingly different both syntactically and semantically. Whereas movement of the underlying predicate in the analysis of specificational copular clauses by scholars such as Mikkelsen (2005b) and Moro (2006) targets the subject position, i.e. SpecTP, movement of the predicate in inverse copular clauses targets SpecTopP.[44] To illustrate, an adjectival predicate has inverted with the underlying subject in the inverse copular clause below:

(197) *Turut hadir adalah Pengarah Pelancongan Negeri Kelantan…*
also present COP director tourism state K.
'Also present was the Director of Tourism of the state of Kelantan…'
(S. R. Idris, 2017)

[44] Although Moro (2006) calls specificational copular clauses inverse copular clauses, I shall use the term inverse copular clause to specifically refer to copular clauses in which a non-DP predicate has inverted with the subject. These inverse copular clauses do not have a specificational reading as the inverted predicate does not introduce a variable, as per Akmajian (1970a).

The inverted predicate in inverse copular clauses also needs to be topical or given, as supported by the pre-copular constituent of many such inverse copular clauses including adverbs such as *turut/juga* (also). The ungrammaticality of example (198) illustrates that a non-topical predicate may not participate in copular inversion.

Context: *Walaupun tidak kelihatan…* (Even though out of sight…)

(198) * … *hadir* *adalah* *Pengarah* *Pelancongan* *Negeri* *Kelantan…*
present COP director tourism state K.
'… present was the Director of Tourism of the state of Kelantan…'

However, other than the lack of a specificational reading due to the inverted predicate being an AP, the choice of copula is indicative of the predicational nature of the relationship between the two constituents flanking the copula. Presumably, that the predicate and underlying subject has inverted and that the post-copular constituent is a referential human DP should warrant the use of *ialah*. To the contrary, *adalah* is used, which confirms that cases of copular inversion are not identical to specificational copular clauses, given that the copula is invariable in specificational copular clauses, as illustrated in example (191), repeated below:

(199) a. *Orang yang dia suka* {*ialah* /??*adalah*} *Zul.* (Human)
person REL 3.SG like COP COP Z.
'The person that he likes is Zul.'

b. *Benda yang dia suka* {*ialah* /??*adalah*} *kari.* (Non-Human)
thing REL 3.SG like COP COP curry
'The thing that he likes is curry.'

The difference in the choice of copula reveals that the derivations of specificational and inverse copular clauses in Malay differ. The derivation of the former involves movement of the underlying predicate to SpecTP via checking of the [*u*Top] feature on T^0, as described in Section 3.3.2, whilst in the latter, movement of the predicate targets the left periphery as the T^0 head lacks the [*u*Top] feature. Therefore, an inverse copular clause starts out as a canonical predicational copular clause, judging from the choice of copula being *adalah*. Undoing the inversion in example (197) produces the canonical predicational copular clause below:

(200) *Pengarah Pelancongan Negeri Kelantan adalah turut hadir.*
director tourism state K. COP also present
'The Director of Tourism of the state of Kelantan was also present.'

Rather than the subject position, the predicate can be said to move to SpecTopP. Raising can be used to diagnose movement to SpecTP, as done by Heycock and Kroch (1997), who provide the following example in which the inverted constituent has undergone raising to a matrix clause:

(201) *'Especially dishonest seems to have been the Rockefeller family.'*
(Heycock & Kroch, 1997, p. 4)

Applying this diagnostic, although it is possible to raise the surface subject of a specificational copular clause (viz. the underlying predicate), the surface subject of an equative copular clause, and the surface subject of a predicational copular clause, raising of the predicate of an inverse copular clause with *adalah* is ungrammatical, as illustrated in example (202d). This finding constitutes evidence that inversion in different kinds of copular clauses in Malay (e.g. predicational, specificational, equative, etc.) involves different derivations. Specificational and equative copular clauses permit A-movement into the subject position of the matrix clause, whereas cases of copular inversion with *adalah* do not.

(202) a. *Pe-laku=nya*$_1$ *di-kata-kan* t$_1$ *ialah Ali.* (Specificational)
AG-do=3 PASS-say-APPL COP A.
'The perpetrator is said to be Ali.'

b. *Ali*$_1$ *di-kata-kan* t$_1$ *ialah pe-laku=nya*$_1$. (Equative)
A. PASS-say-APPL COP AG-do=3
'Ali is said to be the perpetrator.'

c. *Pe-laku=nya*$_1$ *di-kata-kan* t$_1$ *adalah turut hadir.* (Predicational)
AG-do=3 PASS-say-APPL COP also present
'The perpetrator was said to be also present.'

d. * *Turut hadir*$_1$ *di-kata-kan* t$_1$ *adalah pe-laku=nya.* (Inverse)
also present PASS-say-APPL COP AG-do=3
(Also present was said to be the perpetrator.)

Theoretically speaking, given the analysis by Mikkelsen (2005b), it is not possible for non-DP predicates to move to SpecTP as, even though they may be topical, they do not have the necessary features to satisfy the EPP and case features on T^0, which may only be satisfied by a DP. Therefore, apparent copular inversion in non-DP predicational copular clauses must involve the left periphery.

Furthermore, there are other data in Malay that may shed some light on the ongoing debate on copular inversion, and it surprisingly concerns inversion of VP and even AspP. *Adalah* is not usually possible with dynamic verbs, but the inversion of a VP spells out the copula in some cases, nevertheless, as illustrated in (203). Unlike English, Malay does not employ an auxiliary of the same form as the copula in passive and progressive clauses. Therefore, the spell-out of the copula is surprising.

(203) a. *Turut meŋ-meterai adalah PLB Engineering Bhd…*
also ACT-seal COP P.
'Also sealing (the deal) was PLB Engineering Bhd…'
(Abllah, 2019)

b. *Turut meŋ-doa-kan adalah penyanyi asal Singapura Awi Rafael.*
also ACT-pray-APPL COP singer origin S. A.
'Also praying (for him) was Singaporean singer Awi Rafael.'
(Abdul Karim, 2021)

c. *Turut akan di-wujud-kan adalah Taman Rakyat se-luas 85 ekar…*
also PROS PASS-exist-APPL COP park public as-wide 85 acre
'Also to be created is a public park 85-acres wide …'
(Abd Mutalib & Shahrul Annuar, 2019)

Apparently, the copula is obligatory in the inverted examples:

(204) a. *Turut meŋ-meterai *(adalah) PLB Engineering Bhd…*
also ACT-seal COP P.
'Also sealing (the deal) was PLB Engineering Bhd…'

b. *Turut meŋ-doa-kan *(adalah) penyanyi asal Singapura Awi Rafael.*
also ACT-pray-APPL COP singer origin S. A.
'Also praying (for him) was Singaporean singer Awi Rafael.'

c. *Turut akan di-wujud-kan *(adalah) Taman Rakyat se-luas 85 ekar…*
also PROS PASS-exist-APPL COP park public one-wide 85 acre
'Also to be created is a public park 85-acres wide…'

However, the presence of the copula renders the examples ungrammatical if the inversion is undone. This entails that inverse copular clauses with inverted verbal predicates start out as normal verbal clauses.

(205) a. *PLB Engineering Bhd* (**adalah*) *turut meŋ-meterai.*
P. COP also ACT-seal
'PLB Engineering Bhd was also sealing (the deal).'

b. *Penyanyi asal Singapura Awi Rafael* (**adalah*) *turut meŋ-doa-kan.*
singer origin S. A. COP also ACT-pray-APPL
'Singaporean singer Awi Rafael was also praying (for him).'

c. *Taman Rakyat se-luas 85 ekar* (**adalah*) *turut akan di-wujud-kan.*
park public as-wide 85 acre COP also PROS PASS-exist-APPL
'A public park 85-acres wide will also be created.'

It appears that topicalisation of the VP/AspP allows or strengthens the stative interpretation of the verb, as progressives and perfects are typically analysed as states (Comrie, 1976). The same goes for passives, which may get interpreted as states, e.g. adjectival passives (Levin & Malka, 1986). Conversely, when the verbal predicate appears in its usual position following the subject, it normally gets an eventive reading, which appears to be incompatible with the copula (more on the incompatibility of eventiveness with overt encoding of the copula in Chapter 4).

To further make the case for topicality as the main factor in copular inversion is the contrast between focalisation and topicalisation of the VP. When the inverted VP carries focal stress, use of the copula is not possible. Conversely, the copula is possible when it is the constituent "Ali" that carries focal stress.

(206) a. [$_{Foc}$ *Turut di-kuarantin*] (**adalah*) [$_{Top}$ *Ali.*]
also PASS-quarantine COP A.
'ALSO QUARANTINED, Ali is.'

b. [$_{Top}$ *Turut di-kuarantin*] *adalah* [$_{Foc}$ *Ali.*]
also PASS-quarantine COP A.
'Also quarantined is Ali.'

Since examples (203) are originally verbal clauses, they may be analysed to have undergone the standard derivation of verbal clauses whereby the subject (of a transitive clause) is merged in Spec*v*P and moves to SpecTP. After that, the extended VP is then moved to the left periphery. Finally, the copula is either merged directly in Top^0 or moved to it from T^0.[45]

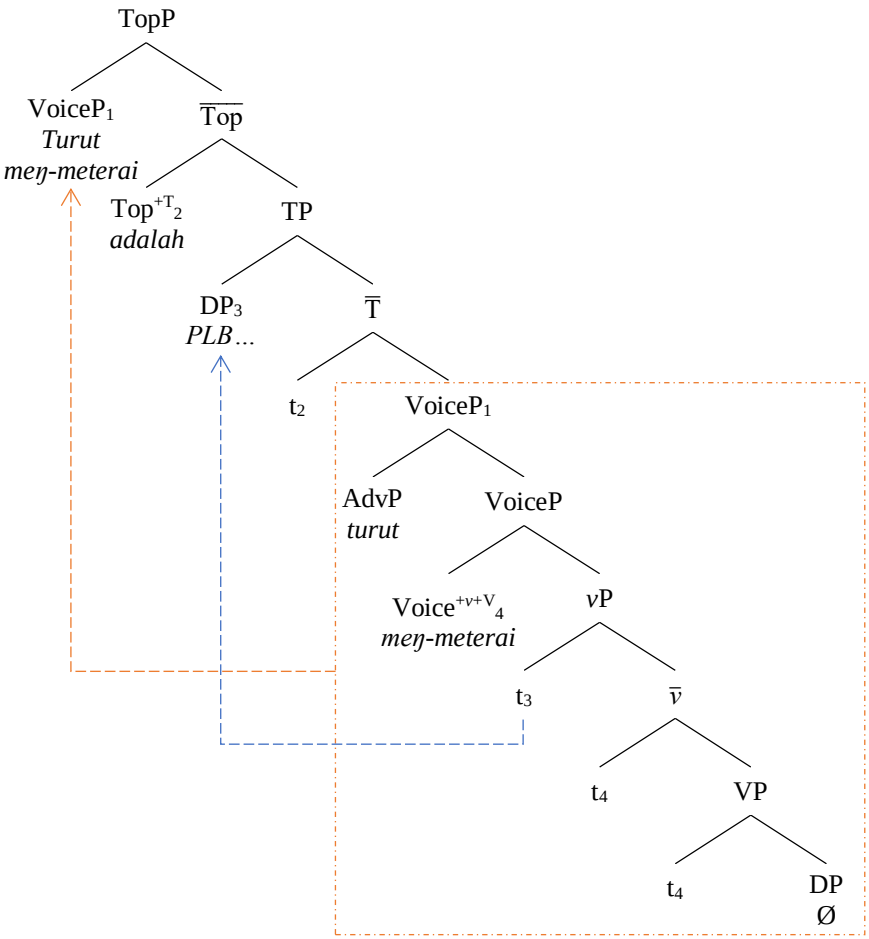


Figure 15: The Derivation of an Inverted Verbal Predicate

On the other hand, inverse copular clauses with nonverbal predicates can be analysed to be derived from canonical predicational copular clauses as illustrated below:

[45] Recall from Section 3.1.2 that the copula *adalah* may co-occur with linking verbs in mono-clausal copular clauses. So, the copula may have been merged in TP.

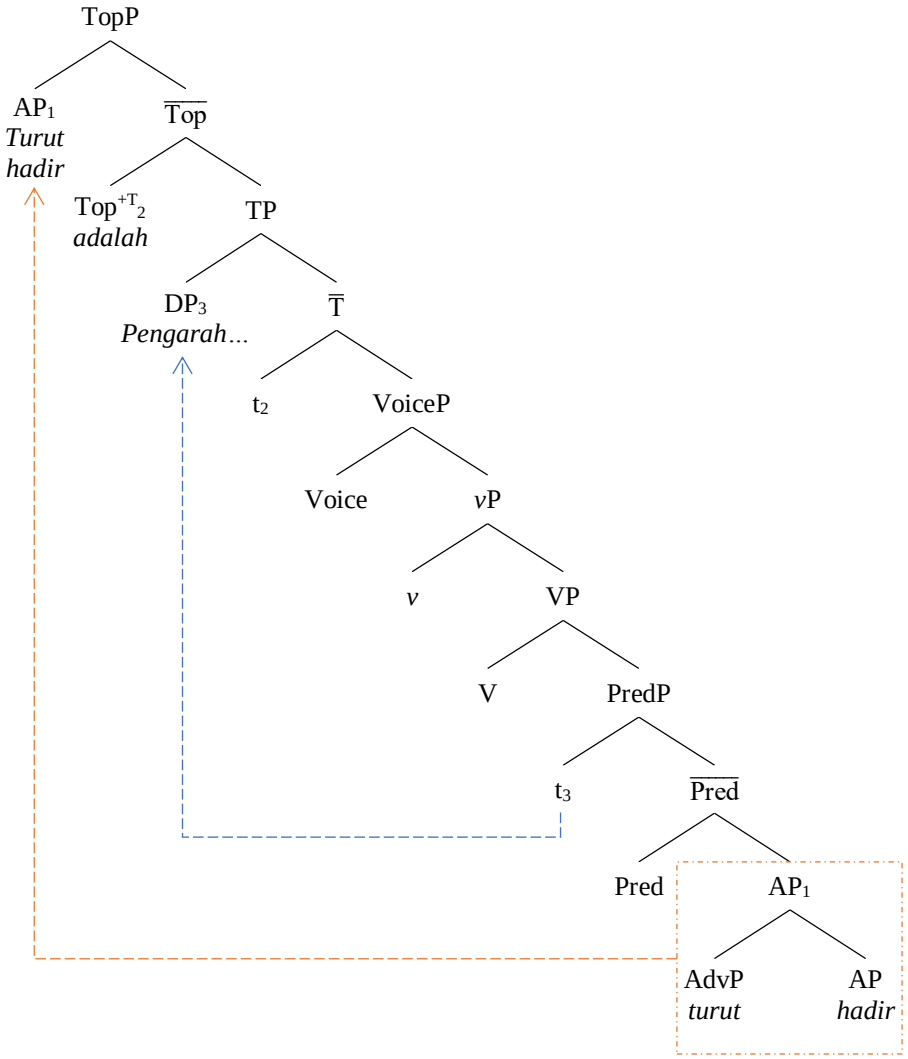


Figure 16: The Derivation of an Inverted Adjectival Predicate

Ultimately, the difference in derivation and choice of copula between specificational and inverse copular clauses demonstrates the sensitivity of the copula to the grammatical subject in the case of specificational copular clauses, but not predicational or inverse copular clauses. The copula in copular inversion is *adalah* due to the fact that the inverted predicate does not correspond to the grammatical subject. As a predicational copula and an elsewhere item, *adalah* has no particular feature that constrains the grammatical subject, unlike the copula in specificational copular clauses, which possesses the [*u*Top] feature that specifically triggers the predicate to move to SpecTP.

3.3.4 Extraction from Copular Clauses

Recall from Section 1.5.13 that extraction of the trigger in verbal clauses results in the formation of a cleft construction, as shown below:

(207) *Siapa*$_1$ *-kah yang* t$_1$ *ter-tumbuk se-orang guru?*
who -Q COMP NVOL-punch one-CLF teacher
'Who (was it that) punched a teacher?'

One interesting fact about copular clauses in Malay is that extraction from other than predicational copular clauses never forms a cleft with *yang*. Even more interesting is the subject vs. non-subject asymmetry, such that only extraction of the subject does this, as illustrated below:

Context: *Jika Zul adalah doktor…* (If Zul is a doctor…)[46]

(208) a. *Siapa-kah yang se-orang guru?* (Pre-Copular)
who-Q COMP one-CLF teacher
'Who is a teacher?'

b. *Apa-kah* (**yang*) *Ali?* (Post-Copular)
what-Q COMP A.
'What is Ali?'

Extraction of neither DP in an equative copular clause spells out *yang*:

Context: *Jika sebenarnya dia ialah Ali…* (If actually he is Ali…)

(209) a. *Siapa-kah* (**yang*) *Zul?* (Pre-Copular)
who-Q COMP Z.
'Who is Zul?'

b. *Siapa-kah* (**yang*) *kamu?* (Post-Copular)
who-Q COMP 2.SG
'Who are you?'

[46] Extraction from copular clauses in Malay is tricky because it is difficult to identify whether it is the pre- or post-copular constituent that has been extracted due to the impossibility of stranding inflectional elements. Using long-distance movement does not help either, as the grammaticality judgements on extracting either the subject or the predicate are bad. To overcome this, a conditional clause preceding the copular clause may be used to prime the kind of copular clause in the following clause to have a parallel structure. For example, if the conditional clause includes a specificational copular clause, the following copular clause is primed to be specificational as well, hence fixing the position of each DP.

The same goes for specificational copular clauses – a cleft is not formed when the post-copular constituent is extracted. Also, extraction of the pre-copular constituent is not possible, regardless of *yang*, due to the fixed information-structural alignment of specificational copular clauses. The fixed topic in pre-copular position may not correspond to a *wh*-phrase, which is a focus.

Context: *Jika nilai x ialah dua…* (If the value of *x* is two…)

(210) a. **Apa-kah* (*yang*) *lima?* (Pre-Copular)
what-Q COMP five
(What is five?)

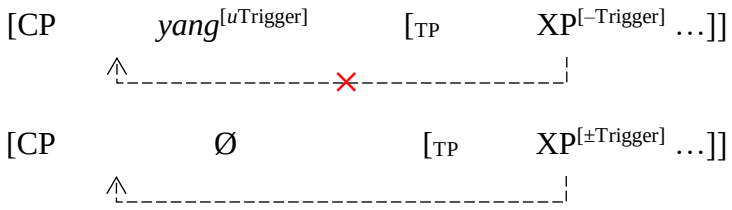
b. *Apa-kah* (**yang*) *nilai y?* (Post-Copular)
what-Q COMP value *y*
'What is the value of *y*?'

The table below summarises these extraction patterns:

	Pre-Copular DP	Post-Copular DP
Predicational	✓	-
Specificational	*	-
Equative	-	-

Table 19: The Spell-Out of 'Yang' in Copular Clauses Following Extraction

Based on the observations above, suppose that there is an agreement feature on *yang* that can only be satisfied by movement of the trigger of the clause to its specifier, much like work by Rizzi (1990) with respect to English *that*. Rather than describing the data observed above as a condition on the spell-out of *yang*, it can be analysed that the C^0 head that is spelt out as *yang* has this feature, say [*u*Trigger], as opposed to a null C^0 head that does not have the feature. Therefore, the merging of *yang* as the head of CP requires movement of the trigger of the clause, whilst merging of Ø allows movement of whatever constituent.



The anything-goes situation with the null C^0 head correctly predicts that it is possible to move the trigger of a clause without *yang*, since movement is free. This makes it seem as though the spell-out of *yang* is optional, as illustrated below with an intransitive verbal clause whose subject has been extracted. However, the apparent optionality is simply an illusion that is borne from the extraction of the trigger in two different clauses, namely one with *yang* and one with a null C^0 head.

(211) *Siapa* (*yang*) *mati?*
who COMP die
'Who died?'

This rule also correctly rules out clefting of adjuncts, as illustrated in (212). The trigger of a clause in Malay always corresponds to the grammatical subject (agent in the active voice and patient in the passive and patient voices), as described in Section 1.5.13. According to Erlewine, Levin, and Urk (2017), this constituent is the "privileged argument". Since adjuncts may not be identified as triggers, as they are not arguments and do not correspond to the grammatical subject, they do not satisfy the [*u*Trigger] feature on *yang* and clefting is impossible.

(212) a. *Bila-kah* (**yang*) *dia mati?*
when-Q COMP 3.SG die
'When did he die?'

b. *Di mana-kah* (**yang*) *dia mati?*
LOC where -Q COMP 3.SG die
'Where did he die?'

c. *Bagaimana-kah* (**yang*) *dia mati?*
how-Q COMP 3.SG die
'How did he die?'

According to this analysis, extraction of the subject of a predicational copular clause correctly satisfies the feature on *yang* as it is the only argument in the clause, and it corresponds to the surface subject. After all, predicational copular clauses are parallel to intransitive verbal clauses with respect to the argument status and syntactic position of the grammatical subject.

Equation does not involve predication, which makes identification of a trigger impossible. Since both the pre- and post-copular DPs are referential, neither DP may be identified as a trigger due to their equal status and due to the lack of a predicate that could help identify the trigger. Therefore, it is not possible to cleft either DP in equative copular clauses. Only extraction without *yang* is possible.

As for specificational copular clauses, things are slightly more complex. Since the grammatical subject of a specificational copular clause corresponds to the predicate and the referential DP does not occupy the subject position, as described in Section 3.3.2 and as argued by other scholars (e.g. Moro, 2006), neither constituent may be identified as the trigger of the clause, hence the impossibility of the referential DP to be clefted. Essentially, this extraction pattern provides evidence for the inversion analysis. The impossibility of extracting either DP from a specificational copular clause with *yang* indicates that there is no trigger, since the underlying subject does not correspond to the grammatical subject, hence there being no trigger. Therefore, only the use of the null C^0 head is possible in equative and specificational copular clauses.

Therefore, in all cases in which clefting is impossible (e.g. extraction of adjuncts, either DP in equative copular clauses, the referential DP in specificational copular clauses, and the predicate in predicational copular clauses) the null C^0 head must be used. Among the three types of copular clauses, it is only the pre-copular constituent of a predicational copular clause, which is unequivocally the trigger and grammatical subject of the clause, that may be clefted.

Ultimately, clefting in both verbal and nonverbal clauses in Malay is highly dependent on the constituent to be extracted being the privileged argument. The only difference between verbal and nonverbal clauses with respect to clefting is that there is an additional filter on extraction in verbal clauses, which is voice morphology. The correct voice morpheme must be ensured before extraction can be applied, whether it be clefting with *yang* or movement without *yang*. To illustrate, regardless of the choice of the C^0 head as either *yang* with the [*u*Trigger] feature or the featureless Ø, the object may not be extracted from a clause whose verb is affixed by the active voice marker.

(213) **Apa* {*yang* /Ø } *Ali meŋ-beli?*
what COMP COMP A. ACT-buy
'What did Ali buy?'

Since copular clauses do not include any sort of voice morphology, either the pre- or post-copular constituent (notwithstanding the pre-copular constituent of specificational copular clauses, which must obey the topic-focus alignment) may be extracted in clauses without *yang*, as is also true of voiceless verbal clauses in spoken Malay, e.g. example (80) repeated below:

(214) a. *Siapa* {*yang* /Ø } *beli nasi itu?* (Subject)
who COMP COMP buy rice DIST
'Who bought that rice?'

b. *Apa* {*yang* /Ø } *Ali beli?* (Object)
what COMP COMP A. ACT-buy
'What did Ali buy?'

This demonstrates that extraction, be it from verbal or nonverbal clauses, is free, provided that the C^0 head is the featureless null variant. What makes verbal and nonverbal clauses differ with respect to extraction is the restriction on the trigger that is voice marking. Meanwhile, when *yang* is involved, an additional restriction comes into play, as only the trigger may be clefted. Clefting seems to act on the same feature that voice does in identifying the trigger of a clause, so both *yang* and voice marking should have the same feature, which is the [*u*Trigger] feature. However, the possibility of extracting and clefting either the subject or object in voiceless verbal clauses in spoken Malay entails that the trigger vs. non-trigger distinction has broken down.

There is further evidence of the conditions on clefting in other areas of Malay grammar. It extends to a couple of related phenomena that involve the use of *yang*. For example, in relative clauses, the constituent relativised by *yang* must also correspond to the trigger. Since the subject of predicational copular clauses can be identified as the trigger, relativisation of the subject is licit, as shown below:

(215) *Ali, yang se-orang guru, akan ber-sara esok.*
A. REL one-CLF teacher PROS INTR-retire tomorrow
'Ali, who is a teacher, will retire tomorrow.'

However, when it is the predicate that is relativised, the resulting relative clause is illicit, as in example (216a). Also, this condition is the reason why there are no relative equative copular clauses in Malay, such as example (216b) in which one of the two referential DPs in the equative copular clause is relativised. Both examples are ungrammatical due to the extraction of a non-trigger, regardless of the copula.

(216) a. **Saya hendak jadi perintis bidang linguistik yang Ali* (*adalah*).
1.SG want become pioneer field linguistics REL A. COP
(I want to become the pioneer in the field of linguistics that Ali is.)

b. **Saya benci watak Zul yang Ali* (*ialah*) *dalam lakon-an itu.*
1.SG hate character Z. REL A. COP in act-NMZ DIST
(I hate the character Zul, whom Ali was in the play.)

One might assume that a null relativiser could be used, as is common in object relatives, given that the null C^0 head allows free extraction; however, it appears that relative copular clauses (other than subject predicational relative clauses) are ungrammatical regardless of the relativiser, as shown below. This is due to the fact that the copulas cannot be stranded. As a result, what remains in the relative copular clause is the referential DP and the clause does not include any material that could be predicated of it. Given the [+Pred] feature on the C^0 head of relative clauses, according to Rizzi (1990), this situation renders the relative clause ungrammatical.

(217) a. **Perintis bidang linguistik Ø Ali…*
pioneer field linguistics A.
(The pioneer in the field of linguistics Ali is…)

b. **Watak Zul Ali Ø dalam lakon-an itu…*
character Z. A. in act-NMZ DIST
(The character Zul, Ali was in the play…)

Even when the copular clause is embedded under a verbal clause, it remains ungrammatical as the copular clause is without its own predicate.

(218) a. **Perintis bidang linguistik* (*yang*) *awak fikir Ø Ali…*
pioneer field linguistics COMP 2.SG think A.
(The pioneer in the field of linguistics that you think Ali is…)

b. **Watak Zul* (*yang*) *awak fikir Ali Ø dalam lakon-an itu…*
character Z. COMP 2.SG think A. in act-NMZ DIST
(The character Zul that you think Ali was in the play…)

To remedy the ungrammaticality, a clear verbal predicate must be used:

(219) a. *Perintis bidang linguistik yang di-cermin-kan Ali…*
pioneer field linguistics REL PASS-reflect-APPL A.
'The pioneer in the field of linguistics that is reflected by Ali…'

b. *Watak Zul yang di-main-kan oleh Ali dalam lakon-an itu…*
character Z. REL PASS-play-APPL by A. in act-NMZ DIST
'The character Zul, who is played by Ali in the play…'

Also related is the use of ligature *yang* to link a NP to a post-modifier. It is commonly used between a NP and a modifier, as illustrated in examples (220a-c). The relationship between *yang* and the modificand is such that the modificand stands in an attributive relation with the modifier. As there is no attribution or modification between possessee and possessor, *yang* cannot be used in possessive DPs, as in (220d).

(220) a. *Kucing yang ini* (Demonstrative – Modifier)
cat LIG PROX
'This cat'

b. *Kucing yang comel* (AP – Modifier)
cat LIG cute
'A cute cat'

c. *Kucing yang peŋ-musnah* (NP – Modifier)
cat LIG AG-destroy
'A destroyer of a cat

d. **Kucing yang dia* (Possessor – Referent)
cat LIG 3.SG
(His cat)

Considering that the possessor is not an attribute, but a referential DP, a feature corresponding to the [*u*Trigger] in NP modification prevents the use of *yang* in possessive DPs in Malay. Instead, *punya* (to own) must be used to act as a predicate.[47]

[47] *Punya* could also arguably be a possessive marker because it may occur DP-internally without the need for a relativiser. Example (xiib) shows that *punya* is spelt out when the possessor has moved to the left of the possessee:

(xii) a. *Kucing dia itu*
cat 3.SG DIST
'His cat'

b. *Dia*$_1$ *punya* t$_1$ *kucing itu*
3.SG POSS cat DIST
'His cat'

(221) *Kucing yang dia punya*
cat LIG 3.SG own
'His cat/the cat that he owns'

In fact, it is quite difficult for clefting to apply in what could be a predicational copular clause if the nominal predicate is a definite or highly specific DP, due to the likeliness of it to be interpreted as a referential expression. The propensity of interpreting the predicate as a referential expression also causes the ambiguity between a predicational/equative/specificational interpretation of a copular clause. A scale from most referential to least referential can be made to formalise which DPs can and cannot be clefted in Malay:

Figure 17: The Scale of Referentiality

Thus, a bare predicative nominal is the easiest to cleft, as it is unambiguously predicative; *yang* cannot be used with referential DPs.

(222) a. *Siapa-kah yang guru?* (Bare)
who-Q COMP teacher
'Who is a teacher?'

b. *Siapa-kah yang se-orang guru?* (Indefinite)
who-Q COMP one-CLF teacher
'Who is a teacher?'

c. *Siapa-kah yang guru itu?* (Definite)[48]
who-Q COMP teacher DIST
'Who is the teacher?'

d. *Siapa-kah (*yang) Profesor Ali?* (Proper)
who-Q COMP Professor A.
'Who is Professor Ali?'

e. *Siapa-kah (*yang) dia?* (Personal Pronoun)
who-Q COMP 3.SG
'Who is he?'

[48] Given that definite DPs may be interpreted as a referential expression or predicates, a DP modified by demonstrative *itu* can only be clefted if it is interpreted as a predicate.

3.4 Summary

This chapter has presented a basic overview of the syntax of copular clauses in Malay. Several misconceptions about copular clauses in Malay have been addressed, one of which is the inadequacy of attributing the choice of copula to the syntactic category of the predicate. Arguably, it is the relationship between the constituents flanking the copula as well as the referentiality of those constituents, which is reflected by those relations, that does this. It is also prescribed that copula *adalah* may not combine with verbs. This is an unfounded claim against which this chapter has argued, citing various natural occurrences of the copula combining with verbs. The preverbal position of the copula, combined with other findings pertaining to word order, has also made it possible to conclude that the copulas are auxiliaries heading TP.

Several syntactic phenomena that demonstrate the privileged status of the subject in copular clauses in Malay have also been discussed in this chapter. The copula *ialah*, which is typically used in equative copular clauses, is reported to possess a selectional property that only allows 3rd person subjects to combine with it. However, the additional requirement that *ialah* combine with human DPs entails that the copula in equative copular clauses is variable. On the other hand, in specificational copular clauses, only copula *ialah* is possible. This fixed choice is analysed to be due to the specific [*u*Top] feature on the copula that attracts the topical predicate to move to the subject position, as confirmed by the lack of the feature on *adalah* in inverse copular clauses. Finally, the correspondence of subject with topic in specificational copular clauses alludes to Austronesian alignment. Clefting is only applicable to the subject of predicational copular clauses, owing to its unmistakeable trigger status (i.e. the argument in subject position). Whilst extraction from all copular clauses is possible (except for the grammatical subject of specificational copular clauses), thanks to the lack of voice morphology, only the subject of predicational copular clauses may be clefted, due to the [*u*Trigger] feature on *yang*. Parallel to the sole argument in intransitive verbal clauses, the subject in predicational copular clauses is identified as the trigger of the clause. Dissimilarly, neither of the DPs in equative copular clauses can be identified as a trigger as they both have equal status, whereas the underlying argument of specificational copular clauses cannot be identified as the trigger as it does not correspond to the grammatical subject.

Chapter 4: Overt Encoding of the Copulas

Copulas are often omitted from nonverbal clauses. This behaviour varies from language to language – the copulas of one language may show more frequent omission than those of another, as suggested by the *Continuum of Zero Copula Encoding* by Stassen (2013). Malay is one of the more extreme languages in the continuum as it allows the optionality of the copulas very frequently. In fact, there appear to be a variety of seemingly unrelated environments in which the copula is banned, as illustrated in (223b-f).[49]

(223) a. *Zul* (*adalah*) *jujur.*
Z. COP honest
'Zul is honest.'

b. *Zul* (**adalah*) *jujur selepas di-hukum.* (Tense)
Z. COP honest after PASS-punish
'Zul was honest after being punished.'

c. *Zul telah* (**adalah*) *jujur.* (Aspect)
Z. PRF COP honest
'Zul has been honest.'

d. *Zul boleh* (**adalah*) *jujur.* (Modality)
Z. can COP honest
'Zul can be honest.'

e. *Saya hendak Zul* (**adalah*) *jujur.* (Control)
1.SG want Z. COP honest
'I want Zul to be honest.'

f. *Saya buat Zul* (**adalah*) *jujur.* (Causative)
1.SG make Z. COP honest
'I made Zul be honest.'

From the perspective of the surface syntax, the distribution of the copulas might seem unpredictable at first; however, the existence of certain environments in which overt copulas are impossible, despite the widely held belief that the copulas are

[49] The restrictions examined in this chapter apparently only affect predicational copular clauses, which are those that employ copula *adalah*, as opposed to specificational or equative copular clauses, which employ copula *ialah*.

optional in all cases, is indicative of there being rules that govern overt encoding of the Malay copulas, i.e. the use of an overt/zero copula.

This chapter analyses the environments in which overt encoding of the copulas is banned. I argue that the semantics of the copular clause plays a salient role in the licensing of an overt copula. Specifically, changes in the aspectual nature of the copular clause affect overt encoding of the copula such that overt copulas may only occur in atemporal clauses, as opposed to clauses that possess a temporally bound interpretation, either by the aspectual properties of the predicate (lexical aspect) or by those of the clause (grammatical aspect). Grammatical aspect is analysed to compositionally interact with lexical aspect to yield differences in overt encoding of the copula, especially in cases of coercion.

4.1 Beyond Economy and Parsing

Copulas are often omitted from very simple copular clauses in Malay, such as examples (224), on the basis of economy and ease of parsing. Although the two factors do allow optionality of the copulas in many cases, they are not solely responsible for the spell-out of the copulas. As I shall argue in this chapter, it is aspect that actually governs overt encoding of the copulas, which greatly differs from the role of economy and parsing as a secondary filter in determining whether the copulas might be spelt out. In other words, aspect is the primary filter that determines whether a copula *could* be spelt out, whereas economy and parsing are an optional secondary filter that determines whether a copula *should* be spelt out.

(224) a. *Nama dia* (*ialah*) *Zul.*
name 3.SG COP Z.
'His name is Zul.'

b. *Dia* (*adalah*) *petani.*
3.SG COP farmer
'He is a farmer.'

The weight of the subject of a copular clause especially is known to be a determining factor in the optionality of a copula: "they are most common in noun clauses where either the subject or predicate is long, in which case they break up a string of nouns and add a smoothness to the construction" (Sneddon, 1996, p. 237).

When a sentence is easily understood, a copula is not used overtly for the sake of brevity, unless formality or register dictates it. The longer the sentence, the more likely a copula will be used.

Compare the examples above with example (225a), which illustrates that, without the copula, the interpretation can be ambiguous due to there not being a clear boundary between the subject and its nonverbal predicate. Although the copula is illustrated in the original example to be obligatory, it is more likely to be the case that the copula is highly preferred (marked with the double question mark), as ambiguity does not necessarily cause ungrammaticality.

(225) a. *Cita-cita Ali untuk meŋ-jadi peŋ-cipta robot* ??(*adalah*) *sia-sia.*
ambition A. for ACT-become AG-invent robot COP fruitless
'Ali's ambition to become a robot inventor is fruitless.'
(Mustaffa, 2015)

b. *Cita-cita Ali untuk meŋ-jadi kaya* (*adalah*) *sia-sia.*
ambition A. for ACT-become rich COP fruitless
'Ali's ambition to become rich is fruitless.'

The copula helps disambiguate the sentence and prevents a garden-path reading in which the adjectival predicate *sia-sia* is interpreted as a post-modifier to the DP *pencipta robot*. Without the copula, the example can be interpreted as "Ali's ambition to become a fruitless robot inventor…" as attributive adjectives in Malay are postnominal. A copula is needed to partition the clause into subject and predicate, thus disambiguating the sentence. When there is no ambiguity, as in (225b), the copula may be omitted freely.[50] Example (226) below presents an even more severe case in which, without the copula, the nominal predicate can be misunderstood as the thematic agent of the clause: "The Japanese man was believed by the North Korean agent."

(226) *Lelaki Jepun itu di-percaya-i* ??(*adalah*) *se-orang ejen Korea Utara.*
man Japan DIST PASS-believe-APPL COP one-CLF agent Korea North
'The Japanese man was believed to be a North Korean agent.'
(BERNAMA, 2017)

[50] Of course, the use of pause-intonation ameliorates the interpretation, but the same cannot be said of written language, which does not have any prosodic cues (other than the use of punctuation, e.g. commas), which is why the copulas are more common in written Malay.

The preference for an overt copula in these ambiguous contexts can be construed to decrease the cognitive strain of parsing heavy constituents and facilitate identification of the subject-predicate boundary. Complex sentences are prone to misunderstanding as parsing becomes increasingly difficult with greater constituent weight and grammatical complexity. This cognitive factor affects the overt realisation of a copula insofar as the use of an overt copula becomes highly preferred, lest the utterance risk being misunderstood. Similarly, such a strain on the cognitive process is known to affect the structure of a sentence and trigger syntactic phenomena such as *it*-extraposition and heavy-NP shift, which occur to ease parsing by postposing a heavy constituent, such as a clausal subject or a long and complex DP to the end of a sentence.

Although it might appear that the copulas in Malay are largely optional, as illustrated in (224), this is not entirely true. There are also cases in which copula *adalah* in particular is impossible, which clearly suggests that there are underlying factors that dictate whether a copula is possible in a given clause. In example (227b), the copula is totally ungrammatical. If something truly is optional, there should not be cases in which it is either obligatory or impossible. *Ceteris paribus*, the difference in the nonverbal predicate of the copular clauses in examples (227a-b) apparently plays some sort of role.

(227) a. *Langit adalah biru.* (Permanent)
sky COP blue
'The sky is blue.'

b. *Langit* (**adalah*) *mendung.* (Temporary)
sky COP cloudy
'The sky is cloudy.'

The difference between the two nonverbal predicates is the lexical aspectual properties of the predicate, i.e. being blue is a permanent property, whereas being cloudy is a temporary state. Therefore, the most obvious place to start searching for answers regarding overt encoding of the copula would be the distinction between stage-level and individual-level predicates.

4.2 Aspect and Eventiveness of the Predicate

This section presents a brief view and application of the individual-level vs. stage-level distinction by Carlson (1977b) and the Davidsonian vs. Kimian states by Maienborn (2008). These notions are very important in this chapter as they have a large (albeit not total) capacity to characterise the restrictions on overt encoding of the copulas in Malay. It is revealed that there is a clear correspondence between the two notions, which predictably affect overt encoding of the copulas in Malay, as summarised below:

Predicate Type	Eventuality	Encoding of Copula
Stage-Level	Event (Davidsonian State)	Zero
Individual-Level	State (Kimian State)	Overt

Table 20: The Correspondence among Predicate Type, Eventuality, and Encoding of the Copula

4.2.1 Individual-Level vs. Stage-Level

Based on examples (227), overt encoding of the copulas is affected by lexical aspect. It is then possible to approach the data using the individual-level vs. stage-level framework by Carlson (1977b). The permanent interpretation of an individual-level predicate is attributed to the property being ascribed directly to the individual, as opposed to it being ascribed to a stage of the individual. Stage-level predicates concern a temporally bound slice in time of the individual, whilst individual-level predicates concern the individual itself regardless of time (Carlson, 1977a).

In accordance with the framework, overt Malay copulas are analysed to combine with individual-level predicates, viz. properties that describe the individual without any change over time. Stage-level predicates do not warrant the use of an overt copula as they describe a stage of an individual in time. So, the inability of the copulas to combine with stage-level predicates explains the example provided by Arka (2013), whose ungrammaticality is erroneously judged to be a sign that *adalah* cannot combine with adjectives in general. *Sakit* is a stage-level predicate, which does not warrant overt encoding of the copula.

(228) *John* (**adalah*) *sakit.*
John COP ill
'John is ill.'
(Arka, 2013, p. 31)

This pattern extends to all categories of predicates. The following examples contrast an individual-level predicate with a stage-level predicate:

(229) a. *Zul adalah se-orang lelaki.* (Nominal)
Z. COP one-CLF man
'Zul is a man.'

b. *Zul* (**adalah*) *ke-hujan-an.*
Z. COP NMZ-rain-NMZ
'Zul is drenched.'

c. *Zul adalah matang.* (Adjectival)
Z. COP mature
'Zul is mature.'

d. *Zul* (**adalah*) *risau.*
Z. COP worried
'Zul is worried.'

e. *Zul adalah seperti bidadari.* (Prepositional)
Z. COP like angel
'Zul is like an angel.'

f. *Zul* (**adalah*) *dalam ke-sakit-an.*
Z. COP in NMZ-hurt-NMZ
'Zul is in pain.'

Apparently, this phenomenon has not gone unnoticed in the languages of the world. That copulas in different languages exhibit this sensitivity is not a new finding, as it has been noted by Stassen (1997) and Pustet (2003) that factors such as the permanence or temporal stability of the predicate influences nonverbal predication in one way or another. In some languages, permanence might influence the syntactic category of the predicate itself. Consider the Māori examples in (230), cited in Stassen (1997), in which the permanence of the predicate is concomitant with how a property-concept word is encoded – as either a verbal or nonverbal predicate. Example (230a) is a clause employing the so-called verbal strategy in encoding the property-concept word *pai* (good). *Pai* is encoded as a verb with aspectual particle *ka*, commonly

associated with verbs. On the other hand, in example (230b), it is encoded as an indefinite nominal with indefinite marker *he*, hence the nominal strategy. This difference in encoding strategy is argued to be borne from the application of what Stassen (1997) calls the *Permanency Parameter*, such that a non-permanent interpretation of a property licenses verbal encoding of the predicate, whereas a permanent interpretation licenses nominal encoding of said predicate.

Māori (Austronesian – Polynesian)

(230) a. *Ka pai te whare nei.* (Temporary)
INCEP good ART house PROX
'This house is good.'

b. *He pai te koorero.* (Permanent)
ART.NDEF good ART talk
'The talk is good.'
(Biggs, 1969, p. 24)

In Spanish, copulas *ser* and *estar* are used in accordance with the individual-level vs. stage-level distinction, as argued by Arche et al. (2019). With the use of *estar* in example (231a), the predicate is interpreted as a stage-level predicate – the subject looks handsome at the time of the utterance, e.g. after a nice shave and haircut – whereas the predicate is interpreted as an individual-level predicate with the use of *ser* in example (231b) – the subject is always handsome, e.g. has nice facial features.

Spanish (Indo-European – Romance)

(231) a. *Roberto Alcázar está guapo.* (Temporary)
R. COP^{estar} handsome
'Roberto Alcázar is handsome (now).'

b. *Roberto Alcázar es guapo.* (Permanent)
R. COP^{ser} handsome
'Roberto Alcázar is handsome (always).'

Most interestingly, in Akawaio, this permanent-temporary dichotomy is argued by Overall, Vallejos, and Gildea (2018) to have an effect such that temporary states license the use of an overt copula, and vice versa, which is the total opposite of the observations regarding the Malay copulas.

Akawaio (Cariban)

(232) a. *Juwaŋ be maŋ.* (Temporary)
hunger ATTR 3.COP
'He is hungry (now; a fact).'

b. *Juwaŋ kɨrə-rə.* (Permanent)
hunger 3.AN-EMPH
'He's hungry (always).'
(Meira & Gildea, 2009, p. 109)

In all of the crosslinguistic examples above, the difference in the copula or category of the predicate correlates with permanence, which can be said to reflect the lexical aspect of the predicate. These properties have a one-to-one correspondence with the stage-level vs. individual-level distinction such that stage-level predicates are temporary and individual-level predicates are permanent. An overt copula in Malay may never combine with stage-level predicates, as they only describe the subject at a moment in time, which entails that the duration for which the predicate applies to the subject is limited and will cease, hence a change of state. In contrast, an overt copula is able to combine with individual-level predicates, which describe the subject as a whole, never-ceasingly, and at all times, hence no change of state. The (im)permanence of the stage-level vs. individual-level distinction is a property that is inherent to the predicate, as opposed to the surrounding elements such as the subject, the syntactic structure, or the grammatical construction in which the predicate occurs.

4.2.2 States vs. Events

Other than the stage-level vs. individual-level distinction, one can consider the notion of eventuality in the Davidsonian sense to have a part to play, as it does appear that aspect adds a dimension of event semantics into copular clauses, especially in examples involving a change of state. After all, there has been a large body of research linking changes of states to events, dating back to the 90s, e.g. Lakoff (1965), McCawley (1968), Dowty (1979). Maienborn (2008) provides the following tests to diagnose Davidsonian states (i.e. events), which are the opposite of what she calls Kimian states. According to the diagnostics, stage-level predicates in Malay are confirmed to be eventive. As illustrated in (233), *sakit* (ill) may occur with the diagnostics which identify it as an event.

- Locative modifier: ✓
- Manner adverb: ✓
- Non-finite complement of perception verb: ✓

(233) a. *Zul sakit di hospital.* (Locative)
Z. ill LOC hospital
'Zul is ill at hospital.'

b. *Zul sakit dengan sengsara=nya.* (Manner Adverb)
Z. ill with miserable=3
'Zul is ill miserably.'

c. *Tadi saya tengok Zul sakit.* (Perception Verb)[51]
just.now 1.SG watch Z. ill
'Just now I watched Zul be ill.'

This even applies to nominal predicates that encode temporary states, as illustrated below:

(234) a. *Zul ke-hujan-an di luar.* (Locative)
Z. NMZ-rain-NMZ LOC out
'Zul is drenched outside.'

b. *Zul ke-hujan-an sepenuhnya.* (Manner Adverb)
Z. NMZ-rain-NMZ completely
'Zul is drenched completely.'

c. *Tadi saya tengok Zul ke-hujan-an.* (Perception Verb)
just.now 1.SG watch Z. NMZ-rain-NMZ
'Just now I watched Zul get drenched.'

Considering that stage-level predicates in Malay correspond to events, individual-level predicates should be the opposite of events, i.e. Kimian states. As expected, individual-level predicates are shown to be unable to accommodate the diagnostics for events, as illustrated in (235). Regardless of whether the copula is overt or null, the sentences are ungrammatical, which confirms that individual-level predicates are Kimian states, and not events.

[51] The use of a spatiotemporal adverbial is required to create a context in which a situation is perceived because perception verbs in Malay can take finite clausal complements with null complementisers, which are difficult to distinguish from non-finite clauses due to the lack of morphological marking.

(235) a. **Zul* (*adalah*) *raksasa di dalam bilik=nya.* (Locative Adverb)
Z. COP monster LOC in room=3
(Zul was a monster in his room.)

b. **Zul* (*adalah*) *raksasa laju-laju.* (Manner Adverb)
Z. COP monster fast-RED
(Zul was a monster fast.)

c. **Tadi saya tengok Zul* (*adalah*) *raksasa.* (Perception Verb)
just.now 1.SG look Z. COP monster
(Just now I watched Zul be a monster.)

That the diagnostics confirm the eventuality of stage-level predicates is not all too surprising as there have been analyses linking them to events. Most prominently, Kratzer (1995) argues that stage-level predicates are "*Davidsonian* in that they have an extra argument position for *events* or *spatiotemporal locations*" (p. 126).

In sum, stage-level predicates in Malay are Davidsonian states, or simply events, which makes it possible for them to be located spatiotemporally. Conversely, individual-level predicates in Malay are Kimian states. Clearly, the lexical aspectual properties of the two types of predicates in Malay also correspond to a difference in eventuality, both of which subsequently affect overt encoding of the copulas in Malay. The eventive nature of stage-level predicates is incompatible with overt copulas, whilst individual-level predicates are compatible. These findings reveal a clear-cut distinction and satisfying correspondence between stage-level vs. individual-level predicates, Davidsonian vs. Kimian states, and overt vs. null encoding of the Malay copulas.

4.3 Coercion and Grammatical Aspect

Although the lexical aspect of the predicate indeed plays a part in the licensing of overt copulas, it is very much possible for a copula to be rendered null by the grammatical aspect of the clause. The Malay copulas are sensitive to changes in information that expresses time, as aspectual markers and certain temporal adverbials, among other material, affect the copulas in a way that renders overt copulas ungrammatical. To illustrate, the following examples show how overt encoding of the copulas is affected by altering the temporal or aspectual information of the clause by adding adverb *dulu* (previously) and aspectual marker *belum* (IMPRF).

(236) a. *Kawasan ini adalah selamat.*
area PROX COP safe
'This area is safe.'

b. *Kawasan ini belum* (**adalah*) *selamat.*
area PROX IMPRF COP safe
'This area is not yet safe.'

c. *Dulu kawasan ini* (**adalah*) *selamat.*
previously area PROX COP safe
'Previously, this area was safe.'

Selamat (safe) is typically an individual-level predicate that encodes a permanent property; however, the copula is not possible in examples (236b-c) because the use of the adverb and the imperfective marker creates a temporally bound interpretation of the predicate such that the state denoted by the predicate has a starting or ending point, as opposed to an atemporal reading in which the duration of the state is indefinite. The presupposed starting or ending point entails a change to or from the state denoted by the predicate, which apparently is not compatible with the strictly atemporal nature of the copula. The lack of any specification with regard to time in example (236a) allows the atemporal interpretation of the predicate, which licenses an overt copula. This atemporality gives nonverbal predicates the same effect as habitual verbal predicates such that statements are true regardless of time.

Examples (236b-c) additionally demonstrate that the use of the individual-level predicate in a temporally bound environment has coerced it into a stage-level predicate, meaning that the predicate *selamat* only applies to a stage of the subject, and not the subject as a whole. In consequence of the stage-level interpretation, overt encoding of the copula is disallowed. In contrast, consider *sakit* (ill), a temporary state that does not warrant overt encoding of the copula. One could hypothesise that creating a context to induce a permanent or atemporal reading of the nonverbal predicate should allow an overt copula to surface; however, this prediction is not borne. In example (237b) the copula remains null due to the inherent lexical aspect of the predicate.

(237) a. *Kucing itu* (**adalah*) *sakit.* (Stage-Level – Temporary)
cat DIST COP ill
'That cat is ill.'

b. *Kucing itu* (**adalah*) *kekal sakit.* (Stage-Level – Permanent)
cat DIST COP permanent ill
'That cat is permanently ill.'

Examples (236) and (237) demonstrate that, although constructing a context that allows a temporary reading of an individual-level predicate results in the impossibility of overt encoding of the copula, the reverse does not hold; coercing a permanent reading of a stage-level predicate does not allow the copula to surface. Strictly speaking, it is not possible to coerce an atemporal reading of a stage-level predicate because, by definition, stage-level predicates concern a temporally bound slice in time of the individual, as opposed to individual-level predicates, which concern the individual itself regardless of time (Carlson, 1977a). Trying to coerce a permanent interpretation of a stage-level predicate merely equates to combining stages to create a series of consecutive temporary states, and not a description of the individual as a whole. Essentially, coercing a temporally bound interpretation of an individual-level predicate transforms it into a stage-level predicate, whereas coercing an atemporal interpretation of a stage-level predicate does not transform it into an individual-level predicate. As such, stage-level predicates are always temporally bound. Thus, no matter how the context is constructed, *adalah* can never combine with *sakit*.

As for the eventiveness of coerced predicates, the diagnostics for events reveal that individual-level predicates that have been coerced are eventive. To illustrate, the predicate *jujur* (honest) in examples (238) is interpreted to be eventive. The sentences are grammatical with *jujur*, but the copula must be omitted as the copula may not occur in eventive clauses.

(238) a. *Zul* (**adalah*) *jujur di mahkamah.* (Locative Adverb)
Z. COP honest LOC court
'Zul was honest at court.'

b. *Zul* (**adalah*) *jujur dengan tenang=nya.* (Manner Adverb)
Z. COP honest with calm=3
'Zul was honest calmly.'

c. *Tadi saya tengok Zul* (**adalah*) *jujur.* (Perception Verb)
just.now 1.SG watch Z. COP honest
'Just now I watched Zul be honest.'

It should be noted that the examples above no longer possess an atemporal or generic reading. If the predicate truly has become stage-level, it should be possible for it to occur in existential clauses, provided that it occurs with the diagnostics that identify it as an event. Being able to occur in existential clauses entails that a predicate is stage-level, as individual-level predicates are banned from existential clauses. Using the locative adverbial, *jujur* may occur in an existential clause and can be concluded to be a stage-level predicate when in an existential clause.

(239) *Ada orang telah jujur di mahkamah semalam.*[52]
EXIST person PRF honest LOC court yesterday
'There was a person having been honest at court yesterday.'

On the other hand, even when given an atemporal interpretation, stage-level predicates are able to accommodate the diagnostics for events. As illustrated below, the temporal adverbs used promote an atemporal reading of the stage-level predicates and the examples are grammatical:

(240) a. *Zul* (**adalah*) *selalu lapar di rumah.* (Locative Adverb)
Z. COP always hungry LOC home
'Zul is always hungry at home.'

b. *Zul* (**adalah*) *selalu cepat lapar.* (Manner Adverb)
Z. COP always fast hungry
'Zul always (gets) hungry fast.'

c. *Saya tengok Zul* (**adalah*) *selalu lapar.* (Perception Verb)
1.SG watch Z. COP always hungry
'I see Zul always hungry.'

They are also grammatical in existential clauses, hence are eventive.

(241) *Ada orang sentiasa lapar di Afrika sana.*
EXIST person always hungry LOC Africa there
'There are people always hungry in Africa.'

[52] The perfective marker is used to make sure that the predicate is not interpreted as a postnominal modifier.

To summarise, individual-level predicates transform into stage-level predicates and become eventive via coercion, which prevents overt encoding of the copula. Conversely, stage-level predicates are inherently temporally bound and eventive regardless of coercion, which likewise prevents overt encoding of the copula. Setting the context for a permanent interpretation of a stage-level predicate does not alter the lexical aspect that is inherent to the state. The following table summarises the spell-out patterns of the copula following interaction between lexical and grammatical aspect:

Grammatical Aspect	**Lexical Aspect**	
	Stage-Level	Individual-Level
Atemporal	–	State COP → *adalah*
Temporally Bound	Event COP → Ø	(Coercion) Event COP → Ø

Table 21: Aspectual Interaction and the Resulting Eventiveness

Clearly, lexical aspect alone does not account for the full range of spell-out patterns observed of the copula in Malay, as is clear in cases of coercion. Grammatical aspect also affects overt encoding of the copula as it does eventiveness. Otherwise, the copula should be able to surface when an individual-level predicate undergoes coercion to obtain a temporally bound reading. The context in which the predicates occur and the additional linguistic material that may construct those contexts lie beyond the stage-level vs. individual-level framework. The context and the additional linguistic material introduce another dimension of aspect to the copular clause, namely grammatical aspect, in addition to the predicate-inherent lexical aspect.

4.4 Inner AspP as the Main Factor

Reconsider examples (223b-f), repeated below:

(242) a. *Zul* (**adalah*) *jujur selepas di-hukum.* (Tense)
Z. COP honest after PASS-punish
'Zul was honest after being punished.'

b. *Zul telah* (**adalah*) *jujur.* (Aspect)
Z. PRF COP honest
'Zul has been honest.'

c. *Zul boleh* (**adalah*) *jujur.* (Modality)
Z. can COP honest
'Zul can be honest.'

d. *Saya hendak Zul* (**adalah*) *jujur.* (Control)
1.SG want Z. COP honest
'I want Zul to be honest.'

e. *Saya buat Zul* (**adalah*) *jujur.* (Causative)
1.SG make Z. COP honest
'I made Zul be honest.'

These examples present a curious case as, rather than being merely optional, the copula is banned altogether. It is apparent that the copula is impossible in a wide range of constructions that are not syntactically related. For example, control clauses and causative clauses are syntactically distinct, but the copula is impossible in both environments, nonetheless. From a syntactic perspective, it does not seem possible to generalise over all the examples to explain the ban on the copula as they are all too varied. On the other hand, a semantic approach seems more promising; examples (242) share the common semantic property that makes available an interpretation of the predicate presupposing a change of state or its possibility. There is an aspectual factor that presupposes information regarding the (in)completion of an event that would culminate in the achievement of the state or property denoted by the predicate or its potentiality. For instance, *jujur* is not typically a temporary state, but the use of the control clause in (242d) implies that the subject does not possess the permanent property at the time of the utterance but is wanted to achieve the property at an unspecified time, hence the change-of-state presupposition, i.e. in wanting Zul to be honest, there is a presupposed change of state from dishonesty to honesty.

Following from the discovery of this semantic relation, I argue that the root of the cause for the ban on overt copulas is aspect. The copula may only be overtly realised in copular clauses with predicates that remain constant and do not change over time, i.e. lack an interpretation that involves or presupposes changes of states (carried either by grammatical or lexical aspect). For example, *lapar* (hungry) is a temporary state that changes and does not warrant overt encoding of the copula, whilst *benar* (true) is a permanent property that does not change and warrants overt encoding of the copula. Clauses that carry the change-of-state interpretation are eventive in nature, which clashes with the strictly stative property of the copula. In other words, if a clause has some grammatical or lexical aspectual factor that carries a change-of-state interpretation, it is eventive and does not permit overt encoding of the copula. Conversely, if a clause does not have any such aspectual factor, it is stative and allows overt encoding of the copula.

This reasoning is straightforward for predicates that encode either permanent properties or temporary states, as they either correspond to states or events. Therefore, an analysis that combines the wisdom from the individual-level vs. stage-level distinction by Carlson (1977b) and the Davidsonian vs. Kimian states by Maienborn (2005) should prove fruitful in modelling the ban on the copula from certain copular clauses. Simply put, overt copulas may only combine with individual-level predicates in Malay, which are strictly stative, and are incompatible with stage-level predicates, which are events, as illustrated below:

(243) a. *Zul adalah jujur.* (Individual-Level
Z. COP honest
'Zul is honest.'

b. *Zul* (**adalah*) *lapar.* (Stage-Level)
Z. COP hungry
'Zul is hungry.'

Stage-Level (Event)	Individual-Level (State)
COP → Ø	COP → *adalah*

Table 22: The Spell-Out of the Copula According to Predicate Type

In accordance with scholars such as Batllori (2011), Gallego and Uriagereka (2016), and Arche, Fábregas, and Marín (2017) with regard to copular clauses in Spanish, I extend the analysis to Malay that copular clauses with stage-level predicates are more syntactically complex than those with individual-level predicates. Their treatment of the distinction between copulas *ser* and *estar* in Spanish – the former combines with individual-level predicates, whilst the latter combines with stage-level predicates – involves different underlying syntactic structures associated with either copula. Specifically, the structure of copular clauses with *estar*, whose underlying form is one and the same as *ser*, is more complex as the copula combines with an aspectual head to yield *estar*. However, rather than having the copula combine with Asp^0, I argue that it is the projection of AspP within the *v*P, i.e. *inner aspect*, as per Travis (2010), that is responsible for the spell-out pattern of the copulas in Malay.[53] Given that stage-level predicates concern stages of an individual, they are always temporally bound, entailing the presence of the inner AspP in all copular clauses with stage-level predicates. Consequently, the presence of the inner AspP prevents overt encoding of the copula with stage-level predicates as the atemporal property of overt copulas in Malay clashes with the temporally bound property of clauses that include the inner AspP. On the other hand, copular clauses with individual-level predicates do not project an inner AspP, which permits overt encoding of the copula.

Only in cases of coercion to a temporally bound reading do copular clauses with individual-level predicates project an inner AspP. For instance, the following examples contrast an atemporal vs. temporally bound reading of the individual-level predicate *honest*. The individual-level predicate gets coerced into carrying a temporally bound interpretation as the state denoted by the predicate applies for a limited duration of time, owing to the temporal specification by the *when*-clause. The same phenomenon is discussed by Comrie (1976) and Partee (1977) in what is called the *active-be* construction: "He is honest" vs. "He is being honest".

(244) a. *He is honest.* (Individual-Level)

b. *He is honest when he is drunk.* (Coerced)

[53] Besides, the copulas in Malay are merged in T^0, and lowering does not seem to be an option that is viable in Malay.

The diagram below illustrates the presence of the inner AspP in temporally bound copular clauses.

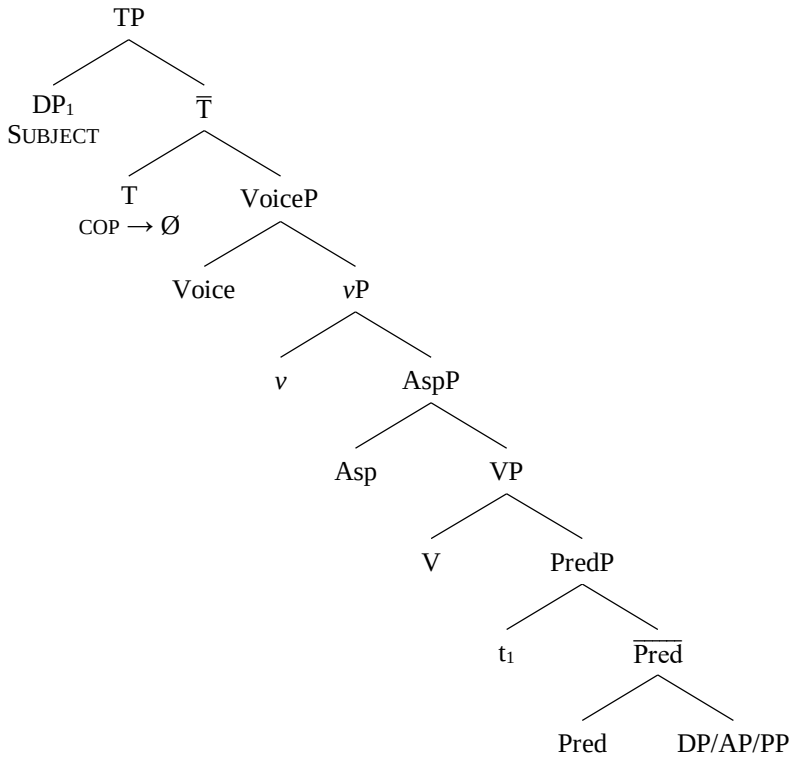


Figure 18: The Structure of Copular Clauses with Inner AspP

4.4.1 Motivation for Inner AspP

Although the notions of stage-level vs. individual-level predicates and Davidsonian vs. Kimian states are adequate to characterise overt encoding of the copula with regard to the inherent lexical aspectual properties of the predicate (especially in simplex copular clauses without any linguistic material other than the subject, the optional copula, and the predicate), there exist certain environments in which a temporally bound interpretation is not associated with the predicate itself. The covert grammatical aspect argued for in this chapter is detectable in certain environments in which neither the predicate nor the inflectional layer may be responsible for introducing a temporally bound interpretation of the clause. For instance, although the change-of-state interpretation is clearly available in the modal, control, and causative environments in (223d-f) and many more, how the clauses obtain it is not so obvious as no overt temporo-aspectual material is detected, i.e. no visible grammatical aspect. Consider examples (245):

(245) a. *Saya cuba untuk* (**akan*) *jujur dengan=nya.* (Control)
1.SG try for PROS) honest with=3
'I tried to be honest with him.'

b. *Saya buat Zul* (**akan*) *jujur.* (Causative)
1.SG make Z. PROS) honest
'I made Zul be honest.'

In spite of the individual-level predicate and the impossibility of outer grammatical aspect (i.e. aspectual markers in the inflectional layer) the examples presuppose a change of state such that there is a beginning to the state denoted by the predicate. Given that *jujur* is an individual-level predicate that denotes a permanent property, it cannot be the predicate that is responsible for the inchoative aspect. Likewise, given that outer aspect cannot be projected in these constructions, outer aspect is out of the question. Therefore, inner grammatical aspect is the remaining agent responsible for introducing the change of state in the clauses and is analysed to be an invisible factor in the calculation of the overt encoding of the copula in Malay.

Furthermore, instances in which the copula co-occurs with *jadi* without any inchoative aspect, e.g. (246), points to something else being responsible for the change-of-state interpretation. Apparently, the verb as such (which usually remains null) is not responsible for the interpretation as the reading is unavailable despite its overt realisation.[54] Therefore, *jadi* needs to combine with something else that adds the inchoative aspect to the verb, which is comparable to the analysis of inchoative verbs as the composition of a predicate and other elements that carry causative and inchoative information, e.g. *enrage* → [CAUSE [BECOME [very angry]]] (Lakoff, 1976).

(246) *Ke-selamat-an rakyat negeri ini adalah* *(*meŋ-*)*jadi ke-utama-an.*
NMZ-safe-NMZ people state PROX COP ACT- become NMZ-first-NMZ
i. 'The safety of the people of this state is a priority.'
ii. #'The safety of the people of this state is to become a priority.'
(Riduan, 2020)

[54] As indicated by the incoherence of (246ii), it is not the case that *jadi* introduces a clausal constituent that functions as the complement of the copula. It is an optional element, and the meaning of the sentence does not change without it (also see Section 3.1.2 for a discussion of the syntax of this kind of sentence.) Therefore, the DP *keutamaan* remains the predicate, rather than *jadi*, since it is meaningless and optional. However, in sentences in which *jadi* encodes an event with a change of state, then it should be identified as the verbal predicate as it is an obligatory and meaningful linking verb.

The verb gets spelt out despite its lack of aspect by the attraction of active voice marker *meŋ-*, which forces the spell-out and movement of the verb to VoiceP to serve as its host to prevent a violation of the stray affix filter, as suggested by the complete absence of the string "*adalah jadi*" where *jadi* is not affixed by *meŋ-*.[55] [56] Conversely, in the example below, the voice marker is shown to be unable to affix to the meaningful inchoative verb, which entails that the verb has not moved to VoiceP.

(247) *Nampak=nya perut saya boleh kembung (*meŋ-)jadi=nya.*
see=3 stomach 1.SG can bloated ACT- become=3
'It seems my stomach can become bloated.'
(Mohamad, 2015)

What examples (246) and (247) demonstrate is that the inchoative aspect of the verb is not inherent to the verb but is obtained elsewhere. Thus, it is reasonable to attribute the change-of-state interpretation to something other than the verb, the predicate, or outer aspect. It is therefore inner aspect that is responsible. Specifically, the Asp^0 head combines with the null verb to give it the inchoative aspect, which subsequently spells out *jadi*, as in (247). Considering that the predicate has undergone movement to a position intervening between the verb and modal *boleh* in (247), which is still within the extended VP, the inner AspP responsible for the inchoative aspect of the verb can be located below VoiceP but above VP.

[55] There being an underlying null form of *jadi* in Malay then reduces *ialah* and *adalah* to auxiliaries that can be construed as supporting elements in the copular clause. As we have seen, *ialah, adalah,* and *jadi* are not one and the same underlying copula, since *adalah* is observed to co-occur with *jadi* in certain environments. *Ialah* and *adalah* head TP, whereas the null and overt forms of *jadi* head VP. Therefore, *jadi* and its underlying null form can be construed as a copular verb, whereas *ialah* and *adalah* are copular auxiliaries.

[56] The string here refers to the copula and the verb co-occurring in the same clause. It is possible for *jadi* to be the head of a separate clause that forms the complement of the copula, which implicates a bi-clausal structure. In this case, the string "*adalah jadi*" is possible, as shown below:

(xiii) *Paling ter-baik adalah jadi diri-sendiri…*
most SUP-good COP become self-RED
'(The) best is to be yourself…'
(Fauzi, 2019)

Figure 19 illustrates head movement of the verb to Asp^0 to obtain its inchoative aspect before moving to Voice^0 to serve as the host to the voice marker. In the case of example (246), there is no inner AspP to provide the verb with inchoative aspect. Nevertheless, it gets spelt out to host the voice marker.

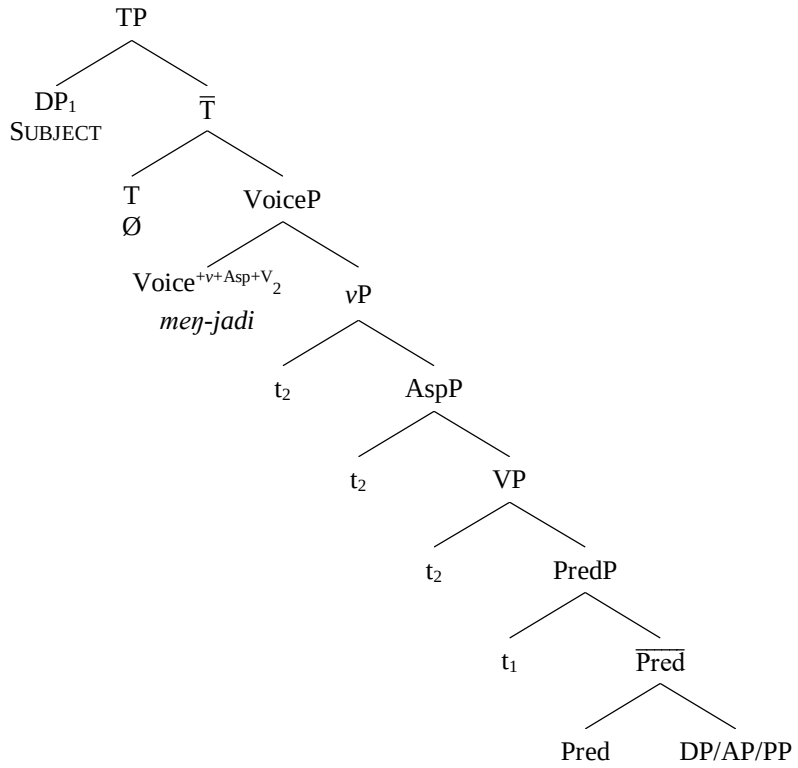


Figure 19: The Movement of 'Jadi' to Asp^0 and Voice^0

Apparently, there are various environments in which a temporally bound interpretation of the clause is available despite the absence of overt temporo-aspectual material. For example, mood is also likely to alter the aspectual properties of the clause and subsequently affect the copula, i.e. depending on the mood, a temporally bound interpretation can also be achieved. To illustrate, the moods in examples (248b-e) presuppose a change of state from dishonesty to honesty, which is not compatible with overt encoding of the copula:

(248) a. *Zul adalah jujur.* (Declarative)
Z. COP honest
'Zul is honest.'

b. (*Adalah*) *jujur* *sekarang!* (Imperative)
COP honest now
'Be honest now!'

c. *Jangan* (**adalah*) *jujur!* (Prohibitive)
PROH COP honest
'Don't be honest.'

d. *Semoga Zul* (**adalah*) *jujur.* (Optative)
OPT Z. COP honest
'May Zul be honest.'

e. *Jom kita* (**adalah*) *jujur.* (Hortative)
HORT 1.PL.INCL COP honest
'Let's be honest.'

Whether it is the mood of the clause or the aspectual properties associated with it that is at work in the examples above can be tested by slightly altering the interpretation, but not the mood of the sentence. The examples below do not force a reading that implies a change of state. Rather, they more appropriately concern the veracity or truth value of the proposition "Zul's statement is honest". As expected, overt encoding of the copula is possible, despite the examples being of a non-declarative mood. Whatever the mood is, there is no change in the aspectual property of the clause as the subject "Zul's statement" is not something whose honesty can change, in contrast to a human subject who can change how honest they are.

(249) a. *Jangan-jangan kenyataan Zul adalah jujur.* (Prohibitive)
PROH-RED statement Z. COP honest
'Possibly (doubtfully) Zul's statement is honest.'[57]

b. *Semoga kenyataan Zul adalah jujur.* (Optative)
OPT statement Z. COP honest
'May Zul's statement be honest.'

This finding should also make it possible for overt encoding of the copula not to be dictated by the mood of a clause other than its own. For example, a copular clause with an overt copula can be embedded in a non-declarative clause.

[57] The epistemic use of prohibitive *jangan* (optionally reduplicated) does not have a straightforward translation in English but it can be thought of as a phrase used to convey doubt. Statements using *jangan-jangan* lean towards the negative, as opposed to those using *entah-entah*, which can be used in the same way but is more positive on the epistemic scale.

(250) a. *Meŋ-aku sekarang bahawa dia adalah putera!* (Imperative)
ACT-admit now COMP 3.SG COP prince
'Admit now that he is a prince!'

b. *Semoga benar bahawa dia adalah putera.* (Optative)
OPT true COMP 3.SG COP prince
'May it be true that he is a prince.'

The data presented above make it clear that it is not solely the stage-level vs. individual-level distinction that govern overt encoding of the copula. Grammatical aspect is also responsible for this phenomenon, and the effect is seen on several different levels of linguistic representation, such as syntax, semantics, and pragmatics. These equivocal environments lie outside frameworks that analyse the predicate on its own as the locus of differences in aspectual information. It is obvious that the lexical aspect of the predicate does not solely govern overt encoding of the copula; grammatical aspect also has a part to play. By postulating a covert AspP projection that carries the change-of-state interpretation, grammatical aspect can be analysed to compositionally interact with lexical aspect to yield differences in overt encoding of the copula, as well as the verb *jadi*, which typically remains null in atemporal copular clauses.

4.5 A Further Problem: Verbalisation of the Predicate

Even with the analysis above, there remain some additional questions that need answering, especially when coercion is concerned. As illustrated below, when an individual-level predicate is used in an environment in which it receives a temporally bound reading, overt encoding of the copula is illicit, as in (251a), as expected; however, regardless of the overt or null encoding of the copula, the copular clause is rendered ungrammatical. In order to ameliorate the construction, inchoative verb *jadi* must be used, as in (251b).

(251) a. **Zul telah* (*adalah*) *wanita.* (Individual-Level)
Z. PRF COP woman
(Zul has become a woman.)

b. *Zul telah* *(*jadi*) *wanita.*
Z. PRF become woman
'Zul has become a woman.'

Meanwhile, stage-level predicates do not seem to behave very well, as the use of *jadi* is optional, as shown below:

(252) *Zul telah* (*jadi*) *lapar.* (Stage-Level)
Z. PRF become hungry
'Zul has become a hungry.'

Even more mysterious is the existence of an ambiguous class of predicates that assume the guise of individual-level predicates but behave like stage-level predicates when coerced. These predicates warrant overt encoding of the copula in atemporal contexts, as in (253a). Therefore, as individual-level predicates, one would expect that they pattern with the individual-level predicate in (251b) in making the use of *jadi* obligatory when such predicates occur in temporally bound contexts. On the contrary, *jadi* is optional, as shown in (253b). This spell-out pattern of *jadi* makes these ambiguous predicates unexpectedly pattern with stage-level predicates, as in (252).

(253) a. *Zul adalah jujur.* (Individual-Level)
Z. COP honest
'Zul is honest.'

b. *Zul telah* (*jadi*) *jujur.* (Coerced)
Z. PRF become honest
'Zul has been/become honest.'

The table below lists down some of these predicates. Remarkably, the degree to which the copula and the inchoative verb are obligatory increases with the permanence of the predicate. Stage-level predicates are temporary, whereas individual-level predicates are permanent. Ambiguous predicates can either denote permanent properties when used atemporally, which warrants overt encoding of the copula, or temporary states or even processes (e.g. with the use of imperfective marker *belum*) when used in temporally bound contexts, which bans overt encoding of the copula. Preceding stage-level predicates in the table should then be dynamic verbal predicates such as *main* (play) and *lari* (run), which are [–*adalah*] and [–*jadi*].

Stage-Level: [–*Adalah*] [±*Jadi*]	Ambiguous: [±*Adalah*] [±*Jadi*]	Individual-Level: [+*Adalah*] [+*Jadi*]
Pening (dizzy)	*Kotor* (dirty)	*Haram* (prohibited)
Dahaga (thirsty)	*Bersih* (clean)	*Halal* (permissible)
Takut (scared)	*Kecil* (small)	*Robotik* (robotic)
Risau (worried)	*Mahal* (expensive)	*Percuma* (free)
Sedih (sad)	*Tinggi* (tall)	*Pendek* (short)
Gembira (happy)	*Tua* (old)	*Muda* (young)

Table 23: Predicates According to Their Occurrence with the Copula and Inchoative Verb[58]

The following table summarises the spell-out patterns of the copula and the inchoative verb. The class of ambiguous predicates appears in the intersection.

Grammatical Aspect	**Lexical Aspect**	
	Stage-Level	Individual-Level
Atemporal (Inner AspP Absent)	–	COP → *adalah* V → Ø
Temporally Bound (Inner AspP Present)	COP → Ø V → (*Jadi*)	Amb. (Coercion) COP → Ø V → *(*Jadi*)

Table 24: The Spell-Out Patterns of the Copula and the Inchoative Verb

[58] Notice that *mahal* (expensive), *tinggi* (tall) and *tua* (old) are [±*adalah*, ±*jadi*], whilst their antonyms are [+*adalah*, +*jadi*]. On a temporally bound reading, becoming tall is a durative process, whereas becoming short cannot be conceived to be durative as it is not something that happens naturally, so it is construed as a sudden occurrence whose use of inchoative *jadi* is obligatory.

(xiv) a. *Zul telah* (*jadi*) *tinggi.*
Z. PRF become tall
'Zul has become tall.'

b. *Zul telah* *(*jadi*) *pendek.*
Z. PRF become short
'Zul has become short.'

I argue that the apparent optionality of *jadi* is a reflection of the syntactic category of the predicate. It is a result of a verbal-nonverbal alternation of the predicate such that the spell-out of *jadi* occurs when the predicate is nonverbal, whereas no spell-out occurs when the predicate is verbal. By *verbal*, the predicate is analysed to be verbalised via combination of the root of the predicate with a verbalising v^0 head. As such, there is essentially no optionality. The following diagram illustrates the change in category of the predicate: [59]

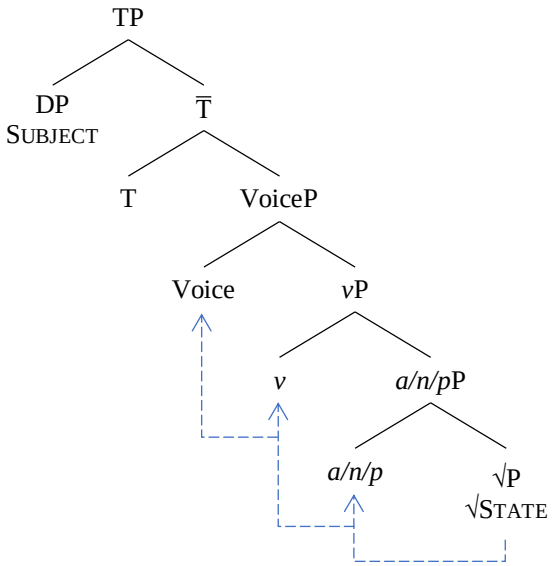


Figure 20: The Verbalisation of a Temporary State

Strictly individual-level predicates like *benar* (true) may never occur in temporally bound environments unless they combine with inchoative *jadi*, which entails that such predicates are exclusively nonverbal. Stage-level predicates like *lapar* (hungry) may or may not combine with *jadi*, depending on whether the predicate is verbal or nonverbal – nonverbal with *jadi* and verbal without *jadi*. Although ambiguous predicates such as *jujur* (honest) or *sempurna* (perfect) may or may not combine with *jadi*, similar to stage-level predicates, their lexical aspectual properties

[59] Depending on the thematic structure of the resulting verb, the subject of the clause could originate in Spec*v*P. This is the case with verbs such as *bersedih* and *berlapar* as they are agentive, as evidenced by the grammaticality of using the adverbial *dengan sengaja* (with intention).

(xv) *Mereka* { *ber-sedih* / *ber-lapar* } *dengan sengaja.*
3.PL INTR-sad INTR-hungry with intention
'They are grieving/starving (themselves) intentionally.'

differ depending on their syntactic category. For instance, they are individual-level predicates when nonverbal, as they must occur with *jadi*, whereas they are stage-level predicates when they occur as verbs, as they do not require the assistance of *jadi* despite gaining a temporally bound reading. The following table summarises the spell-out patterns according to whether the predicate is verbal or nonverbal:

	Ambiguous		
	Stage-Level / Davidsonian State		**Individual-Level / Kimian State**
	Nonverbal	Verbal	Nonverbal
Atemporal: (AspP Absent)	–	Ø	COP → *Adalah* V → Ø
Temporally Bound: (AspP Present)	COP → Ø V → *Jadi*	Ø	COP → Ø V → *Jadi*

Table 25: The Spell-Out of the Copula and the Inchoative Verb According to Category

4.6 Verbal Encoding of Temporary States

The patterns associated with the spell-out of the copula and *jadi* are a double-faceted illustration of the *Permanency Parameter* by Stassen (1997), which concerns how a property-concept word is encoded according to the permanence of the denoted state. In some languages, the permanency parameter manifests as a difference in overt encoding of the copula, as in Akawaio and Malay, whereas in others, the syntactic category of the predicate changes according to permanence, as in Māori.

In light of this and given that stage-level predicates pass as events, it is likely that stage-level predicates in Malay have the capacity to formally appear as verbs, instead of adjectives. Surely, this would explain why overt encoding of the copula is not licensed and why *jadi* is optional with stage-level predicates. Therefore, the presence of *jadi* can be analysed to reflect the syntactic category of the property-concept word as either a verb or an adjective. With *jadi*, the property-concept word is obviously a nonverbal element; without it, it is a verb, as it patterns with other verbs in neither requiring a copula nor some other verbal element to function as a predicate. The same can be said of ambiguous predicates, which require a copula and inchoative *jadi* when used as a predicative adjective in atemporal and temporally bound contexts respectively, but neither when it is used as a verb, regardless of temporality.

To gain a better understanding of the dual verbal-adjectival nature of some of these predicates, consider examples (254) and (255). In (254), *lapar* and *jujur* pattern with *makan* in not requiring *jadi*, which suggests that they are verbs.[60]

(254) a. *Saya ti-ada masa untuk *(jadi) warak.* (Individual-Level)
1.SG NEG-have time for become pious
'I have no time to be pious.'

b. *Saya ti-ada masa untuk lapar.* (Stage-Level)
1.SG NEG-have time for hungry
'I have no time to be hungry.'

c. *Saya ti-ada masa untuk jujur.* (Ambiguous)
1.SG NEG-have time for honest
'I have no time to be honest.'

d. *Saya ti-ada masa untuk makan.* (Dynamic Action)
1.SG NEG-have time for eat
'I have no time to eat.'

Conversely, *lapar* and *jujur* pattern with *warak* in (255) in requiring *jadi*, which entails that they are adjectives:

(255) a. *Zul ber-ubah *(jadi) warak.* (Individual-Level)
Z. INTR-change become pious
'Zul changed to become pious.'

b. *Zul ber-ubah *(jadi) lapar.* (Stage-Level)
Z. INTR-change become hungry
'Zul changed to become hungry.'

c. *Zul ber-ubah *(jadi) jujur.* (Ambiguous)
Z. INTR-change become honest
'Zul changed to become honest.'

d. **Zul ber-ubah (jadi) makan.* (Dynamic Action)
Z. INTR-change become eat
(Zul changed to eating.)

[60] *Jadi* is optional with *lapar* and *jujur* in these examples, which entails that they could either be encoded as verbs or adjectives in this context, but the inchoative verb is not included here to illustrate the similarity of the two predicates with the verbal predicate in a clearer manner.

Naturally, there is a change in interpretation as well, according to the occurrence of *jadi* in the clause. With *jadi*, the change of state is interpreted to be punctual, i.e. a sudden change. For example, the change to the state of hunger in example (256a) could have occurred instantaneously, e.g. triggered by seeing of a tantalising bowl of ramen, as opposed to the durative change in (256b) which could have taken however long it takes for a person to naturally feel the sensation of hunger.

(256) a. *Zul telah jadi lapar.* (Punctual)
Z. PRF become hungry
'Zul has become hungry.'

b. *Zul telah lapar.* (Durative)
Z. PRF hungry
'Zul has become hungry.'

We once again see the conditioning of a constituent based on aspect; the aspect of the nonverbal predicate (stage-level vs. individual-level) determines overt encoding of the copula, whilst the aspect of the change of state (punctual vs. durative) determines the syntactic category of the predicate. Also related is the nominalisation of temporary states to imply a prolonged state, as illustrated below.

(257) *Zul ke-lapar-an / ke-panas-an / ke-letih-an.*
Z. NMZ-hungry-NMZ NMZ-hot-NMZ NMZ-tired-NMZ
'Zul has been hungry/hot/tired (for a long time).'

This difference in permanence is reminiscent of the Time-Stability scale by Givón (1984), which grades the lexical categories – Noun, Adjective, Verb – based on their dynamicity, with verbs being the most dynamic and nouns the least. However, as seen with the example above, the notion of permanence does not have a one-to-one correspondence with syntactic category. Dynamicity transcends syntactic category – certain nominals can encode temporary states (*ke-hujan-an*; NMZ-rain-NMZ; drenched) instead of kinds, whereas certain verbs can encode permanent properties (*meŋ-risau-kan*; ACT-worry-APPL; worrying) instead of events. What is important is the semantic distinction among kinds, states, and events, which only roughly but not necessarily correspond to the syntactic categorical distinction among nouns, adjectives, and verbs. Kinds are the least dynamic as they are abstractions over individuals, which are strictly atemporal, whereas events are the most dynamic as they are temporally bound.

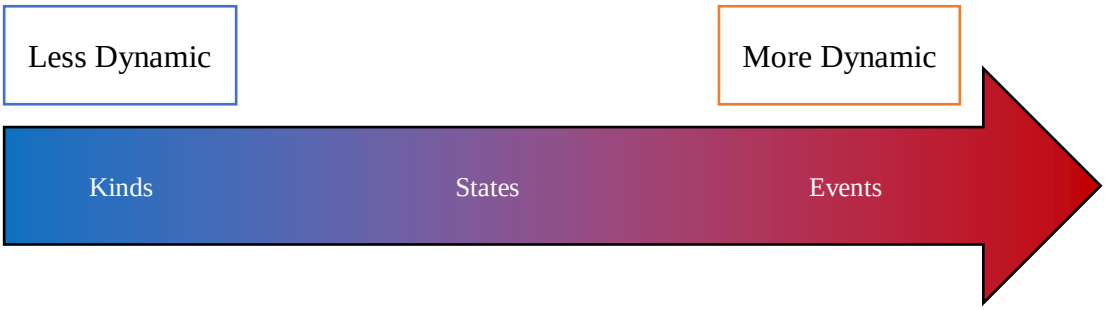


Figure 21: The Revised Time-Stability Scale Adapted from Givón (1984)

The scale above could be translated and reduced to the distinctions discussed in Section 4.2 to yield differences in overt encoding of the copula and verbal or nonverbal encoding of the predicate. Thus, in Carlson and Maienborn's terms, kinds are individual-level predicates and K-states, whereas events are stage-level predicates and D-states. Lastly, the class between kinds and events on the scale could form either stage-level or individual-level predicates and Kimian or Davidsonian states depending on their dynamicity.

The intermediate position of states on the scale, which roughly correspond to adjectives, is interesting and relevant to the discussion of verbal vs. nonverbal encoding of temporary states in Malay. Temporary states such as *lapar* (hungry), *gembira* (happy), *panas* (hot), etc. may be encoded verbally or nonverbally, with the latter option allowing the use of inchoative verb *jadi* in temporally bound contexts. This phenomenon follows the Adjective Principle by Stassen (1997), who observes that words that typically carry an adjectival role, e.g. properties and states, get encoded either as a verb in having verbal traits like tense, or as something that appears as a noun in requiring an overt copula.

> The Adjective Principle:
>
> *Predicative adjectives have no prototypical encoding strategy of their own. In their predicative encoding, adjectives will align themselves either with verbs (and, as a result, have a verbal strategy), or with nominals, or with locationals. In no language is it possible to have an encoding strategy which is used exclusively for predicative adjectives.*
>
> (Stassen, 1997, p. 30)

Stative verbs are common throughout Austronesian as an alternative to predicative adjectives, as many Austronesian languages do not have a fixed adjectival class to begin with.[61] Compare Urak Lawoi', a very closely related Malayic language, with Malay – examples (258) and (259) show that the two languages employ different encoding strategies with the same property-concept word. In Urak Lawoi', the property-concept word is encoded as a stative verbal predicate, whereas, in Malay, it is encoded as a nonverbal predicate.

Urak Lawoi' (Austronesian – Malayic)

(258) *Nya suka naq budö.*
3.SG like to stupid
'He likes to be stupid.'
(Hogan & Pattemore, 1988, p. 47)

Malay (Austronesian – Malayic)

(259) *Dia suka nak *(jadi) bodoh.*
3.SG like to become stupid
'He likes to be stupid.'

It is evident that even very closely related languages may have different encoding strategies. Farther from Malayic, Tagalog makes use of a stative marker on property-concept words in copular clauses, regardless of permanence, as shown below:

Tagalog (Austronesian – Philippine)

(260) a. *Ma-talino ang bata.* (Individual-Level Predicate)
STAT-intelligence ANG child
'The child is intelligent.'

b. *Ma-lungkot ang bata.* (Stage-Level Predicate)
STAT-sadness ANG child
'The child is sad.'

The examples above show that the predicates are derived from nouns *talino* (intelligence) and *lungkot* (sadness). This phenomenon is also very common in Malay. In fact, both the stative marker *ma-* in Tagalog and intransitive verbal prefix

[61] Adjectives are a kind of verb, according to Hogan and Pattemore (1988) for Urak Lawoi' (p. 45); Arnold (2018) for Ambel (p. 39); van den Heuvel (2006) for Biak (p. 108); Teng (2007) for Puyuma (p. 87).

ber- in Malay (whose relevance and importance shall be made clear in the following section) arguably originate from the same ancestral root, Proto-Austronesian stative prefix **ma-*: "Modern Malay *ber-* is a predicative affix compounded of the well-established Austronesian element *ma-* 'predicator'/'actor-orientation' and *r* (whatever its source and original meaning)" (Benjamin, 2009, p. 309).

Considering the Adjective Principle, it is clear that states may occur as either verbs or adjectives in Malay. Aspect is seen to have a part to play in the verbal or nonverbal encoding of the predicate, with punctual changes of states allowing nonverbal encoding of the state and durative changes of states allowing verbal encoding of the predicate.

4.6.1 Differentiating Adjectives and Verbs

Temporary states in Malay may be encoded as either adjectives or verbs. Property-concept words that denote temporary states in Malay are notoriously difficult to distinguish from verbs. Omar (1968) states "in terms of their co-occurrence with the aspect and the modal verbs, the adjectives and the verbs behave entirely in the same way" (p. 17). In fact, some words do not even show a morphological distinction between nouns, adjectives, and verbs. As shown in examples (26), repeated below, *marah* could mean anger (N), angry (A), or to scold (V):

(261) a. *Marah dia belum reda.* (Nominal)
anger 3.SG IMPRF subside
'His anger has not subsided.'

b. *Dia sangat marah.* (Adjectival)
3.SG very angry
'He is very angry.'

c. *Dia marah saya.* (Verbal)
3.SG scold 1.SG
'He scolded me.'

However, there are morphosyntactic properties that tease adjectives apart from verbs in Malay. The possibility of temporary states to be affixed by *ber-* is the strongest evidence for the verbal encoding of the predicate, as illustrated in (262). The use of the prefix implies agentivity or intentionality, presumably because it is commonly used with unergative verbs.

(262) a. *Mereka ber-sedih* (*dengan sengaja.*)[62]
3.PL INTR-sad with intention
'They are grieving (intentionally).'

b. *Mereka ber-lapar* (*dengan sengaja.*)
3.PL INTR-hungry with intention
'They are starving (intentionally).'

Either verbal or adjectival encoding of the temporary state happens to make a difference in argument structure, as suggested by the agentivity of the subject. As an agentive verb, the verbal predicate does not require the assistance of a copula or a linking verb as it is capable of predicating itself. In having an agent as an argument, the *v*P should project a specifier from which the subject originates, as opposed to the nonverbal encoding of the predicate, whose *v*P should not project a specifier, due to the unaccusative nature of copular clauses. The predicative adjective evidently requires the assistance of a linking verb, which usually remains null but surfaces as *jadi* when an aspectual factor is present.

Next, gradeability is typically a property that is associated with adjectives. Although it is possible to quantify over events, the expression *se-banyak* (one-much) must be used with a verbal predicate in a comparative construction in Malay. Conversely, *se-* can directly attach to an adjectival predicate, as illustrated below:

(263) a. *Se-kejam firaun* (IL Adjective)
one-cruel pharaoh
'As cruel as a pharaoh'

---

[62] Although the example illustrates verbal encoding of the predicate, there do exist certain predicates affixed by *ber-* that are nonverbal. These predicates are deverbal adjectives, as evidenced by the possibility of the copula to surface.

(xvi) *Semua vaksin adalah ber-kesan.*
all vaccine COP INTR-effect
'All vaccines are effective.'
(Adnan, 2021)

In fact, predicates modified by all sorts of verbal affixes may be deverbalised. All the predicates in the following example carry verbal morphology but are interpreted as properties that can be ascribed to a subject.

(xvii) *Forex adalah*{ *di-larang* / *ter-kawal* / *meŋ-untung-kan* / *ber-bahaya.* }
forex COP PASS-prohibit NVOL-control ACT-profit-APPL INTR-danger
'Forex is prohibited/controlled/profitable/dangerous.'

b. *Se-sedih kisah Puteri Gunung Ledang* (SL Adjective)
one-sad story princess G.
'As sad as the story of the Princess of Gunung Ledang'

c. *Se-ber-tuah anak raja* (IL Deverbal Adjective)
one-INTR-luck child king
'As lucky as a king's child'

d. *Se-ter-tekan hamba abdi* (SL Deverbal Adjective)
one-NVOL-press slave slave
'As stressed as a slave'

e. **Se-ber-renang ikan lumba-lumba* (Verb)
one-INTR-swim fish race-RED
'Swim as much as a dolphin'

f. *Ber-renang se-banyak ikan lumba-lumba* (Verb)
INTR-swim one-much fish race-RED
'Swim as much as a dolphin'

Considering that the expression *sebanyak* is used with verbs, it may also be used with *lapar* and *sakit*, attesting to their dual verbal-adjectival nature, as shown below:

(264) a. *Kita lapar se-banyak sepuluh kali se-hari.*
1.PL.INCL hungry one-much ten time one-day
'We get hungry as many as ten times a day.'

b. *Aku sakit se-banyak dua kali se-minggu.*
1.SG ill one-much two time one-week
'I get ill as much as twice a week.'

Recall from Section 4.2.1 that stage-level predicates in Malay, such as *sakit* (ill), pass diagnostics for events. The eventiveness of stage-level predicates is perhaps the reason why they may be affixed by *ber-*, which makes them verbs in form. Dynamic verbs may not be modified by *sangat*, which is a degree adverb used with gradable adjectives, as they clearly are not gradable, whilst individual-level adjectival predicates may not be affixed by *ber-* as they are not eventive, as illustrated in (265). Therefore, verbs that denote actions do not overlap with adjectives that denote permanent properties, unlike temporary states, which overlap with both classes.

(265) a. **Mereka sangat ber-lari.* (Action Verb + Degree Adverb)[63]
3.PL very INTR-run
(They run a lot.)

b. **Mereka sangat ber-gaduh.*
3.PL very INTR-fight
(They fight a lot.)

c. **Mereka ber-bijak.* (Permanent Property + *Ber-*)
3.PL INTR-intelligent
(They are being intelligent.)

d. **Mereka ber-sempurna.*
3.PL INTR-perfect
(They are being perfect.)

The observations so far point towards stage-level predicates being able to form an adjectival class as well as a verbal class. However, both stage-level and individual-level predicates behave similarly when used as a postmodifier of a noun. They pattern with each other to the exclusion of dynamic verbs, which need to form relative clauses in order to function as a postmodifier of a noun. Therefore, stage-level predicates can be said to be more adjective-like.

(266) a. *Orang sakit itu jatuh.* (Stage-Level Predicate)
person ill DIST fall
'That ill person fell.'

b. *Orang gembira itu jatuh.*
person happy DIST fall
'That happy person fell.'

c. *Orang bijak itu jatuh.* (Individual-Level Predicate)
person intelligent DIST fall
'That intelligent person fell.'

d. *Orang ber-kira itu jatuh.*
person INTR-count DIST fall
'That calculative person fell.'

[63] The appropriate degree adverb to quantify events is *banyak* (much).

(xviii) *Mereka banyak ber-lari.*
3.PL much INTR-run
'They run a lot.'

e. *Orang* *(*yang*) *ber-bual itu jatuh.* (Dynamic Action)
person REL INTR-chat DIST fall
'That person who is chatting fell.'

f. *Orang* *(*yang*) *ber-jalan itu jatuh.*
person REL INTR-walk DIST fall
'That person who is walking fell.'

Even though stage-level predicates might possess some verbal traits, they seem to have a closer affinity to the adjectival class. Their more adjectival nature indicates that they start out as adjectives but may be derived into verbs. In fact, all adjectives in Malay may be used as transitive verbs, provided that they are affixed by an appropriate applicative suffix, as shown in (267).

(267) a. *Dia risau-kan saya.* (Stage-Level → Verb)
3.SG worried-APPL 1.SG
'He worries me.'

b. *Dia gembira-kan saya.*
3.SG happy-APPL 1.SG
'He makes me happy.'

c. *Dia haram-kan arak.* (Individual-Level → Verb)
3.SG prohibited-APPL alcohol
'He banned alcohol.'

d. *Dia digital-kan jam itu.*
3.SG digital-APPL watch DIST
'He digitised the watch.'

Thus, a verbal stage-level predicate like *bersedih* is analysed to be initially merged as a root without a syntactic category, as illustrated in Figure 20. It gets derived into an adjective (as is apparent from its adjective-like gradeability) by the adjectival categoriser a^0 before getting derived into a verb by v^0. From there, it moves higher up into VoiceP to obtain the verbal *ber-* prefix, which is argued by Bril (2005) to be a multi-functional prefix that can express durative verbs, states and properties, inter alia. *Ber-* is also often referred to as the middle voice prefix, e.g. in Benjamin (2009), so it is assumed to be the head of VoiceP.

Ultimately, there is no actual optionality of *jadi*, e.g. (268a). The apparent optionality of *jadi* is attributed to the structural ambiguity of the clause, especially since the *ber-* prefix is also optional. With the prefix, the clause is obviously verbal, and it gets the agentive interpretation. However, without it, the sentence deceivingly looks like a copular clause. Due to the optionality of *ber-*, the sentence could take the guise of a copular clause with a zero copula when, in fact, the temporary state is a verb that need not be accompanied by a copula or *jadi*, be it overt or zero, as in (268c). Meanwhile, in the actual nonverbal use, inchoative verb *jadi* always gets spelt out due to the presence of the inner AspP, as is apparent in all temporally bound copular clauses with individual-level predicates, as in (268b). Lastly, combining the verbal stage-level predicate with *jadi* results in ungrammaticality due to the co-occurrence of two verbs in a single clause, as in (268d).

(268) a. *Dia* (*jadi*) *sedih.* (Apparent Optionality)
3.SG become sad
'She became sad.'

b. *Dia jadi sedih.* (Adjectival)
3.SG become sad
'She became sad.'

c. *Dia* (*ber-*) *sedih.* (Verbal)
3.SG INTR-sad
'She is sad/grieving.'

d. **Dia jadi ber-sedih.*
3.SG become INTR-sad
(She became sad.)

This analysis has the power to explain predicates in Malay that originate as nominal or prepositional elements but are later derived into verbs via verbalisation and subsequent affixation of the *ber-* prefix, as illustrated below.

(269) a. *Kucing itu ber-anak.* (Denominal Verb)
cat DIST INTR-child
'That cat is giving birth.'

b. *Petani itu ber-panas.* (Deadjectival Verb)
farmer DIST INTR-hot
'That farmer is scorching (in the sun).'

c. *Lubang itu ber-dalam-dalam.* (Deprepositional Verb)
hole DIST INTR-in-RED
'That hole is deepening.'

4.7 Summary

To summarise, the copulas have been found to be sensitive towards the notion of aspect, whether it be grammatical aspect or lexical aspect. It is apparent that environments that allow overt encoding of the copulas are those that are atemporal. The preference of the copulas for atemporal environments goes beyond the level of the syntax, as the semantics of the predicate itself and change-of-state presuppositions via coercion are also responsible for the ban on overt encoding of the copula. The aspectual factors that bring about change of state translates into eventuality – stage-level predicates are eventive, whereas individual-level predicates are stative. Subsequently, the eventuality of the clause makes overt encoding of the copula impossible.

The Interpretation of the change of state is reflected in the syntax such that it is carried by an AspP that is located within the layers of the *v*/VP, as in Travis (2010). The presence of this *v*/VP-internal AspP is substantiated by the availability of the change-of-state interpretation in clauses that cannot accommodate grammatical aspect, such as non-finite clauses. Besides, this aspectual factor is observed to interact with the verb to allow inchoative verb *jadi* to surface, which usually remains null when stative.

Building on the observations made pertaining to aspect and overt encoding of the copula is the finding that predication in Malay obeys the Adjective Principle by Stassen (1997). States are likely encoded as nonverbal predicates, whereas events are likely encoded as verbal predicates. However, the syntactic category of the predicate does not always have a one-to-one correspondence with what Givón (1984) calls Time-Stability, as there do exist nouns that correspond to stage-level predicates and verbs that correspond to individual-level predicates, which affects overt encoding of the copula. Due to the hybrid status of stage-level predicates between more dynamic events and less dynamic states, they can either be encoded as verbs or adjectives. However, they are found to be more adjective-like in form, which is analysed to be an

indication that they are formally adjectives. When verbal, stage-level predicates are initially merged as adjectives but get derived into verbs via categoriser v^0 and finally moves to VoiceP to obtain the middle voice marker *ber-*. In their nonverbal use, they do not license overt encoding of the copula but must combine with inchoative *jadi*, as the presence of the *v*/VP-internal AspP forces the linking verb to surface.

This chapter has provided an exposition of those various environments and generalised the pertinent data to identify the conditions that are conducive to overt encoding of the copulas. It has elaborated on the conditions that govern the choice of copula in copular clauses in Malay. It can be concluded that the choice between a null or zero copula is not arbitrary but is principled. The misconception that the Malay copulas are largely optional is disproven as, in many cases, the use of an overt copula is ungrammatical, due to strict semantic conditioning.

Chapter 5: Cleft Constructions

Cleft constructions in Malay are poorly understood despite their use in some common syntactic operations such as the formation of interrogatives. As illustrated below, the structure of a cleft is identical to that of a *wh*-question, which itself involves clefting.[64]

(270) a. *Apa yang Ali minum?* (Interrogative Cleft)
what COMP A. drink
'What was it that Ali drank?'

b. *Air yang Ali minum.* (Declarative Cleft)
water COMP A. drink
'It was water that Ali drank.'

Prima facie, clefts and pseudoclefts in Malay appear to be one and the same construction, but with different constituent orders, especially when one omits the optional copula. As is the general case, one construction is roughly a paraphrase of the other, meaning that there is not much difference between the two constructions in terms of semantics and information structure.

(271) a. [$_{\text{Focus}}$ *Ali*] [$_{\text{Cleft Clause}}$ *yang sedang meŋ-nyanyi.*] (Cleft)
A. COMP PROG ACT-sing
'It is ALI who is singing.'

b. [$_{\text{Cleft Clause}}$ *Yang sedang meŋ-nyanyi*] ~~*ialah*~~ [$_{\text{Focus}}$ *Ali.*] (Pseudocleft)
COMP PROG ACT-sing COP A.
'Who is singing is ALI.'

This chapter explores the syntax of cleft constructions in Malay and argues for the derivation of pseudoclefts from clefts. To begin, clefts are derived via focus movement of a DP from a root clause to the left periphery, past the C^0 head *yang* that forms the cleft clause. Novel data pertaining to the discovery of a covert matrix copular clause above the visible part of the cleft makes the case for a structure that is reminiscent of the *it*-cleft in English. This additional structure forms the copular portion of the pseudocleft to which the cleft clause moves to derive a pseudocleft.

[64] Although clefted and non-clefted questions (e.g. *wh*-in-situ) may have different interpretations, scholars usually provide a non-clefted translation for both. For example, in (270a), the clefted question may be given a normal *wh*-question translation such as "what did Ali drink?".

Upon further inspection of pseudoclefts in Malay, 2 varieties of pseudoclefts are identified: pseudoclefts whose subject is a DP and pseudoclefts whose subject is a CP (I shall henceforth call the former a canonical copular clause and the latter a pseudocleft). Scholars such as Kader (1981) and Cole and Hermon (2000) argue that clefts are derived from canonical copular clauses; however, I argue that a cleft may not be derived as such on the grounds that the cleft clause of a cleft shows non-DP properties. Rather, the reverse is true such that a pseudocleft is derived from a cleft. Furthermore, it is the pseudocleft with a CP subject that is derived from a cleft, owing to the cleft clause of both constructions being of the same syntactic category.

5.1 On Cleft Constructions in Malay

To date, the analysis of cleft constructions in Malay remains an open question due to the complexity of their syntactic structure and their branching into the formation of other different focus-related constructions that are associated with extraction in general. The structure of a cleft in Malay involves the separation of a focus from the rest of the clause by *yang*, which introduces the cleft clause, as shown below:

Context: *Zul fikir bahawa Ali minum kopi.* (Zul thought that Ali drank coffee.)

(272) *Tetapi sebenarnya* [$_{\text{Focus}}$ *jus*] [$_{\text{Presupposition}}$ *yang Ali minum.*]
but actually juice COMP A. drink
'But actually, it was JUICE that Ali drank.'

Compare clefting with topicalisation below, which does not spell out *yang*:

Context: *Ali mengidam jus.* (Ali was craving juice.)

(273) *Oleh itu,* [$_{\text{Topic}}$ *jus*] [$_{\text{Comment}}$ *Ali minum.*]
by DIST juice A. drink
'Therefore, juice Ali drank.'

There has not been unanimous agreement among Austronesianists with regard to the syntactic structure of cleft constructions in Malay, despite the breadth and depth of research done in the field of extraction and the formation of interrogatives in Austronesian languages in general, whose application involves clefting. The inclusion of interrogatives in this chapter is essential as clefting in Malay is a basic operation that usually applies in the formation of interrogatives. Mentioning this is important as clefts in Malay have only been studied within the context of the derivation

of interrogatives. Therefore, this chapter heavily relies on the conclusions drawn from research on interrogative clefts by previous scholars, which shall be extrapolated to the analysis of declarative clefts. Nonetheless, new findings are presented here which concern the structure of cleft constructions in both the interrogative and declarative.

As constructions that are intimately associated with focus and *wh*-phenomena, cleft constructions are widely assumed to involve movement of the focus to the left periphery. This is the position taken by scholars such as Aldridge (2007) and Fortin (2007), who assume that the *wh*-phrase moves past the C^0 head *yang*. As illustrated in Figure 22, *wh*-movement is a simple operation that targets SpecCP. This analysis is nothing out of the ordinary since questions in most languages with *wh*-ex-situ are derived likewise. It is only due to the use of *yang* in extraction that interrogatives in Malay form apparent cleft constructions, which is also quite a common strategy in the languages of the world, especially among the Austronesian languages.

(274) *Siapa yang meŋ-beli buku=nya?*
Who COMP ACT-buy book=3
'Who bought the book?'
(Aldridge, 2007, p. 1448)

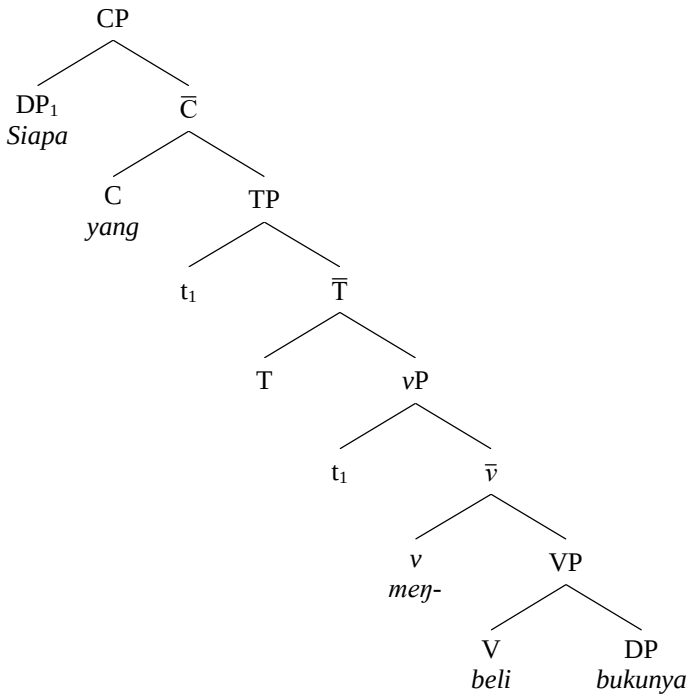


Figure 22: The Structure of an Interrogative Cleft According to Aldridge (2007)

Intuitively, one would be inclined to construe clefts and pseudoclefts to be related in how their derivation proceeds, since both of them have so many similarities, as considered by scholars such as Akmajian (1970b), Percus (1997), etc. One observation that suggests that the derivation of the two constructions in Malay is closely related is that it is not possible to pseudocleft a constituent if said constituent cannot be clefted, i.e. whatever cannot be clefted also cannot be pseudoclefted. For instance, focus movement of a non-DP argument from a clause to its left periphery does not derive a cleft construction. In examples (275), both clefting and pseudoclefting have failed because *yang* cannot be used when a non-DP is focalised. If the two constructions were not related in terms of their derivation, pseudoclefting should not be affected by the restrictions that govern clefting — primarily the use of *yang* in focusing DPs only. Otherwise, there should exist no restriction on pseudoclefting a non-DP constituent, which is apparently impossible.

(275) a. * *Semalam yang Ali di-rompak.* (Illicit Non-DP Cleft)
yesterday COMP A. PASS-rob
(It was YESTERDAY that Ali was robbed.)

b. * *Yang Ali di-rompak ialah semalam.* (Illicit Non-DP Pseudocleft)[65]
COMP A. PASS-rob COP yesterday
(When Ali was robbed was YESTERDAY.)

Indeed, scholars such as Kader (1981) and Cole and Hermon (2000) argue that clefts in Malay are derived from what appears like a pseudocleft, i.e. a cleft construction whose focus is separated from the cleft clause by a copula, as shown below. However, the structure and derivation according to these scholars is radically different from the those of scholars such as Aldridge (2007) and Fortin (2007).

[65] Substituting the illicit *yang*-clause subject of the pseudocleft with a free relative clause remedies the ungrammaticality because forming a free relative clause with an adjunct *wh*-phrase head does not involve clefting or the spell-out of *yang*.

(xix) *Bila Ali di-rompak ialah semalam.*
when A. PASS-rob COP yesterday
'When Ali was robbed was yesterday.'

(276) [Cleft Clause *Yang Ali minum*] *ialah* [Focus *jus.*]
COMP A. drink COP juice
'What Ali drank was JUICE.'

Kader (1981), one of the first scholars to have analysed interrogatives in Malay, proposes that a cleft is derived from a canonical copular construction whose subject is a relative clause adjoined to a null dummy noun. The cleft clause and the *wh*-phrase to be clefted are mediated by a null underlying copula – *ia* or *ada*. As shown below, the focus originates as the complement of the copula and moves to the front of the clause to derive a cleft:

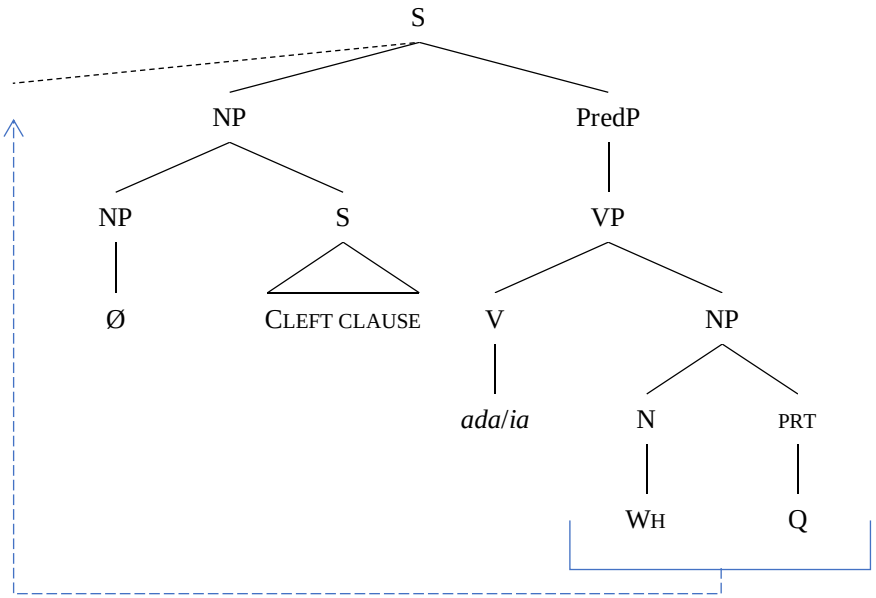


Figure 23: The Structure of a Cleft According to Kader (1981)

This analysis is also shared by Cole and Hermon (2000), who also claim that the cleft clause is a headless relative clause and that clefted questions are base-generated as canonical copular clauses. Similarly, the focus is analysed to move from the complement position of the copula to the left periphery. In either structure, a declarative counterpart can be formed by replacing the *wh*-phrase with a focused DP.

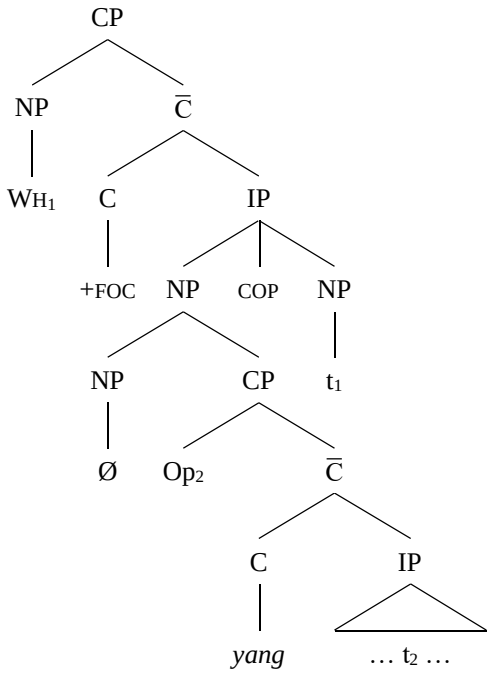


Figure 24: The Structure of a Cleft According to Cole and Hermon (2000)

The two analyses are very different and the structures they derive should reflect different properties, behaviours, and interpretations, especially with respect to whatever the constituent headed by *yang* is – part of a complex DP or a root clause.

5.1.1 The Structure of the Cleft Clause

I analyse the cleft clause to be part of the root CP, akin to the structure by Aldridge (2007) and Fortin (2007). First, the focus is moved to the front of the clause. Using the COMP-trace effect, movement is confirmed to have occurred as a null complementiser must be used when a subject has successive-cyclically raised from a complement clause into a matrix clause, as illustrated below:

(277) a. *Zul*$_1$ *dia kata* (**bahawa*) t$_1$ *makan nasi itu.* (Subject Topicalisation)
Z. 3.SG say COMP eat rice DIST
'Zul, he said, ate the rice.'

b. *Siapa*$_1$ *yang dia kata* (**bahawa*) t$_1$ *makan nasi itu?* (Subject *Wh*-Ex-Situ)
who COMP 3.SG say COMP eat rice DIST
'Who was it that he said ate the rice?'

c. *Zul*$_1$ *yang dia kata* (**bahawa*) t$_1$ *makan nasi itu.* (Subject Focus Ex-Situ)
Z. COMP 3.SG say COMP eat rice DIST
'It was Zul that he said ate the rice.'

Yang is then spelt out, which only occurs when *wh*-ex-situ is involved, which forms a clefted question, as illustrated in Section 1.5.14. It does not surface when the *wh*-phrase does not move. Naturally, the same pattern is also observed in instances of focus movement with focus marker *lah*:

(278) a. *Saya kata dia makan nasi-lah.* (Focus In-Situ)[66]
1.SG say 3.SG eat rice-FOC
'I said he ate RICE.'

b. *Saya kata nasi*$_1$ *-lah yang dia makan* t_1. (Partial Focus Ex-Situ)
1.SG say rice -FOC COMP 3.SG eat
'I said it was RICE that he ate.'

c. *Nasi*$_1$ *-lah yang saya kata dia makan* t_1. (Focus Ex-Situ)
rice -FOC COMP 1.SG say 3.SG eat
'RICE, I said it was that he ate.'

Spell-out of *yang* can be seen as a reflex of focus movement, comparable to the *wh*-agreement marker in Chamorro (Chung, 1982). Its use is not licensed when the focus remains in-situ, as in (278a). Also, it is not required in topicalisation, as in (277a), meaning that not all forms of Ā-movement license *yang*. Therefore, the co-occurrence of *yang* with focus marker *lah* or interrogative marker *kah*, all three of which are heads in the CP layer of a clause, necessitates multiple CP projections.

Due to its inherent focal property and its correspondence to focus *lah*, interrogative particle *kah* can be assumed to head FocP. That *lah* and *kah* head FocP is supported by the distribution in and the discussion around examples (41), repeated below as (279). In cases in which an auxiliary optionally remains in TP in a polar interrogative, *kah* is left without a host. Therefore, *ada* is merged in Foc^0 to support the stray affix, as in (279c). The grammaticality of example (279c) indicates that *ada* is merged directly in CP, as opposed to having moved from TP, which would violate the head movement constraint as it would need to skip *mesti*.

[66] It seems that focus marker *lah* may attach to the focus in-situ, unlike interrogative marker *kah*. At first blush, *kah* and *lah* appear to have different syntactic restrictions such that *kah* must be local to the *wh*-phrase. However, it is assumed here that movement is obligatory with both *lah* and *kah*. In the case of the apparent focus in-situ with *lah*, the whole TP has undergone movement to the left periphery in such instances and the focus marker only associates with the constituent intended to be focused, in line with analyses of *only* in which the adverb takes scope over the whole clause but only focuses the associated constituent (Jackendoff, 1972; McCawley, 1996; Rooth, 1985).

(279) a. *Ali mesti ada baca novel.*
A. must AUX read novel
'Ali must have read a novel.'

b. *Mesti*$_1$ *-kah Ali* t$_1$ *ada baca novel?*
must -Q A. AUX read novel
'Must Ali have read a novel?'

c. *Ada-kah Ali mesti baca novel?*
AUX-Q A. must read novel
'Must Ali read a novel?'
#'Must Ali have read a novel?'

The view that *kah* is merged in the CP layer goes against some scholars who analyse it to directly attach to the *wh*-phrase where the *wh*-phrase is merged, e.g. Kader (1981), which makes incorrect predictions with regard to *wh*-in-situ in Malay. If *kah* could directly attach to a *wh*-phrase, it should also be able to directly attach to an auxiliary in a polar interrogative. However, as we have seen in (279c), *kah* is merged in the left periphery. If it directly attached to its host, *mesti* would necessarily have to move to the front of the clause as the only auxiliary available because it would preclude the merging of *ada* as a support morpheme for the stray *kah* affix. Furthermore, we would expect the perfective reading of *ada* to be possible in (279c), as there would be no need to merge a semantically vacuous variant of *ada* in CP for the same reason.[67]

[67] We know that the *kah* morpheme in this case is the one that occupies CP, and not the interrogative disjunctive marker in an &P. Depending on its pronunciation (as either [kah] or [kə]), it gets a perfective reading. The [kah] variant gets a habitual reading, which entails that it is merged in CP as a support morpheme, whereas the [kə] variant gets a perfective reading, which entails that it is merged in TP. Besides, it is only with the [kə] variant that the null disjunct immediately following the morpheme can be spelt out. With the [kah] variant, the disjunct occurs at the end of the clause and declarative disjunctive marker *atau* is required, since the interrogative [kah] morpheme in CP takes scope over the whole question.

(xx) a. *Ada-kah* (**tidak*) *Ali baca novel* (*atau tidak*)? (Habitual)
AUX-Q NEG A. read novel DISJ NEG
'Does Ali read novels (or not)?'

b. *Ada-kə* (*tidak*) *Ali baca novel* (**atau tidak*)? (Perfective)
AUX-DISJ.Q NEG A. read novel DISJ NEG
'Did Ali read a novel (or not)?'

Consider the following example. *Kah* is affixed to *ada*, which is higher than the moved epistemic modal, which entails that there is another projection below the one that is headed by *kah*. *Mesti* is analysed to have moved to and landed in a lower CP projection, e.g. FinP, whereas *ada* is merged in Foc^0 to support the hostless *kah.*

(280) *Ada-kah mesti*$_1$ *Ali* t$_1$ *baca novel?*
AUX-Q must A. read novel
'Must Ali read a novel?'

As part of the extended CP, *kah* occurs between different heads that are associated with the CP layer, such as the complementiser, the topic marker, and *yang,* as shown below:[68]

(281) *Dia kata bahawa surat ini kan, Zul -lah yang bagi kepada Ali.*
3.SG say COMP letter PROX TOP Z. -FOC COMP give to A.
'He said that, this letter, it was Zul who gave (it) to Ali.'

Treating these markers as heads of their own projections is supported by Fortin (2007), who argues that the structure of content interrogatives with a *wh*-subject in Indonesian is as shown below:

[$_{FocP}$ SUBJECT [$_{Foc}$ *kah* [$_{CP}$ [$_{C}$ *yang* [$_{TP}$ ~~SUBJECT~~ [$_{vP}$ ~~SUBJECT~~ [$_{v}$ *v*+V [$_{VP}$ [$_{V}$ OBJECT]]]]]]]]

Based on the clausal structure above, the Foc^0 probe searches for an appropriate focus-carrying goal. Therefore, to check the uninterpretable [*u*Foc] feature on Foc^0, it agrees with the focus and triggers it to move to SpecFocP. The findings so far allow us to generate a structure such as the one below:

[68] Declarative *lah* is used here because complementiser *bahawa* only occurs in the declarative, as shown below. Since *kah* and *lah* are different moods of the same head, they can be assumed to occur in the same position.

(xxi) *Dia kata* (**bahawa*) *siapa-kah yang bagi surat ini kepada Ali?*
3.SG say COMP who-Q COMP give letter PROX to A.
'Who did he say gave this letter to Ali?'

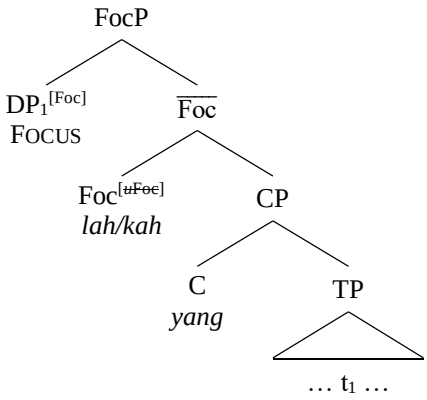


Figure 25: The Structure of the Cleft Clause in Malay

5.1.2 The Malay Cleft as a Copular Construction

Although the cleft in Malay does not resemble a copular construction, there is reason to believe that it is. Specifically, I argue that clefts in Malay are bi-clausal with the visible part of the cleft being embedded under a phonologically null copular clause. Therefore, the syntactic structure of a cleft in Malay should roughly resemble the *it*-cleft in English, as analysed by scholars such as Heggie (1989), Svenonius (1998), Hedberg (2000), Kiss (1998), etc. (notwithstanding the internal structure of the cleft clause):

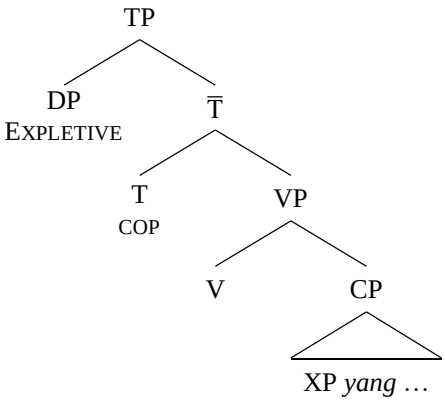


Figure 26: The Structure of the Matrix Copular Clause of Clefts in Malay

Due to the single visible predicate within the cleft clause, clefts in Malay appear to be mono-clausal. However, as should be clear by now, copulas are elements that often do not have any phonological realisation in Malay. Paired with the fact that expletives in Malay are also null, as argued by Mustaffa (2020), it can be quite tricky to determine whether there exists a matrix copular clause in a cleft construction in

Malay. Nevertheless, by using the relevant conditions on overt encoding of the copula, it can be demonstrated that clefts in Malay are also copular constructions. For example, the use of the modal *boleh* makes it obligatory for inchoative *jadi* (become) to be used in a copular clause, as described in Section 4.2. Although a clefted DP is usually a referential expression, as opposed to a predicate, the surfacing of *jadi* nonetheless applies, as clearly seen in the following examples:

(282) a. *Boleh jadi perkara sukar itu-lah yang meŋ-datang-kan ke-puas-an.*
can become matter tough DIST-FOC COMP ACT-come-APPL NMZ-satisfied-NMZ
'It could be THE TOUGH MATTER that brings satisfaction.'
(Shahimi, 2018)

b. *Boleh jadi ibu sendiri yang leka.*
can become mother self COMP negligent
'It could be THE MOTHER HERSELF who was negligent.'
(Shahimi, 2019)

Given that there is a matrix clause, it follows that the subject position should be occupied, since Malay obeys the EPP, as shown in Section 1.5.2. Preceding the matrix verb should be a null expletive and a null copula. In fact, the notion of a null expletive in a cleft is quite common. There are other Austronesian languages in which an expletive can be made visible in certain cases. For example, in clefts in Marshallese, an expletive surfaces when the cleft is negated, as shown below:

Marshallese (Austronesian – Micronesian)

(283) a. *Leddik eo en e=ar koot-e leo jer-a.*
girl DEF.SG that 3.SG=PST steal-OBJ man friend-1.SG
'It is the girl who stole my boyfriend.'

b. *E=j jab leddik eo en e=ar koot-e leo jer-a.*
3.SG=PRS NEG girl DEF.SG that 3.SG=PST steal-OBJ man friend-1.SG
'It is not the girl who stole my boyfriend.'
(Willson-Sturman, 2014, p. 10)

In Malay, there are also special cases in which an overt expletive becomes observable. The examples below illustrate that, when an extraposition construction is derived into a polar interrogative, the null expletive is spelt out.

(284) a. *Ada-kah ia ke-perlu-an untuk kita keluar*?
AUX-Q 3.SG.NH NMZ-need-NMZ for 1.PL.INCL exit
'Is it a necessity for us to exit?'
(Abd Mutalib & Halid, 2020)

b. *Bukan-kah ia prosedur am untuk tidak ber-kongsi butiran?*
CNTR-Q 3.SG.NH procedure general for NEG INTR-share detail
'Isn't it a general procedure not to share details?'
(Ngah, 2018)

Although scarce in newspapers, clefts with overt expletives are possible to find when subjected to the same condition. The example below shows the use of an overt expletive in the matrix clause of a cleft when derived into a polar interrogative.

(285) *Ada-kah ia Sarawak Report yang banyak meŋ-laku-kan serang-an…?*
AUX-Q 3.SG.NH S. COMP much ACT-do-APPL attack-NMZ
'Was it SARAWAK REPORT who committed attacks a lot…?'
("Najib, jangan lupa saman Tun M," 2015)

Prior to this, the silent matrix copular clause above the cleft in Malay had gone undetected. Based on the presence of a null expletive and a zero-copula preceding the visible part of the cleft, the structure of clefts in Malay corresponds to Figure 26. This revelation goes against all other analyses of clefts and *wh*-movement in general in Malay.[69]

Further evidence of the matrix copular clause can be found when negation and focus markers come into view. Consider the following negated copular clauses and assume that the head of the DP subject has undergone NP-ellipsis:

(286) a. ~~*Orang*~~ *yang kirim surat itu bukan Ali.*
person REL send letter DIST CNTR A.
'(The person) who sent that letter is not Ali.

b. **Ali* ~~*orang*~~ *yang kirim surat itu bukan* t_1.
A. person REL send letter DIST CNTR
(Ali, [the person] who sent the letter is not.)

[69] Cheng (1991) does suggest that interrogatives in Indonesian are *reduced clefts* in which the cleft clause is a sort of relative clause that modifies the focus, which is base-generated in SpecCP, but without the matrix copular clause and an expletive.

If the focus were truly merged in a position following the cleft clause, its movement to CP would strand the contrastive negator, as in (286b). However, stranding the negator is ungrammatical, which is an unexpected result if one subscribes to the analyses by Kader (1981) and Cole and Hermon (2000). Rather, the contrastive negator should occur preceding the whole cleft construction, as shown below:

(287) *Bukan Ali yang kirim surat itu.*
CNTR A. COMP send letter DIST
'It was not ALI who sent that letter.'

Even if the negator were analysed to have moved to the left periphery, the subsequent movement of the focus would be expected to target a position preceding the negator. The same can be said of other auxiliaries. On the contrary, the negator and other auxiliaries occur preceding the clefted XP, as illustrated below:

(288) *Mungkin tetap bukan Ali yang kirim surat itu.*
might CONT CNTR A. COMP send letter DIST
'Maybe it still was not ALI who sent the letter.'

These word order facts are unexpected if a cleft truly were derived via movement of the complement of the copula to the left periphery. With the discovery of the matrix copular clause, it is clear how the constituent order in example (288) can be derived. It is the matrix copular clause that hosts the inflectional elements and the contrastive negator, not the CP layer of the visible part of the cleft, as shown in (289). If it were the latter, it would not be possible to explain how they could have moved across the clefted XP and *yang* heading the CP.

(289) [$_{TP}$ *Ø Mungkin tetap bukan* [$_{CP}$ *Ali yang kirim surat itu.*]]
EXPL might CONT CNTR A. COMP send letter DIST
'Maybe it still was not ALI who sent the letter.'

In addition to that, it is ungrammatical for a single clause to have more than one focus marker. Even more ungrammatical is a single clause that has multiple focus markers of different moods, i.e. declarative *lah* and interrogative *kah* (or interrogative disjunctive marker *kah*). For instance, example (290a) is severely ill-formed due to two violations: multiple focus markers in a single clause; focus markers of mismatching moods. In contrast, examples (290b-e) are grammatical as each clause only includes one focus marker of either the declarative or interrogative mood.

(290) a. *_Mesti-kah surat ini di-baca oleh kamu-lah?_
must-Q letter PROX PASS-read by 2.SG-FOC
(Must this letter be read by you?)

b. _Mesti-kah surat ini di-baca oleh kamu?_
must-Q letter PROX PASS-read by 2.SG
'Must this letter be read by you?'

c. _Surat ini mesti di-baca oleh kamu-kah_ (_dia_)?
letter PROX must PASS-read by 2.SG-DISJ.Q 3.SG
'Must this letter be read by you (or her)?'

d. _Mesti-lah surat ini di-baca oleh kamu._
must-FOC letter PROX PASS-read by 2.SG
'This letter MUST be read by you.'

e. _Surat ini mesti di-baca oleh kamu-lah._
letter PROX must PASS-read by 2.SG-FOC
'This letter MUST be read by you.'

Separate clauses may each contain one focus marker of different moods, as shown below:

(291) _Ada-kah dia ber-fikir bahawa Zul-lah yang kirim surat itu?_
AUX-Q 3.SG INTR-think COMP Z.-FOC COMP send letter DIST
'Does he think that it was ZUL who sent that letter?'

Following from this observation, it is peculiar why the example below is grammatical despite the multiple focus markers of different moods in the left periphery of the root copular clause, according to the structures proposed by Kader (1981) and Cole and Hermon (2000).

(292) _Bukan -kah Zul -lah yang kirim surat itu?_
CNTR -Q Z. -FOC REL send letter DIST
'Isn't it ZUL who sent that letter?'

If a cleft is analysed to have the structures by Kader (1981) and Cole and Hermon (2000), the two constituents affixed by the *kah* and *lah* should occupy the left periphery of the same clause, as illustrated below, which cannot possibly be correct:

(293) *[$_{CP}$ _Bukan$_1$-kah Zul$_2$ -lah_ [$_{TP}$ ~~_orang_~~ _yang kirim surat itu_ t_1 t_2?]]
CNTR -Q Z. -FOC person REL send letter DIST
(Isn't it ZUL who sent that letter?)

The existence of the matrix copular clause has the power to explain how it is possible for a cleft to have multiple focus markers of different moods. Each focus marker legitimately occupies a different clause, as shown below:

(294) [$_{CP}$ *Bukan*$_1$*-kah* [$_{TP}$ Ø t$_1$ [$_{CP}$ *Zul -lah yang kirim surat itu*?]]]
CNTR -Q EXPL Z. -FOC COMP send letter DIST
'Isn't it ZUL who sent that letter?'

It has become clear that clefts in Malay are a kind of copular construction, judging from the surfacing of *jadi* (become) in certain contexts. In addition to that, the existence of a matrix copular clause with expletive *ia* as its subject, albeit usually covert, above the main visible part of the cleft, constitutes conclusive evidence that the cleft construction in Malay is bi-clausal. Combining the structure of the cleft clause in Figure 25 and the structure of the matrix copular clause in Figure 26 allows us to generate the full structure of a cleft in Malay below:

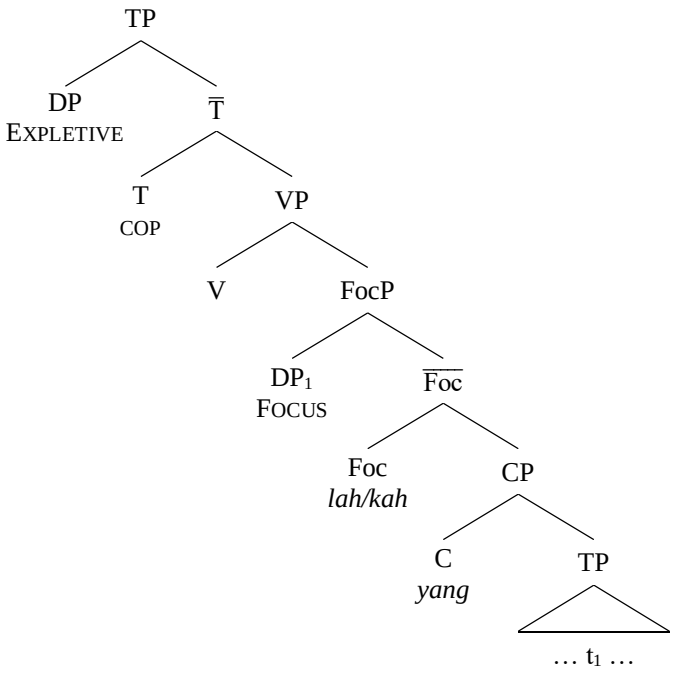


Figure 27: The Structure of a Cleft in Malay

5.2 The Derivation of Cleft Constructions in Malay

Cleft or pseudocleft first? I argue that it is pseudoclefts that are derived from clefts, contrary to the analyses by Kader (1981) and Cole and Hermon (2000). According to their analyses, the subject of a pseudocleft is a D/NP. However, a comparison of the cleft clause of clefts and pseudoclefts in Malay yields that the cleft clause of a cleft does not form a complex DP; rather, it is a bare CP. This finding makes it impossible for a cleft to be derived from a canonical copular clause whose post-copular DP has been focalised.

A further examination of pseudoclefts reveals that there exist pseudoclefts whose subject is a bare CP, as opposed to a complex DP (i.e. a canonical copular clause). The two varieties of pseudoclefts have different syntactic, semantic, and information-structural properties. Most importantly, the cleft clause of a cleft and a pseudocleft with a CP subject being of the same syntactic category permits an analysis of the derivation of the latter from the former, as opposed to the analysis by Kader (1981) and Cole and Hermon (2000).

I argue that the derivation of pseudoclefts from regular clefts first involves the steps explained in the previous section, namely movement of the focus to SpecFocP, which spells out complementiser *yang* in a lower CP projection, as in (295a), and the merging of a matrix copular clause, as in (295b). If an expletive is merged in the matrix copular clause, a regular cleft is derived. As for pseudoclefts, the cleft clause headed by *yang* may move to the matrix subject position to derive a pseudocleft, provided that no expletive is merged, as in (295c). This movement corresponds to *intraposition*, as in Williams (1980), whereby a clausal constituent is preposed to the beginning of a clause, as opposed to *extraposition*, whereby a clausal constituent is postposed to a position at the end of a clause.

(295) a. $[_{FocP}\ [\text{FOCUS}]_1\ [_{CP}\ \textit{yang}\ldots\ [_{TP}\ t_1]]]$

b. $[_{TP}\ (\text{EXPLETIVE})\ \text{COP}\ [_{FocP}\ [\text{FOCUS}]_1\ [_{CP}\ \textit{yang}\ldots\ [_{TP}\ t_1]]]]$

c. $[_{TP}\ [_{CP}\ \textit{yang}\ldots\ [_{TP}\ t_1]]_2\ \text{COP}\ [_{FocP}\ [\text{FOCUS}]_1\ t_2\ ]]$

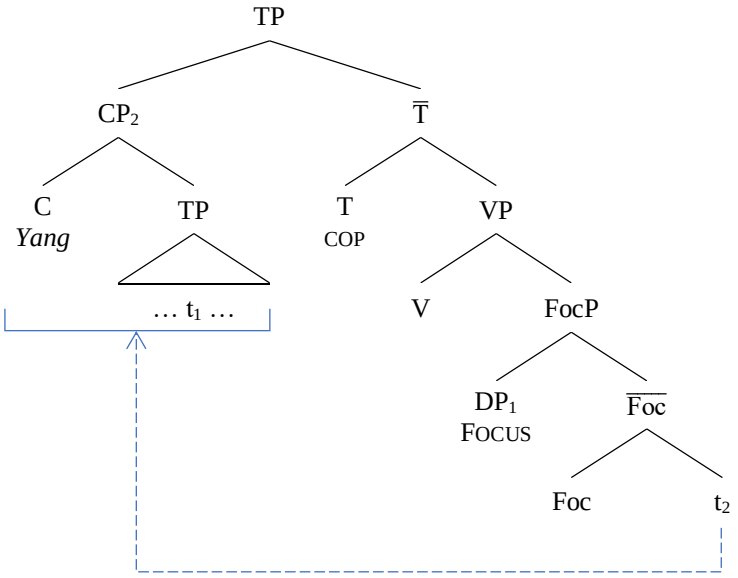


Figure 28: The Structure of a Pseudocleft in Malay

5.2.1 Against Extraposition of the Cleft Clause to Derive Clefts

Clefts appear similar to extraposition structures, whereas pseudoclefts appear similar to intraposition structures, as illustrated below:

(296) a. [$_{TP}$ Ø Ø *benar* [$_{CP}$ *bahawa Ali di-rompak.*]] (Extraposition)
EXPL COP true COMP A. PASS-rob
'It is true that Ali was robbed.'

b. [$_{TP}$ Ø Ø *Ali* [$_{CP}$ *yang di-rompak.*]] (Cleft)
EXPL COP A. COMP PASS-rob
'It was ALI that was robbed.'

c. [$_{TP}$ [$_{CP}$ *Bahawa Ali di-rompak*] *adalah benar.*] (Intraposition)
COMP A. PASS-rob COP true
'That Ali was robbed is true.'

d. [$_{TP}$ [$_{CP}$ *Yang di-rompak*] *ialah Ali.*] (Pseudocleft)
COMP PASS-rob COP A.
'Who was robbed was ALI.'

Construction	Subject	Copula	XP	Clause
Extraposition	Null expletive	Null copula	Predicate	*Bahawa*-clause
Cleft	Null expletive	Null copula	Referential DP	*Yang*-clause
Intraposition	*Bahawa*-clause$_1$	*adalah*	Predicate	t$_1$
Pseudocleft	*Yang*-clause$_1$	*ialah*	Referential DP	t$_1$

Table 26: The Similarities between Cleft Constructions and Extra- and Intra-position

Akmajian (1970b) and Percus (1997) have analysed clefts to involve extraposition, as illustrated below. The cleft clause, which starts out as a relative clause, is extraposed to adjoin to T/IP, making it occupy a position that is structurally higher than the rest of the construction. The spell-out rule then takes place such that the definite determiner and the null NP surfaces as *it*.

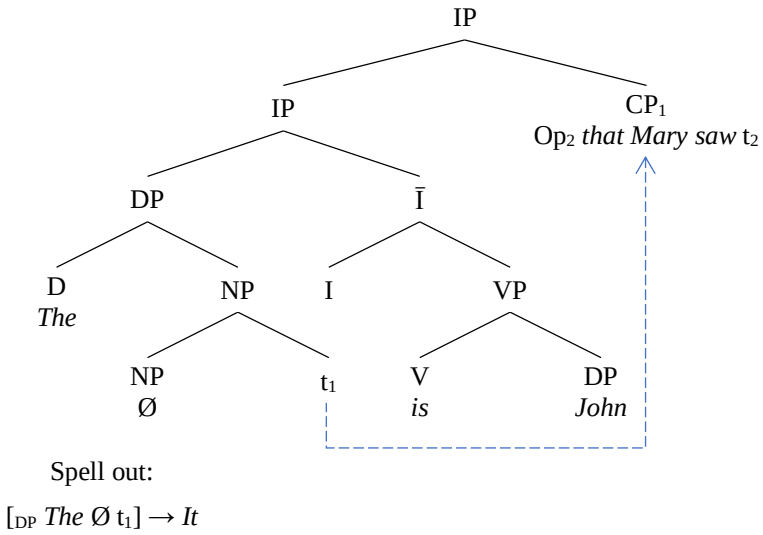


Figure 29: Extraposition in the Derivation of Clefts (Percus, 1997)

However, the structure that I have proposed in Figure 27 is in opposition to analyses of clefts that involve extraposition, especially with regard to the position of the cleft clause. Using the anti-c-command property of Negative Polarity Items (NPIs) – that NPIs cannot c-command their licensors – in Heycock and Kroch (2002), it can be discerned whether the cleft clause is higher than the focus. For instance, the topicalised NPI in example (297b) c-commands the negator that licenses it, rendering the example ungrammatical.

(297) a. *Ali tak boleh makan sebarang jenis kacang.*
A. NEG can eat any type nut
'Ali can't eat any type of nuts.'

b. *[*Sebarang jenis kacang*]$_1$ *Ali tak boleh makan* t$_1$.
any type nut A. NEG can eat
(Any type of nuts, Ali can't eat.)

Therefore, given the structure of clefts in Malay in Figure 27, clefting an NPI should be impossible. Indeed, example (298) is ungrammatical, which entails that the focus c-commands into the cleft clause. This finding shows that topicalisation and clefting are similar in that both operations move a constituent to a high position in the left periphery and that the rest of the clause occurs below the moved constituent.

(298) * *Sebarang jenis kacang yang Ali tak boleh makan.*
any type nut COMP A. NEG can eat
(It is any type of NUTS that Ali can't eat.)

Furthermore, considering that the subject in Figure 29 is a definite DP, the *it*-subject that is spelt out from the definite description cannot be semantically vacuous, as argued by Elbourne (2001). However, this (non-)spell-out is unable to reflect facts about pronominal subjects in Malay in general. For instance, the actual application of extraposition produces a sentence in Malay that does not resemble a cleft, especially with respect to the subject. As shown in the example below, extraposition of the relative clause does not allow the deletion of the remnant DP. Rather, *yang* is spelt out to support the demonstrative that has been left stranded by the null NP.

(299) *Ø yang itu Ali, yang saya nampak tadi.*
N LIG DIST A REL 1.SG see just.now
'That one is Ali, whom I saw just now.'

Besides, 3rd person pronouns generally cannot be deleted, unlike 1st person and 2nd person pronouns. As shown below, the optionally deleted subject may only refer to the speaker or the listener. However, inanimate antecedents do allow optional pro-drop of the pronoun:

(300) a. *Ø lapar-kah?*
Pro hungry-DISJ.Q
i. 'Am I hungry?' (Pro = 1.SG *saya*)
ii. 'Are you hungry?' (Pro = 2.SG *awak*)
iii. # 'Is he hungry?' (Pro = 3.SG.H *dia*)

b. *Ø mahal-kah?*
Pro expensive-DISJ.Q
'Is it expensive?' (Pro = 3.SG.NH *ia*)

If it were a contentful 3rd person pronoun, the subject of the matrix copular clause of a cleft in Malay should be spelt out normally, regardless of the animacy of the antecedent, given that pro-drop is optional. On the contrary, spell out of the pronoun in a cleft in Malay is ungrammatical, as shown below. The subject of the matrix copular clause of a cleft in Malay is confirmed to be an expletive, as it is the only pronominal form in the language that does not have any phonetic realisation in normal circumstances.

(301) a. (**Ia*) *kacang yang Ali tak boleh makan.*
3.SG.NH nut COMP A. NEG can eat
'It is NUTS that Ali can't eat.'

b. (**Dia*) *Ali yang tak boleh makan kacang.*
3.SG.H A. COMP NEG can eat nut
'It is ALI that can't eat nuts.'

So far, we have concluded that the cleft clause is structurally lower than the clefted focus and we have evidence for a matrix copular clause, which has allowed the generation of the structure in Figure 27. In consideration of these two findings, I have also argued against extraposition of the cleft clause in the derivation of a cleft to defend the structure in Figure 27.

5.2.2 For Intraposition of the Cleft Clause to Derive Pseudoclefts

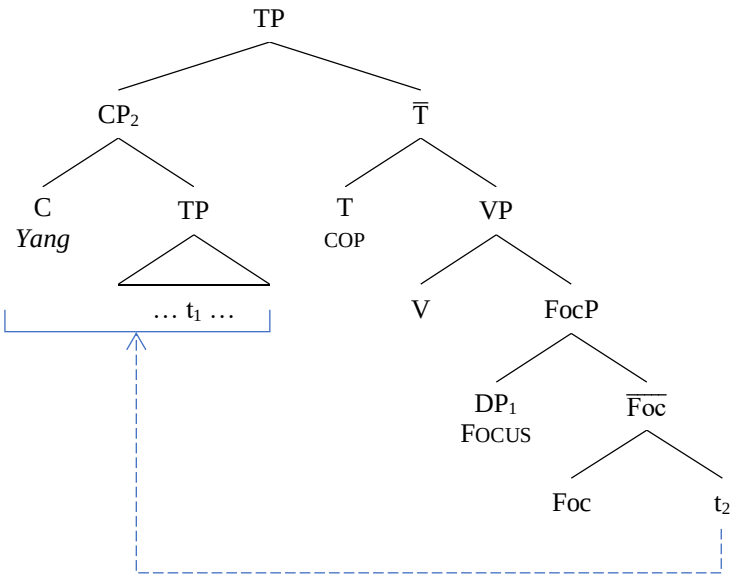
Moving on to the derivation of pseudoclefts in Malay, to recapitulate, a matrix copular clause is merged following movement of the focus and spell-out of *yang* in CP. In lieu of merging an expletive in the subject position of the matrix copular clause to derive a cleft, the cleft clause is moved to the subject position to derive a pseudocleft.

In contrast to clefts and extraposition structures, pseudoclefts are similar to intraposition structures, which involve movement of the clausal constituent to the subject position (Williams, 1980). As opposed to extraposition, it is more likely for the cleft clause to undergo intraposition, especially considering that the cleft clause is structurally lower than the clefted focus.

(302) a. [TP [CP *Bahawa Ali di-rompak*] *adalah benar.*] (Intraposition)
COMP A. PASS-rob COP true
'That Ali was robbed is true.'

b. [$_{TP}$ [$_{CP}$ *Yang* *di-rompak*] *ialah* *Ali.*] (Pseudocleft)
COMP PASS-rob COP A.
'Who was robbed is ALI.'

Granted the structure of a cleft in Malay in Figure 27, the derivation of a pseudocleft can proceed quite straightforwardly with movement of the cleft clause to the matrix subject position, instead of the merging of an expletive. Technically speaking, a pseudocleft is not derived *from* a cleft, but it is a further continuation of a derivation that could have resulted in a cleft.



As seen in Figure 28 (repeated above), the structure of a pseudocleft is similar to the analysis of intraposition by Williams (1980), but departs from it when it comes to the position of the clausal constituent. It is argued by Koster (1978), among others, based on the following observations on the syntactic behaviour of sentential subjects, that they occupy the topic position, instead of the subject position:

- a sentential subject cannot undergo subject-auxiliary inversion;
- a sentential subject cannot co-occur with a fronted constituent;
- a sentential subject cannot occur in a subordinate clause.

As shown below, the English data by Koster (1978) are ungrammatical:

(303) a. **Did that John showed up please you?*

b. **Such things that he reads so much doesn't prove.*

c. **Although that the house is empty may depress you, it pleases me.*

However, sentential subjects in Malay do not conform to such restrictions observed in English. The sentential subject of an intraposition structure and the cleft clause of a pseudocleft can be demonstrated to behave in ways that are contrary to these observations, which suggests that they occupy the subject position. The three restrictions are tested on the following examples:

(304) a. *Bahawa Zul meŋ-curi duit itu adalah benar.* (Intraposition)
COMP Z. ACT-steal money DIST COP true
'That Zul stole the money is true.'

b. *Yang meŋ-curi duit itu ialah Zul.* (Pseudocleft)
COMP ACT-steal money DIST COP Z.
'Who stole the money is ZUL.'

Subject-auxiliary inversion: ✓

(305) a. *Ada-kah bahawa Zul meŋ-curi duit itu benar?*
AUX-Q COMP Z. ACT-steal money DIST true
'Is it true that Zul stole the money?'

b. *Ada-kah yang meŋ-curi duit itu Zul?*
AUX-Q COMP ACT-steal money DIST Z.
'Was who stole the money ZUL?'

Occurring with a fronted constituent: ✓

(306) a. *Duit itu kan, bahawa Zul meŋ-curi=nya adalah benar.*[70]
money DIST TOP COMP Z. ACT-steal=3 COP true
'The money, that Zul stole it is true.'

b. *Duit itu kan, yang meŋ-curi=nya ialah Zul.*
money DIST TOP COMP ACT-steal=3 COP Z.
'The money, who stole it was ZUL.'

Functioning as the subject of a subordinate clause: ✓

(307) a. *Meskipun bahawa Zul meŋ-curi duit itu adalah benar, ia di-pertikai-kan.*
although COMP Z. ACT-steal money DIST COP true 3.SG PASS-dispute-APPL
'Although it is true that Zul stole the money, it is disputed.'

[70] Additionally, examples (306) make it even more convincing that the sentential subject and the cleft clause occupy the subject position, because the fronted constituents originate in the clausal constituents themselves. The use of resumptive *-nya* is needed to repair the violation of the subject condition.

b. *Meskipun yang meŋ-curi duit itu ialah Zul, ia di-per-tikai-kan.*
although COMP ACT-steal money DIST COP Z. 3.SG PASS-CAUS-different-APPL
'Although who stole the money was ZUL, it is disputed.'

Furthermore, it is possible to raise the sentential subject and cleft clause into the subject position of a matrix clause via raising and passivisation, which are attested to be A-movement and triggered by the EPP.

(308) a. *Bahawa Zul meŋ-curi duit itu di-sah-kan adalah benar.*
COMP Z. ACT-steal money DIST PASS-verify-APPL COP true
'That Zul stole the money was verified to be true.'

b. *Yang meŋ-curi duit itu di-sah-kan ialah Zul.*
COMP ACT-steal money DIST PASS-verify-APPL COP Z.
'Who stole the money was verified to be ZUL.'

The subject-like behaviour of the clausal constituent in intraposition structures and pseudoclefts, as summarised in Table 27, supports the claim that they occupy the subject position, as opposed to a topic position. Therefore, it is possible to extend the analysis of intraposition to the derivation of a pseudocleft in Malay, owing to the parallel structures of the two constructions.

	Intraposition	Pseudocleft
Subject-Auxiliary Inversion	✓	✓
With Fronted Constituent	✓	✓
Subject of Subordinate Clause	✓	✓
Raising/Passivisation	✓	✓

Table 27: The Subject-like Behaviour of the Clausal Constituent

5.2.3 Connectivity as a Corollary of Intraposition

Pseudoclefts are well-known to exhibit connectivity effects. For example, the focus in the English pseudocleft below is a constituent that contains a reflexive pronoun whose use apparently violates Condition A of Binding Theory; the antecedent is not local to and does not c-command the reflexive. The seemingly unlicensed use of the reflexive should render the example ungrammatical, but it is nonetheless grammatical.

(309) ‘*What he is is a danger to himself*.’

The focus is interpreted to originate in the cleft clause, e.g. “He is a danger to himself”. However, the gap in the cleft clause corresponds to the *wh*-phrase. Unlike in English, the clausal subject of a pseudocleft in Malay is not introduced by a *wh*-phrase, which means that the two gaps within it could correspond to the focus.

(310) ____$_1$ *Yang Abu pukul* ____$_1$ *ialah diri=nya sendiri*$_1$.
COMP A. hit COP self=3 alone
‘Who Abu hit was himself.’

The analysis proposed in the previous sections makes it possible to attribute the connectivity in pseudoclefts to reconstruction in a straightforward fashion, as illustrated below. The movement of the focus from the cleft clause and subsequent movement of the cleft clause to subject position do not bleed reconstruction.

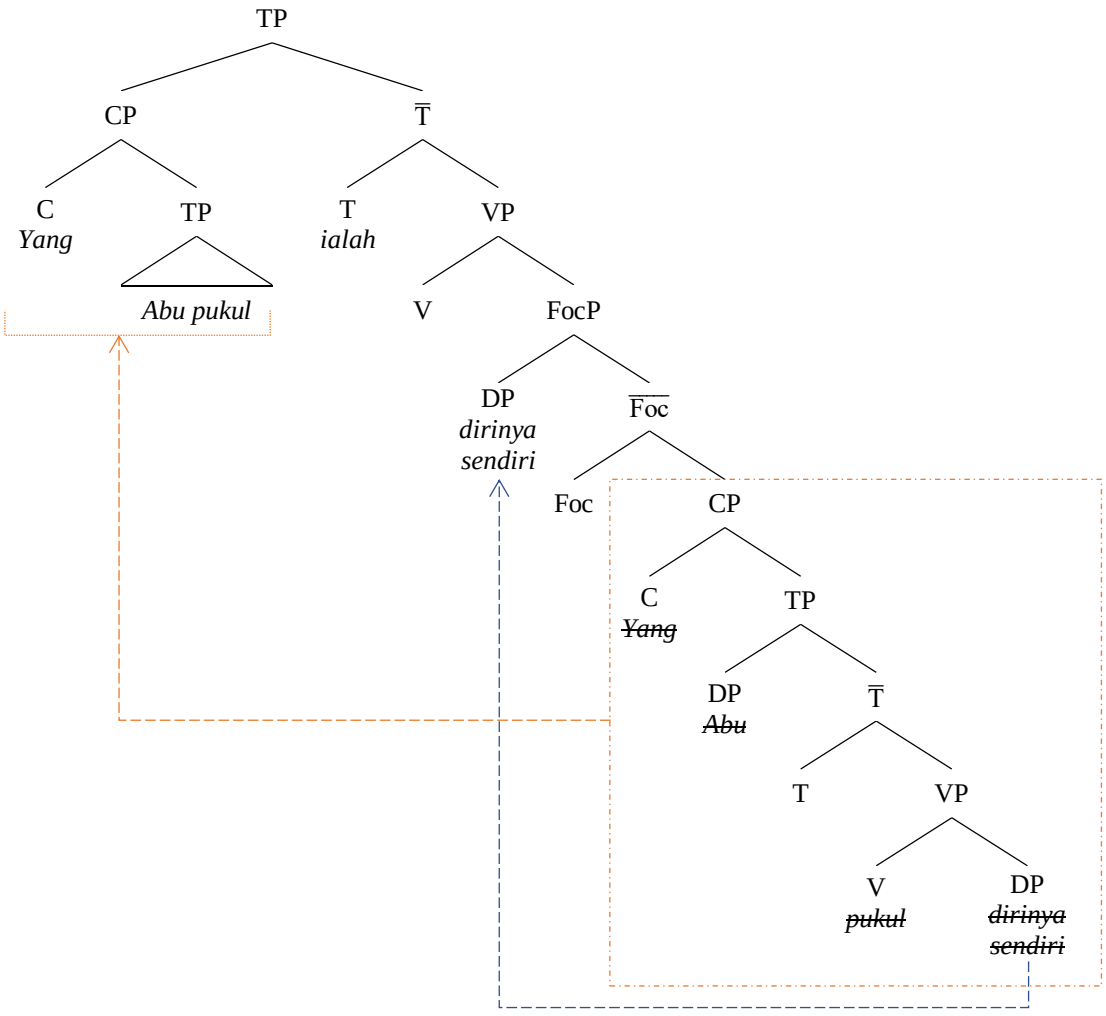


Figure 30: The Reconstruction and Connectivity Caused by Movement

The same connectivity effect is observed with NPI licensing:

Context: *Dia boleh makan apa-apa sahaja.* (He can eat anything.)

(311) ____$_1$ *Yang dia tak boleh makan* ____$_1$ *ialah sebarang jenis kacang.*
COMP 3.SG NEG can eat COP any type nut
'What he can't eat is any type of nuts.'

The structures by Kader (1981) and Cole and Hermon (2000) are not capable of accounting for the reconstruction of the reflexive DP because the structures analyse the focus as having been base-generated outside the cleft clause that contains the DP that should locally c-command it. Neither can an analysis involving extraposition of the cleft clause account for this reconstruction, as at no point in the derivation is the gap within the cleft clause not c-commanded by the focus.

5.2.4 On the Split-CP Hypothesis and Phasehood

Given the Split-CP hypothesis by Rizzi (1997) and the Phase Impenetrability Condition by Chomsky (2001), the question arises: which of the projections within the split CP constitute phases? This question is important as it concerns the derivation of pseudoclefts in Malay, which involves multiple CP projections and raising of the cleft clause to the subject position of the matrix copular clause, as proposed in Section 5.1.1.

With regard to Malay, I agree with Totsuka (2015), who argues that TopP and ForceP are phases, as opposed to FocP and FinP, based on facts concerning islandhood, among others. Using the following examples from Koizumi (1999), it is argued that extraction is impossible out of TopP, but not out of FocP. Examples (312a-c) illustrate that topicalisation has induced syntactic islands out of which *wh*-phrases and relative operators cannot move. Conversely, examples (312d-f) are grammatical since no island is formed following focalisation.

(312) a. **On which table did Lee say that* these books *she will put?*

b. **Which books did Becky say that* to Aaron *she will give?*

c. **This is the book that John said that* Mary$_1$ *he would inform* t$_1$ *that I had read.*

d. *On which table did Lee say that* ***only these books*** *would she put?*

e. *Which books did Becky say that **only to Aaron** will she give?*

f. *This is the book that John said that **only Mary**$_1$ would he inform* t$_1$ *that I had read.*
(Totsuka, 2015, p. 28)

Using islandhood as a diagnostic for phasehood, it can be concluded that TopP constitutes a phase in Malay, in contrast to FocP. As illustrated below, topicalisation has induced a syntactic island and rendered the clause inaccessible to further syntactic operations. Therefore, extraction of the *wh*-phrase is banned.

(313) a. *Di atas meja mana-kah*$_1$ *Ali fikir* [$_{CP}$ *dia letak buku itu* t$_1$ *semalam?*]
LOC on table which-Q A. think 3.SG put book DIST yesterday
'On which table did Ali think he put that book yesterday?'

b. *Di atas meja mana-kah*$_1$ *Ali fikir* [$_{FocP}$ *semalam -lah dia letak buku itu* t$_1$?]
LOC on table which-Q A. think yesterday -FOC 3.SG put book DIST
'On which table did Ali think, YESTERDAY, he put that book?'

c. **Di atas meja mana-kah*$_1$ *Ali fikir* [$_{TopP}$ *semalam kan, dia letak buku itu* t$_1$?]
LOC on table which-Q A. think yesterday TOP 3.SG put book DIST
(On which table did Ali think, yesterday, he put that book?)

The phasehood of TopP correctly predicts that the derivation of a pseudocleft is made impossible if the topic field is activated during the derivation of a regular cleft. To illustrate, topicalisation of the adverb in example (314a) prevents raising of the cleft clause to the subject position of the matrix copular clause to derive the pseudocleft in example (314b). Given the phasehood of TopP, topicalisation from a cleft creates an island, which entails that remnant movement of FinP to SpecTP to derive a pseudocleft should be impossible. Probing by the matrix T^0 is not possible as FinP is too deeply embedded within the split CP whose phasal layer is TopP.

(314) a. [$_{TP}$ Ø Ø [$_{TopP}$ *semalam kan* [$_{FocP}$ *Ali* [$_{FinP}$ *yang* [$_{TP}$ *baca buku itu.*]]]]]
EXPL COP yesterday TOP A. COMP read book DIST
'Yesterday, it was ALI who read that book.'

b. *[$_{TP}$ [$_{FinP}$ *Yang* [$_{TP}$ *baca buku itu*]]$_1$ *ialah* [$_{TopP}$ *semalam kan* [$_{FocP}$ *Ali* t$_1$.]]]
COMP read book DIST COP yesterday TOP A.
(Who read that book was, yesterday, Ali.)

As for FinP, it appears not to be a phase. Based on example (315), movement of the adverb to FinP (which does not seem to be topicalisation due to the ungrammaticality of the topic marker) does not prevent movement of the focus to a position above it.[71]

(315) *Buku itu -lah semalam* (**kan*) *yang Ali baca.*
book DIST-FOC yesterday TOP COMP A. read
'It was that book that yesterday Ali read.'

Considering the optionality and independence of the activation of the topic and focus fields, I assume that ForceP is not lexicalised needlessly. When it is activated, its head usually surfaces as *bahawa*. For example, ForceP is projected and headed by *bahawa* in the following example alongside activation of the topic and focus fields. Also, given that the higher CP projections have been activated, FinP should be activated as well, by implication, even if it is unpronounced.

(316) *Ali kata*[$_{\text{ForceP}}$ *bahawa* [$_{\text{TopP}}$ *semalam kan* [$_{\text{FocP}}$ *di sini-lah* [$_{\text{FinP}}$ [$_{\text{TP}}$ *dia tidur.*]]]]]
A. say COMP yesterday TOP LOC here-FOC 3.SG sleep
'Ali said that, yesterday, HERE he slept.'

On the same note, it can be assumed that the CP need not be split at all if none of the other fields are activated. This can be said of complement clauses with *bahawa* that do not involve topicalisation or focalisation, as well as relative clauses in Malay, which allow neither case of movement, regardless of its landing position within the left periphery, as illustrated below:

(317) a. *Saya suka buku* [$_{\text{CP}}$ *yang Ali baca semalam.*]
1.SG like book REL A. read yesterday
'I like the book that Ali read yesterday.'

[71] Recall that extraction of an adverb does not spell out *yang*. Only after focalisation of the object in the following example does *yang* surface. The examples below also entail that a focus does not cyclically move to FinP, which *yang* heads, before landing in FocP.

(xxii) a. *Semalam* (**yang*) *Ali baca buku itu.* → b. *Buku itu -lah semalam yang Ali baca.*
yesterday COMP A. read book DIST book DIST -FOC yesterday COMP A. read
'Yesterday Ali read that book.' 'It was that book that yesterday Ali read.'

b. **Saya suka buku* [$_{CP}$ *yang, semalam kan, Ali baca.*]
1.SG like book REL yesterday TOP A. read
(I like the book that, yesterday, Ali read.)

c. **Saya suka buku* [$_{CP}$ *yang semalam -lah Ali baca.*]
1.SG like book REL yesterday -FOC A. read
(I like the book that YESTERDAY Ali read.)

d. **Saya suka buku* [$_{CP}$ *semalam* {*kan* /*-lah* } *yang Ali baca.*][72]
1.SG like book yesterday TOP -FOC REL A. read
(I like the book that, yesterday, Ali read.)

In cases in which ForceP is not projected, the properties of Force are carried by the head of the highest activated projection within the split CP. For example, the head of the FocP in a cleft can be said to be a combination of Foc and Force, i.e. is not split into two different heads/projections. Given that both *wh*-movement and focalisation target FocP in Malay – e.g. by the correspondence of *kah* as the interrogative counterpart of focus marker *lah* according to Kader (1981) – there is no actual syntactic function for ForceP other than to accommodate complementiser *bahawa* if it is merged at all. One piece of evidence for the union of the two heads is the incompatibility of *bahawa* with the interrogative mood, which indicates that illocutionary force and focus are both carried by the single head *kah*, as illustrated below. As *kah* carries the features of both Force and Focus, there is no merit to assuming that ForceP is projected, at least in interrogative clauses.

(318) *Ali fikir* (**bahawa*) *semalam-kah dia baca buku itu*?
A. think COMP yesterday-Q 3.SG read book DIST
'Did Ali think that he read that book yesterday?'

Also, the fact that *bahawa* cannot occur in the cleft clause of a cleft suggests that ForceP is not and, in fact, cannot be activated in that environment.[73]

[72] Without the topic or focus marker, this example would be grammatical but only on the reading that *semalam* is a possessive, e.g. *buku semalam yang*… (yesterday's book that…)

[73] This is in spite of its possibility in a normal complement clause with an activated focus field, which suggests different licensing properties for ForceP.

(xxiii) [$_{TP}$ *Ali mesti fikir* [$_{ForceP}$ *bahawa* [$_{FocP}$ *semalam -lah* [$_{TP}$ *dia baca buku itu.*]]]]]
A. must think COMP yesterday -FOC 3.SG read book DIST
'Ali must have thought that yesterday he read that book.'

(319) [$_{TP}$ Ø *mesti* Ø [$_{ForceP}$ (**bahawa*) [$_{FocP}$ *Ali -lah* [$_{FinP}$ *yang* [$_{TP}$ *baca buku itu.*]]]]]
EXPL must COP COMP A. -FOC COMP read book DIST
'It must be ALI who read that book.'

Along with its features, the phasehood of ForceP is carried by the highest associated left-peripheral projection. Although FocP does not constitute a phase on its own, e.g. (313b), its combination with the properties of Force grants it phasehood. This should also be true of TopP. If there is no need for ForceP, it can be said to combine with TopP as the highest left-peripheral projection when the topic field is activated.

Finally, the possibility of raising FinP entails that it is accessible to the matrix T^0 probe to trigger movement to the matrix subject position. Although the PIC states "at a phase, only the next lower phase head and its specifier(s) are available for operations" (Chomsky, 2001), it should technically be possible for the whole complement of the phase head, to be visible as a whole constituent, much like a quotative that itself could be moved but not its contents. Otherwise, it should not be possible for the cleft clause to be topicalised, despite having undergone transfer, as shown below. As a constituent that is merged where it is and has not itself undergone any type of movement, FinP is open to becoming the tail of either an A- or Ā-chain.

(320) *Yang baca buku itu kan, Zul tahu sebenarnya Ali.*
COMP read book DIST TOP Z. know actually A.
'Who read that book, Zul knows it was actually Ali.'

5.2.5 Topic as the Trigger for Movement

As described in Section 3.3.2, specificational copular clauses exhibit a fixed topic-focus alignment. The complement of $Pred^0$, which normally corresponds to the nonverbal predicate in a predicational copular clause, is topical in nature and ends up being the subject of a specificational copular clause. Based on the analysis by Mikkelsen (2005b), T^0 of a specificational copular clause has an uninterpretable Topic feature [*u*Top] in addition to the conventional uninterpretable case and EPP features. It happens that the constituent that serves as the predicate possesses all three features. Therefore, the predicate is raised to SpecTP to check all the uninterpretable features on T^0 and the derivation of the specificational copular clause converges.

Specificational pseudoclefts have the same information-structural alignment too. Given the specificational nature of pseudoclefts, the cleft clause can be analysed to be triggered to move by the same [*u*Top] feature.[74] In fact, the constituents flanking the copula in a pseudocleft in Malay cannot undergo copular inversion, unlike pseudoclefts in English. The cleft clause, which corresponds to the topic, always occurs in pre-copular position, whereas the focus always occurs in post-copular position. Conversely, canonical copular clauses may undergo copular inversion, provided that the appropriate copula is used, as illustrated below:

(321) a. *Yang akan di-dakwa ialah Ali.* (Pseudocleft)
COMP PROS PASS-prosecute COP A.
'Who will be prosecuted is ALI.'

b. **Ali ialah yang akan di-dakwa.*
A. COP COMP PROS PASS-prosecute
(Ali is who will be prosecuted.)

c. *Orang yang akan di-dakwa itu ialah Ali.* (Copular Clause)
person REL PROS PASS-prosecute DIST COP A.
'The person who will be prosecuted is Ali.'

d. *Ali adalah orang yang akan di-dakwa itu.*
A. COP person REL PROS PASS-prosecute DIST
'Ali is the person who will be prosecuted.'

The analysis by Mikkelsen (2005b) can be extended to pseudoclefts in Malay, but with slight adjustments: T^0 probes for a topic in its c-command domain, but the complement of the copula is apparently the split CP that contains both the focus and the cleft clause, granted the structure of a cleft in Figure 27. Arguably, the topical element is the cleft clause to the exclusion of the focus, which is the FinP. Therefore, FinP agrees with T^0, checks the [*u*Top] feature, and gets raised to SpecTP. As illustrated below, the T^0 head of the matrix copular clause possesses the [*u*Top] feature that triggers the topical FinP to move to its specifier:

[74] Considering that Malay is a topic-prominent language that puts importance on what is construed as the topic of a clause, as per Austronesian alignment, this information-structural trigger to obtain the topic-focus alignment should be relevant.

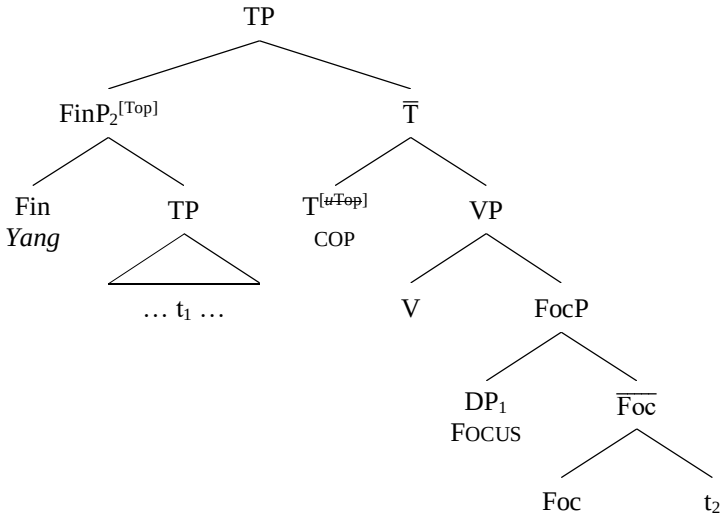


Figure 31: The [uTop] Feature in Pseudoclefts in Malay

Although clefts and pseudoclefts in Malay are derived the same way, the difference between a cleft and a pseudocleft in Malay is the presence of the [*u*Top] feature on T^0. The absence of this feature prevents FinP from raising to SpecTP, so an expletive is merged to satisfy the EPP, and a regular cleft is derived.

There is an interesting observation that provides evidence for this hypothesis. It appears that once the focus has undergone clefting, it can no longer move, i.e. it is frozen in its position below the copula, which is an effect known as Criterial Freezing (Rizzi, 2010). Example (322c) illustrates the impossibility of the focus to Ā-move across the auxiliary in the matrix copular clause or a higher clause.

(322) a. [$_{TP}$ *Ali curi duit itu.*] (Simplex)
A. steal money DIST
'Ali stole the money.'

b. [$_{TP}$ Ø *mesti* Ø [$_{FocP}$ *Ali-lah yang curi duit itu.*]] (Cleft)
EXPL must COP A.-FOC COMP steal money DIST
'It must be ALI who stole the money.'

c. *[$_{FocP}$ *Ali-lah* [$_{TP}$ *saya kata* [$_{CP}$ [$_{TP}$ Ø *mesti* Ø [$_{FocP}$ *yang curi duit itu.*]]]]]
A.-FOC 1.SG say EXPL must COP COMP steal money DIST
(ALI, I said it must be who stole the money.)

Now it is only a matter of explicating what makes the cleft clause topical to allow it to move to the matrix TP to check the [*u*Top] feature.

5.2.6 *Yang* and Topicality

Yang is associated with different functions when its distribution outside the uncontroversial domains of relativisation and complement clauses is examined. In other words, the function of *yang* in complement and relative clauses can be identified quite straightforwardly, but its occurrence in different information-structural environments, such as clefts and focus, as well as givenness and topics, makes it difficult to categorise, especially because it is formally invariable, i.e. syncretic.

A brief survey of the literature pertaining to this lexeme reveals the different guises taken by *yang*, as tabulated below:

Author	Function
Various	Ligature, complementiser, relativiser, nominaliser
(Winstedt, 1913)	Akin to definite article
(Simin, 1988)	Context-sensitive ligature
(Saddy, 1992)	Focus marker
(van Minde, 2008)	(New/contrastive) topic/theme marker
(Yap, 2011)	Topic marker

Table 28: The Functions of 'Yang'

Although *yang* occurs in clefts, it does not serve the purpose of signalling focus, as that is the role taken by focus marker *lah*. Rather, *yang* signals the beginning of the constituent that corresponds to the presupposition or topic, i.e. the element that conveys information that is given, as shown in (323a). Such a partitioning of a cleft construction is attested, as in Gundel and Fretheim (2004): "It is widely accepted that in canonical clefts with a single prominent pitch accent on the clefted constituent…, the clefted constituent is the information focus and the open proposition expressed by the cleft clause… is presupposed and topical" (p. 10). In other words, *yang* clearly defines the information-structural partition of the cleft, according to frameworks of relational givenness (e.g., Gundel & Fretheim, 2004; Halliday, 1967; Reinhart, 1982; Rooth, 1985). Furthermore, the formation of a pseudocleft involves raising of the constituent headed by *yang* to the matrix subject position, which entails that *yang* forms a constituent with the topical clausal constituent, rather than the clefted focus.

(323) a. [Focus *Ali*] [Topic *yang meŋ-curi duit itu.*] (Cleft)
A. COMP ACT-steal money DIST
'It is ALI who stole the money.'

b. [Topic *Yang meŋ-curi duit itu*] *ialah* [Focus *Ali.*] (Pseudocleft)
COMP ACT-steal money DIST COP A.
'Who stole the money is ALI.'

Even in non-cleft constructions, the use of *yang* can be attributed with signalling old and topical information. For instance, the left-dislocated topic in the examples below is seen to be modified by *yang*.

(324) *Yang dia, sedikit pun dia tak kesah.*
TOP 3.SG little.bit also 3.SG NEG care
'As for him, he didn't even care the slightest bit.'
(Yap, 2011, p. 4)

Also, the use of supportive *yang* in NP-ellipsis (discussed further in Section 5.3.6) is related to topicality. The NP to be deleted and supported by *yang* must be given in the discourse to allow recoverability. In the following example, the deleted NP must have an antecedent in the context. Since *monyet* (monkey) is not a given NP, the deleted NP cannot correspond to it.

Context: *Saya benci* [Topic *orang yang suka bergaduh*] (I hate [Topic people who like to fight])

(325) ... *dan* [Topic { ~~*orang*~~ /* ~~*monyet*~~ } *yang suka meŋ-curi.*]
… and person monkey REL like ACT-steal
'… and those who like to steal.'

Therefore, it is no surprise that the whole constituent headed by *yang* can be considered a topic, making plausible the hypothesis that the FinP is a topic.

5.3 Against the Cleft Clause as a Complex DP

A cleft in Malay looks identical to a complex DP. In fact, the resemblance between clefts and relative clauses is very common cross-linguistically, e.g. Percus (1997) analyses the derivation of an *it*-cleft in English to involve extraposition of a relative clause. The cleft clause in Malay appears to be a relative clause introduced by *yang*:

(326) a. [Focus *Air*] [Cleft Clause *yang di-minum.*] (Cleft)
water COMP PASS-drink
'It was WATER that was drunk.'

b. [NP *Air*] [Relative Clause *yang di-minum…*] (Relative Clause)
water REL PASS-drink
'Water that was drunk…'

Due to the uncanny resemblance between the cleft clause and a relative clause, scholars such as Kader (1981) and Cole and Hermon (2000) advocate an analysis of cleft constructions that involves a headless relative clause. As shown in Figure 23 and Figure 24, the cleft clause is a relative clause that forms part of a complex DP that functions as the subject of a canonical copular clause.

[CP FOCUS$_1$ [TP [DP Ø [Relative Clause *yang*…]] COP t$_1$]]

The DP status of the subject makes it possible to liken the subject to a free relative clause, which is traditionally analysed to be a DP (see Bresnan & Grimshaw, 1978). As shown below, the clausal subject may be introduced by a *wh*-phrase:

(327) [FRC *Apa yang di-minum itu*] *ialah* [Focus *air.*]
what COMP PASS-drink DIST COP water
'What was drunk was water.'

For comparison, the examples below show the use of free relative clauses in different positions that are associated with nominal constituents.

(328) a. *Apa yang dia minum itu meŋ-buat saya muntah.* (Subject)
what COMP 3.SG drink DIST ACT-make 1.SG vomit
'What he drank made me vomit.'

b. *Saya tak suka apa yang dia minum itu.* (Object)
1.SG NEG like what COMP 3.SG drink DIST
'I don't like what he drank.'

c. *Saya ter-kesan oleh apa yang dia minum itu.* (Complement of P)
1.SG NVOL-affect by what COMP 3.SG drink DIST
'I am affected by what he drank.'

Given the DP status of free relative clauses, I shall also call copular clauses with free relative clauses canonical copular clauses, not pseudoclefts, which are separate constructions with bare CP constituents as the clausal subject.

To further defend the structure in Figure 27, I argue that clefts in Malay are not derived from canonical copular clauses whose subject is a complex DP (a free relative clause or a complex DP with a null NP head). It can be shown that the cleft clause is not a relative clause, contrary to the analysis by Kader (1981) and Cole and Hermon (2000). Although their analysis suitably describes the structure of a canonical copular clause whose post-copular constituent has been focalised, it does not have the power to account for the behaviour observed of clefts in terms of semantics, licensing, information structure and syntax. There is no evidence of a complex DP in clefts in Malay, judging from the non-DP-like behaviour of the cleft clause: the cleft clause cannot be modified by a definitising element, it does not allow optional spell-out of the NP head, and it does not form a syntactic island, etc. Therefore, the structures proposed by Kader (1981), Cole and Hermon (2000) and similar scholars may not represent the correct analysis for the derivation of clefts in Malay.

5.3.1 Syntactic Islands

The difference between relative clauses and clefts with respect to the structure of the left periphery translates into different phasal properties, as described in Section 5.2.4. Since FocP is not a phase, a topic may escape a cleft and undergo further long-distance movement to the left periphery of the matrix copular clause, as shown below:

(329) [$_{TopP}$ *Ikan itu*$_1$ *kan*, [$_{TP}$ Ø *mesti* Ø [$_{TopP}$ t$_1$ [$_{FocP}$ *Zul* [$_{FinP}$ *yang masak* t$_1$.]]]]]
fish DIST TOP EXPL must COP Z. COMP cook
'That fish, it must be ZUL who cooked it.'

According to Ross (1967), relative clauses are strong syntactic islands out of which linguistic material cannot be moved. When a constituent is extracted from a relative clause, a violation of the *Complex NP Constraint*, which states that "no element contained in a sentence dominated by a noun phrase with a lexical head noun may be moved out of that noun phrase by a transformation" (p. 127), causes ungrammaticality. As described in examples (317), the left periphery of relative clauses in Malay is not split as the focus and topic fields cannot be activated, hence there being a single phasal projection, which gets blocked by the relative operator. The complex NP constraint is demonstrated in the example below in which the object of the relative clause has been illicitly topicalised:

(330) *Ikan itu$_1$ kan, [$_{DP}$ orang yang masak t$_1$] telah pulang.
fish DIST TOP person REL cook PRF return
(That fish, the person who cooked [it] has returned.)

As a means of repairing island violations, resumption must be employed. The following example illustrates repairing of the island violation by resumption:

(331) Ikan itu$_1$ kan, [$_{DP}$ orang yang masak=nya$_1$] telah pulang.
fish DIST TOP person REL cook=3 PRF return
'That fish, the person who cooked it has returned.'

Since the subject of the canonical copular clause by Kader (1981) and Cole and Hermon (2000) prior to clefting is a complex DP, resumption is expected after topicalisation has taken place, as shown below:

(332) Ikan itu$_1$ kan, yang masak=nya$_1$ itu ialah Zul.
fish DIST TOP REL cook=3 DIST COP Z.
'That fish, the (person) who cooked it is Zul.'

Therefore, if the cleft clause truly were a relative clause, the whole DP constituent would form a complex NP island. Resumption should be obligatory to repair the island violation caused by topicalisation out of the cleft clause, according to the analysis by Kader (1981) and Cole and Hermon (2000); however, this prediction is not borne out, as extraction from the cleft clause does not necessitate the use of resumptive clitic *-nya*. Based on the examples below, there is no violation of the constraint when material within the cleft clause has moved out. The examples are grammatical, which signifies that there is no island violation, hence no complex DP.

(333) a. Ikan itu$_1$ kan, Zul yang masak t$_1$.
fish DIST TOP Z. COMP cook
'That fish, it was ZUL who cooked it.'

b. Ikan itu$_1$ kan, siapa yang masak t$_1$?
fish DIST TOP who COMP cook
'That fish, WHO was it that cooked it?'

The lack of resumption in examples (333) challenges the claim by Kader (1981) and Cole and Hermon (2000) that clefts are derived from copular clauses whose subject is a complex DP. It is mysterious why resumption is not required in a cleft, when it is obligatory in a relative clause. If their analysis were correct, movement of linguistic material out of the cleft clause should also violate the complex NP constraint.

5.3.2 Inversion

Pseudoclefts and canonical copular clauses also differ in the syntactic operations that they may undergo. The DPs flanking the copula in a canonical copular clause may undergo inversion, as shown in examples (334a-d). Conversely, the same cannot be said of pseudoclefts in Malay, which happen to have a fixed constituent order in which the clausal subject is always pre-copular.

Context: *Kapal-kapal itu telah hanyut.* (The ships have drifted.)

(334) a. *Apa yang ber-laku ialah tsunami.* (Canonical Copular Clause with *Wh*-Phrase)
what COMP INTR-occur COP tsunami
'What occurred was a tsunami.'

b. *Tsunami ialah apa yang ber-laku.*
tsunami COP what COMP INTR-occur
'A tsunami is what occurred.'

c. *Yang selamat itu ialah yang ber-labuh.* (Canonical Copular Clause with Null NP)
LIG safe DIST COP LIG INTR-anchor
'The ones that are safe are the ones anchored.'

d. *Yang ber-labuh itu ialah yang selamat.*
LIG INTR-anchor DIST COP LIG safe
'The ones anchored are the ones that are safe.'

e. *Bukan, yang telah hanyut ialah rumah.* (Pseudocleft)
CNTR COMP PRF drift COP house
'No, what has drifted is a house.'

f. * *Bukan, rumah ialah yang telah hanyut.*
CNTR house COP COMP PRF drift
(No, a house is what has drifted.)

The inversibility of the DPs in canonical copular clauses suggest that the structure of the two constructions involve a PredP. As for the pseudocleft, its fixed order is possibly the effect of Criterial Freezing by Rizzi (2010), as mentioned in Section 5.2.5. The earlier focalisation of the clefted XP in the derivation of the pseudocleft has frozen it in its post-copular position, unlike the focus in canonical copular clauses.

5.3.3 Relativisation vs. (Pseudo)Clefting

If the cleft clause were a complex DP, it should be possible to see a difference in what introduces the relative clause, depending on whether the postulated null NP corresponds to a relativised argument or adjunct. Owing to the fact that *apa* (what) and *siapa* (who) cannot function as relative operators, a relative clause with a relativised DP argument is always introduced by relativiser *yang* without an overt relative operator, whereas one with a relativised adjunct is introduced by an overt relative operator without *yang*. As illustrated in the range of relative clauses in (335), the item introducing the clause is either an overt operator or relativiser *yang*, and never both at the same time, depending on the status of the relative operator as either an argument or adjunct of the predicate within the relative clause. Otherwise, both the relative operator and relativiser may be omitted, e.g. *orang Ø Ali bunuh semalam*… (the person Ø Ali killed yesterday…)

(335) a. *Pem-bunuh-an* {**apa* / *yang* } *ter-jadi* *semalam…* (Argument)
NMZ-kill-NMZ what REL NVOL-become yesterday
'The murder that happened yesterday…'

b. *Lelaki* {**siapa* / *yang* } *ter-bunuh* *semalam…*
man who REL NVOL-kill yesterday
'The man who was killed yesterday…'

c. *Waktu* { *bila* /**yang* } *Ali* *ter-bunuh* *semalam…* (Adjunct)
time when REL A. NVOL-kill yesterday
'The time when Ali was killed yesterday…'

d. *Tempat* { *di mana* /**yang* } *Ali* *ter-bunuh* *semalam…*
place LOC where REL A. NVOL-kill yesterday
'The place where Ali was killed yesterday…'

e. *Cara* {*bagaimana* /**yang* } *Ali* *ter-bunuh* *semalam…*
way how REL A. NVOL-kill yesterday
'The way how Ali was killed yesterday…'

f. *Sebab* {*kenapa* /**yang* } *Ali* *ter-bunuh* *semalam…*
reason why REL A. NVOL-kill yesterday
'The reason why Ali was killed yesterday…'

The following table summarises the facts regarding the choice of relative operators and relativiser *yang* in headed relative clauses:

Status	Operator	Operator	*Yang*
Argument	*Apa* (what)	✗	✓
	Siapa (who)	✗	✓
Adjunct	*Bila* (when)	✓	✗
	Di mana (LOC where)	✓	✗
	Bagaimana (how)	✓	✗
	Kenapa (why)	✓	✗

Table 29: Comparison between Argument and Adjunct Relative Operators

Provided the structures in Figure 23 and Figure 24, the subject of the copular clause could be a complex DP with a relativised adjunct. If the head NP remained null, the relative clauses should then resemble free relative clauses. Movement of the focus would then yield sentences such as the following examples:

(336) a. *Pada pukul tujuh*$_1$ *-lah* ~~*masa*~~ *bila Ali ter-bunuh semalam* Ø t_1.
at o'clock seven -FOC time when A. NVOL-kill yesterday COP
'7 O'CLOCK, when Ali was killed yesterday was.'

b. *Di sini*$_1$ *-lah* ~~*tempat*~~ *di mana Ali ter-bunuh semalam* Ø t_1.
LOC here -FOC place LOC where A. NVOL-kill yesterday COP
'HERE, where Ali was killed yesterday was.'

c. *Begini*$_1$ *-lah* ~~*cara*~~ *bagaimana Ali ter-bunuh semalam* Ø t_1.
like.this -FOC way how A. NVOL-kill yesterday COP
'LIKE THIS, how Ali was killed yesterday was.'

d. *Kerana ini*$_1$ *-lah* ~~*sebab*~~ *kenapa Ali ter-bunuh semalam* Ø t_1.
because PROX -FOC reason why A. NVOL-kill yesterday COP
'BECAUSE OF THIS, why Ali was killed yesterday was.'

Due to the adjunct status of the relativised NP, *yang* is not used. The absence of *yang* fails to derive a cleft construction in Malay, which is characteristically a focus construction with a focal argument separated from the rest of the clause by *yang*. Although the constructions above are legitimate and grammatical, they are nothing but canonical copular clauses with moved foci, and not cleft constructions. The formation of a cleft construction, whether it be a regular cleft or a pseudocleft, is impossible when the focus corresponds to an adjunct or some other non-DP constituent, due to the absence of *yang*. As shown in the examples below, *yang* is only used with foci that correspond to a DP argument:

(337) a. *Air yang Ali minum.* (Clefted DP Argument)
water COMP A. drink
'It was WATER that Ali drank.'

b. **Kepada Ali yang air di-beri.* (Clefted PP Argument)
to A. COMP water PASS-give
(It was TO ALI that water was given.)

c. **Hari itu yang Ali minum air.* (Clefted DP Adverbial)
yesterday DIST COMP A. drink water
(It was THAT DAY that Ali drank water.)

d. **Di rumah yang Ali minum air.* (Clefted PP Adverbial)
LOC home COMP A. drink water
(It was AT HOME that Ali drank water.)

By extension of pseudoclefts being derived from clefts, the same pattern is observed in pseudoclefts:

(338) a. *Yang Ali minum ialah air.* (Pseudoclefted DP Argument)
COMP A. drink COP water
'What Ali drank was WATER.'

b. **Yang air di-beri ialah kepada Ali.* (Pseudoclefted PP Argument)
COMP water PASS-give COP to A.
(To whom water was given was TO ALI.)

c. **Yang Ali minum air ialah hari itu.* (Pseudoclefted DP Adverbial)
COMP A. drink water COP day DIST
(When Ali drank water was THAT DAY.)

d. **Yang Ali minum air ialah di rumah.* (Pseudoclefted PP Adverbial)
COMP A. drink water COP LOC home
(Where Ali drank water was AT HOME.)

As for free relative clauses, there is no restriction as to which *wh*-phrase may head them. In other words, free relative clauses may be headed by either argument or adjunct *wh*-phrases, as shown below. Also notice that free relative clauses do not obey the doubly-filled COMP filter when *apa* and *siapa* are concerned, e.g. (339a-b), and that the *wh*-phrase is obligatory.

(339) a. *Kami bincang-kan *(apa) yang ter-jadi semalam.*
1.PL.EXCL discuss-APPL what COMP NVOL-become yesterday
'We discussed what happened yesterday.'

b. *Kami* *bincang-kan* *(*siapa*) *yang* *ter-bunuh* *semalam.*
1.PL.EXCL discuss-APPL who COMP NVOL-kill yesterday
'We discussed who was killed yesterday.'

c. *Kami* *bincang-kan* *(*bila*) *Ali* *ter-bunuh* *semalam.*
1.PL.EXCL discuss-APPL when A. NVOL-kill yesterday
'We discussed when Ali was killed yesterday.'

d. *Kami* *bincang-kan* *(*di mana*) *Ali* *ter-bunuh* *semalam.*
1.PL.EXCL discuss-APPL LOC where A. NVOL-kill yesterday
'We discussed where Ali was killed yesterday.'

e. *Kami* *bincang-kan* *(*bagaimana*) *Ali* *ter-bunuh* *semalam.*
1.PL.EXCL discuss-APPL how A. NVOL-kill yesterday
'We discussed how Ali was killed yesterday.'

f. *Kami* *bincang-kan* *(*kenapa*) *Ali* *ter-bunuh* *semalam.*
1.PL.EXCL discuss-APPL why A. NVOL-kill yesterday
'We discussed why Ali was killed yesterday.'

Given the non-compliance of argument free relative clauses to the doubly-filled COMP filter and the obligatory nature of the *wh*-phrase, the derivation of a cleft according to the analyses by Kader (1981) and Cole and Hermon (2000) using extracted adjuncts would yield sentences that are identical to examples (336), which are not cleft constructions. Neither is a cleft construction derived when *apa* or *siapa* head the free relative clause, as the *wh*-phrase intervenes between the focus and *yang*, as shown below:

(340) a. *Ali*$_1$ *-lah* *siapa* *yang* *ter-bunuh* *semalam* *Ø* t_1.
A. -FOC who COMP NVOL-kill yesterday COP
'ALI, who was killed yesterday was.'

b. *Se-orang* *penjenayah*$_1$ *-lah* *apa* *yang* *ter-bunuh* *semalam* *Ø* t_1.
one-CLF criminal -FOC what COMP NVOL-kill yesterday COP
'A CRIMINAL, what was killed yesterday was.'

Table 30 summarises the patterns observed of the different types of clauses. Apparently, the clausal subject of a pseudocleft neither patterns with a headed relative clause nor a free relative clause, but it does pattern with the cleft clause of a regular cleft, which confirms the close relation between the two constructions. Dissimilarly, the subject of canonical copular clauses with moved foci patterns with headed and (to a slightly lesser extent) free relative clauses.

Construction	Status	Operator	*Yang*
Free Relative Clause	Argument	✓	✓
	Adjunct	✓	✗
Headed Relative Clause	Argument	✗	✓
	Adjunct	✓	✗
Subject of Focused Copular Clause	Argument	✗	✓
	Adjunct	✓	✗
Cleft Clause of Regular Cleft	Argument	✗	✓
	Adjunct	-	-
Subject of Pseudocleft	Argument	✗	✓
	Adjunct	-	-

Table 30: Comparison of Wh-Operators and 'Yang' between Relative Clauses and Cleft Clauses

If the supposed null NP corresponds to a relativised adjunct, it should be ungrammatical for the relative clause to be introduced by *yang*. This prediction makes it untenable for one to analyse the cleft clause as a relative clause due to the impossibility of the focus to correspond to anything other than an argument DP. Thus, that headed and free relative clauses can be introduced by adjunct *wh*-phrases, banning the use of *yang*, makes it clear that the clausal constituent of a cleft construction is not a form of relative clause.

5.3.4 Optionality of the NP

NPs are only optionally omissible. As stated by Wok Awang (1981), "for every full relative there is an exact construction without the head" (p. 246). In other words, it is possible for the NP to be deleted, as it is equally possible for the NP to remain overt, as shown below:

(341) *Saya mahu ikan ini, manakala dia mahu* {*ikan*/ Ø} *yang itu.*
1.SG want fish PROX whereas 3.SG want fish N LIG DIST
'I want this fish, whereas he wants that one.'

Provided the structures in Figure 23 and Figure 24, one would expect that the gap preceding *yang* could surface as an overt NP, and this expectation is truly met, as shown below. In fact, a *wh*-phrase may also be used, signalling that the subject truly is a DP. Also notice the demonstrative that modifies the clausal subject, which further confirmd the DP status of the subject.

(342) a. { *Orang* / *siapa* } *yang tidur itu ialah Zul.* (Focus In-Situ)
person who REL sleep DIST COP Z.
'The person that is sleeping is Zul.'

b. *Zul*$_1$*-lah* { *orang* / *siapa* } *yang tidur itu* Ø t_1. (Focus Ex-Situ)
Z.-FOC person who REL sleep DIST COP
'Zul is the person that is sleeping.'

Since the null NPs in examples (342) can be overtly realised, the subject in the structures by Kader (1981) and Cole and Hermon (2000) are confirmed to be DPs, which do not carry a cleft interpretation. The same cannot be said of the cleft constructions below, as they become ungrammatical if an overt NP or *wh*-phrase is plugged into the same position.

(343) a. *Siapa -kah* {**orang* /**siapa* } *yang tidur?*
who -Q person who COMP sleep
'Who is it that is sleeping?'

b. *Zul -lah* {**orang* /**siapa* } *yang tidur.*
Z. -FOC person who } COMP sleep
'It is Zul that is sleeping.'

The absence of a NP supports the claim that the cleft clause of a genuine cleft is not a complex DP. Furthermore, the absence of the demonstrative exacerbates the ungrammaticality of the construction in examples (343). The grammaticality of the noun to surface is contingent on a definitising or specifying element, e.g. a demonstrative, as discussed in further detail in the following section.

5.3.5 Definiteness and Referentiality

Given that the subject in Figure 23 and Figure 24 occurs in what appears to be a specificational copular clause, it is expected to be a definite description. Otherwise, a specific, albeit indefinite, subject may be used, as suggested by Heycock and Kroch (2002). Therefore, the DP requires the use of a definitising element such as a demonstrative or the phrase "*salah se*-CLF", which allows a specific reading. For example, The ungrammaticality in (344c) stems from the lack of either element.

(344) a. { *Orang* / *siapa* } *yang mati* {*itu* / *ini* } *ialah Zul.* [+Def] [+Spec]
person who REL die DIST PROX COP Z.
'The/that/this (person) who died is Zul.'

b. *Salah se-orang yang mati ialah Zul.* [–Def] [+Spec]
one.of one-CLF REL die COP Z.
'One of the persons who died is Zul.'

c. *{*Orang* /?*siapa*} *yang mati ialah Zul.* [–Def] [–Spec][75]
person who REL die COP Z.
(A person who died is Zul.)

Like regular DPs, free relative clauses in Malay require the use of a definitising element, such as demonstrative *itu,* in order to be interpreted as a definite DP. Consider the following examples that contrast two free relative clauses in Malay in terms of referentiality:

(345) a. *Siapa yang telah mati akan masuk syurga.* (Non-Referential)
who COMP PRF die PROS enter heaven
'Those who have died will go to heaven.'

b. *Siapa yang telah mati {itu / ini} akan masuk syurga.* (Referential)
who COMP PRF die DIST PROX PROS enter heaven
'That/this/the (person) who has died will go to heaven.'

The obligatory demonstrative in the example below affirms the DP status of free relative clauses in Malay:

(346) *Kata=nya di-jumpai orang mati dan Ali kebumi-kan siapa yang mati* *(*itu*).
say=3 PASS-find person die and A. bury-APPL who COMP die DIST
'They say that there was found a dead person and Ali buried the person who died.'

In the copular clause in the second conjunct below, the subject free relative clause would be rendered ungrammatical without the demonstrative.

(347) *Kata=nya di-jumpai orang mati dan siapa yang mati* *(*itu*) *ialah Zul.*
say=3 PASS-find person die and who COMP die DIST COP Z.
'They say that there was found a dead person and the person who died was Zul.'

Compare (347), which is a canonical copular clause with a free relative clause subject, with (348), which is a pseudocleft. The absence of the demonstrative indicates that the clausal subject of the pseudocleft is not a DP.

[75] It appears that the *wh*-phrase in the free relative clause allows a specific reading. However, the specificity of the *wh*-phrase is reliant on the context in which it is used. Without a specific context, the use of a free relative clause without a definitising element is incoherent as it is interpreted as a non-specific indefinite.

(348) *Kata=nya di-jumpai orang mati tetapi sebenarnya yang di-jumpai ialah patung.*
say=3 PASS-find person die but actually COMP PASS-find COP doll
'They say that there was found a dead person but actually what was found was a doll.'

As for clefts, according to the proposals by Kader (1981) and Cole and Hermon (2000), the derivation involves the head of the relative clause being silent and movement of the post-copular constituent to the left periphery. Given that the indefinite DP subject in (344c) is ungrammatical, the ungrammaticality caused by the lack of a definitising or specifying element is predicted to persist (by virtue of reconstruction) after movement of the focus. However, without the head of the supposed relative clause and a definitising or specifying element, the examples below are unexpectedly grammatical, suggesting that the subject is not in fact a DP. In addition to that, notice that the interpretations of (349) are that of a pseudocleft and a cleft, rather than a focused copular clause with an indefinite subject. These should not be the actual interpretations if the pre-copular constituents truly were indefinite DPs.

(349) a. *Yang mati ialah Zul.*
COMP die COP Z.
'Who died is Zul.'

b. *Zul yang mati.*
Z. COMP die
'It is Zul that died.'

Consider (350c-d) in which what appear as clefts are derived from the canonical copular clauses in (350a-b) with definite subjects. Despite their grammaticality, they fail to obtain the intended cleft interpretation after movement has taken place. As shown in the translation in examples (350c-d), the interpretation is that of a focused copular clause, rather than a cleft:

(350) a. *Ø Yang mati* { *itu* / *ini* } ~~*ialah*~~ *siapa?* (Focus In-Situ)
N REL die DIST PROX COP who
'Who is the/that/this (person) that died?'

b. *Ø Yang mati* { *itu* / *ini* } *ialah Zul.*
N REL die DIST PROX COP Z.
'The/that/this (person) that died is Zul.'

c. *Siapa*$_1$ *-kah Ø yang mati* { *itu* / *ini* } ~~*ialah*~~ t_1? (Focus Ex-Situ)
who -Q N REL die DIST PROX COP
i. 'Who is the/that/this (person) that died?'
ii. #'Who is it that died?'

d. *Zul*$_1$ *-lah Ø yang mati* { *itu* / *ini* } ~~*ialah*~~ t_1.
Z. -FOC N REL die DIST PROX COP
i. 'Zul, the/that/this (person) that died is.'
ii. #'It is Zul that died.'

This difference in interpretation is due to the demonstrative. The subject of the copular clause is a definite description of a single entity that stands in a one-to-one relation with the focus, hence a singleton set. It is also referential, as it may be substituted with a personal pronoun. As such, it does not presuppose multiple entities, which is different from the presupposition carried by the cleft clause of a cleft, which evokes a set of multiple entities. Compare (350) with the following examples:

(351) a. *Siapa -kah yang mati?*
who -Q COMP die
'Who is it that died?'

b. *Zul -lah yang mati.*
Z. -FOC COMP die
'It is Zul who died.'

Without a demonstrative, examples (351) should be ungrammatical under the assumption that the cleft clause is a DP. On the contrary, they are grammatical. Most importantly, they correctly obtain the cleft interpretation in which the cleft clause presupposes the existence of a set of entities that is greater than one and the focus exhaustively negates all but one alternative, in accordance to proponents of alternative semantics such as Rooth (1985), Krifka (2008) and Neeleman and Vermeulen (2012).

The different definiteness requirements between (350) and (351) hints at the different syntactic categories of the *yang*-clause in the two constructions. The *yang*-clauses in examples (350) is undoubtedly part of a DP constituent as it requires a definitising element. The DP is referential as it refers to an actual person. On the other hand, the ones in (351) are not referential, but they presuppose that there exists a person *x* such that *x* died. Therefore, it should be possible to substitute the *yang*-clause in a canonical copular clause with a demonstrative pronoun, but not in a cleft.

Context: *Lihat. Ada orang mati*. (Look. There's a dead person.)

(352) a. *Siapa-kah yang mati itu?* → *Siapa-kah itu?* (Copular Clause)
who -Q REL die DIST → who -Q DIST
'Who is the (person) that died?' 'Who is that?

b. *Zul -lah yang mati itu.*
Z. -FOC REL die DIST → Z. -FOC DIST
'Zul, the (person) that died is.' ZUL, that is.

Context: *Lihat. Ada majlis pengebumian*. (Look. There's a funeral.)

c. *Siapa-kah yang mati?* → #*Siapa-kah itu?* (Cleft)
who -Q COMP die → who -Q DIST
'Who is it that died?' ('Who is that?)

d. *Zul -lah yang mati.*
Z. -FOC COMP die → #*Zul -lah itu.*
'It is Zul who died.' Z. -FOC DIST
(ZUL, that is.)

As shown above, the cleft clause of a genuine cleft is not referential as it cannot be substituted by a demonstrative pronoun. Although the resulting construction is grammatical, the interpretation has changed into that of a canonical copular clause. So, the cleft clause can be analysed to be a bare CP, a syntactic category that need not be modified by a definitising element.

(353) a. [$_{CP}$ *yang*…] COP [FOCUS] (Pseudocleft)

b. [$_{DP}$ *yang*…]$^{DEF+REF}$ COP [FOCUS] (Canonical Copular Clause)

This difference moreover calls for different syntactic structures. The structure of examples (350) most likely resembles the structures by Kader (1981) or Cole and Hermon (2000), as there is confirmed to be a DP, hence a canonical copular clause. On the other hand, examples (351) have the structure in Figure 27, as there is no complex DP that needs to be definite. Although they look very similar on the surface, they are essentially different constructions.

The difference in syntactic structure reflects a difference in semantics, as confirmed by the difference in their interpretation. The contexts in which canonical copular clauses are used differ from the context in which a genuine cleft is used. Consider the following interrogative copular clause:

(354) *Siapa-kah Ø yang mati* { *itu / ini* }?
who -Q N REL die DIST PROX
'Who is the/that/this (person) that died?'

The question above may be appropriately used in a situation whereby the cadaver in question is visible to the questioner. This is due to the presence of the demonstrative in the DP in the copular clause, which is referential. Depending on the cadavers that are directly visible to the questioner, the number of actual objects denoted by the definite DP is restricted. Therefore, the objects in the answer to the question must be equal to the objects that are visible to the questioner. For example, if the questioner sees one in front of her, the focus may only denote one dead person. It would be infelicitous to include more than one dead person in the answer.

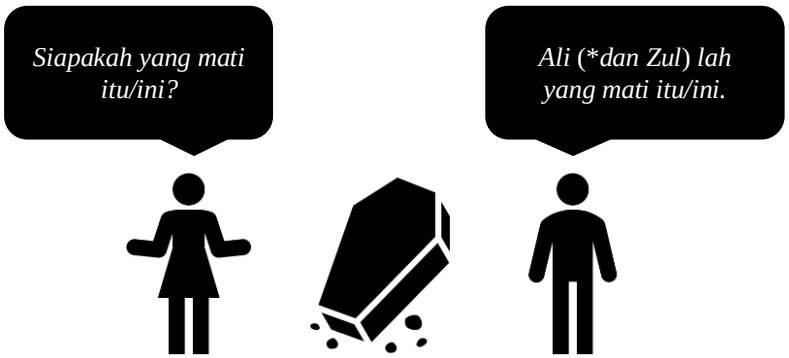


Figure 32: One Dead Person in View

(355) *Ali (*dan Zul) -lah Ø yang mati itu.*
A. and Z. -FOC N REL die DIST
'Ali (*and Zul), the/that/this (person) that died is.'

The use of a cleft, e.g. (356), is infelicitous in such a context. This infelicity is due to the absence of definiteness/referentiality carried by the demonstrative, in the visual context that requires all those notions. Because the cleft clause is not a referential DP, it fails to refer to the cadaver seen by the questioner. Furthermore, the presupposition of a cleft does not restrict the set to a single object. There is an indefinite number of objects in the set and the answer exhaustively identifies one or more objects, depending on context. Since the set is unrestricted, the answer can include any number of cadavers, which is not the goal of the exchange in Figure 32.

(356) *#Siapa-kah yang mati?*
who-Q COMP die
'Who is it that died?'

Therefore, if a cleft truly were a canonical copular clause, the cleft clause should be part of a complex DP that is definite and referential. On the contrary, that the cleft clause is not a constituent that is definite or referential verifies that a cleft is not derived from a canonical copular clause with a complex DP subject.

5.3.6 Licensing of NP-Ellipsis

Ellipsis does not simply happen to any NP. Although it is true that the deletion of a NP in Malay may strand modifiers and make the structure appear as the DPs in Figure 23 and Figure 24, there are rules that govern and license NP-ellipsis.

Consider examples (357a-g), in which optional NP-ellipsis has applied. Following ellipsis, the stranded modifier must be supported by *yang*, which scholars describe as a ligature that connects a NP to an attribute (Simin, 1988; van Minde, 2008; Verhaar, 1982).[76] As illustrated, ellipsis can apply to any of the NPs in (357), leaving behind *yang* and the modifiers following it. Verhaar (1982) states that "*yang* used replacively is invariably obligatory" (p.51).[77] Therefore, supportive *yang* is necessary to indicate that preceding the stranded modifiers is a NP that has been elided.

(357) a. (*Kucing*) *yang hitam* (Adjectival Attribute)
cat LIG black
'A black cat/one'

b. (*Kucing*) *yang peŋ-musnah* (Nominal Attribute)
cat LIG AG-destroy
'A destroyer cat/one'

[76] Without going into the specific details of the so-called ligature, I assume that this variety of *yang* is what Den Dikken (2006) calls a *linker*, which heads a functional projection that takes the attribute as its complement and the NP as its specifier: [$_{FP}$ [$_{NP}$ *kucing*] [$_{F'}$ *yang* [$_{AP}$ *hitam*]]]. In fact, the Tagalog cognates *ang* and *ng* are themselves commonly referred to as linkers, and they stand in a similar linking relationship between a noun and its attribute.

[77] Verhaar (1982) argues that *yang* is strictly a functional head, as is apparent in its functions within the CP layer and its lack of any meaningful semantic content. Therefore, it is likely for the NP to be elided and for *yang* to somehow be triggered to spell out, as opposed to the substitution of the NP by *yang* in the application of what he calls *replacive yang*, which at this point has become a misnomer. Therefore, the term *replacive yang* shall henceforth be called *supportive yang* and *yang-support*.

c. (*Kucing*) *yang seperti harimau* (Prepositional Attribute)
cat LIG like tiger
'A cat/one like a tiger'

d. (*Kucing*) *yang sedang ber-lari* (Relative Clause) [78]
cat LIG PROG INTR-run
'A cat/one that is running'

e. (*Kucing*) *yang ke-lima* (Cardinal Number)
cat LIG CARD-five
'The fifth cat/one'

f. (*Kucing*) *yang mana* (Interrogative)
cat LIG which
'Which cat/one'

g. (*Kucing*) *yang ini /itu* (Demonstrative)[79]
cat LIG PROX /DIST
'This/that cat/one'

Like most kinds of ellipsis phenomena, NP-ellipsis and the application of *yang*-support are licensed by the givenness of the NP to be deleted. Verhaar (1982) states that "the context would be such that the head has been mentioned before, or at least is situationally clearly presupposed… [The head] may be left out in its second occurrence on the strength of the first; or in both occurrences on the strength of a previous occurrence" (p. 51). This requirement entails that ellipsis and supportive *yang* are not possible in contexts in which the deleted NP is not presupposed.

[78] It appears that when the stranded modifier is a relative clause, the readily present relativiser *yang* can sufficiently act as supportive *yang*.

[79] As explained in Section 3.3.4, it impossible to derive a cleft or a relative clause from a non-predicational copular clause, entailing that the string "*yang* DEM" is not part of a cleft or relative clause.

(xxiv) a. *Siapa-kah* (**yang*) *se-benar=nya itu?*
who-Q COMP one-true=3 DIST
'Who actually is that?'

b. *Ali-lah* (**yang*) *se-benar=nya itu.*
A.-FOC COMP one-true=3 DIST
'ALI, that actually is.'

c. * *Orang* (*yang*) *se-benar=nya itu…*
person REL one-true=3 DIST
(The person who that actually is…)

On the other hand, the use of a cleft construction does not hinge on a given NP. In fact, there does not even need to be any meaningful nominal in the previous discourse that forms the presupposition of a cleft construction. To illustrate, neither the context nor the cleft in the example below includes any nominal other than the clefted DP and a presumed null expletive in the weather clause in the context, which is semantically vacuous.

Context: *Sedang hujan.* (It's raining.)

(358) a. *Bukan, ke-bocor-an paip yang sedang berlaku.* (Cleft)
CNTR NMZ-leak-NMZ pipe COMP PROG occur
'No, it is a pipe leakage that is occurring.'

b. *Bukan, yang sedang berlaku ialah ke-bocor-an paip.* (Pseudocleft)
CNTR COMP PROG occur COP NMZ-leak-NMZ pipe
'No, what is occurring is a pipe leakage.'

The examples above make a good case for the fact that the gap in a cleft clause is not traced to some DP from the preceding context or discourse, unlike the licensing of NP-ellipsis. Although cleft constructions and NP-ellipsis require certain presuppositions to license them, their presuppositional properties differ. NP-ellipsis is licensed by givenness of a NP, whereas cleft constructions are licensed by contrast and exhaustivity, such that a whole proposition within the discourse is contrasted. In other words, the presupposition in NP-ellipsis relates to a given NP, whereas that of a cleft construction relates to a given proposition. In fact, the subject of the canonical copular clause does not even need to be a clausal constituent to begin with. As shown below, what appears as a clausal subject is simply a DP that has undergone NP-ellipsis with the spell-out of *yang* to support the stranded modifier:

Context: *Minuman*$_1$ *kuning itu ialah jus epal.* (That yellow drink is apple juice.)

(359) ~~*Minum-an*~~$_1$ *yang putih itu ialah air kelapa.*
drink-NMZ LIG white DIST COP water coconut
'(The white one/the one that is white) is coconut water.'

Therefore, the gap in a cleft construction differs from the one in *yang*-support in that the constituent that corresponds to the gap of the cleft clause is sentence-internal (somewhere within the cleft construction itself), whereas the constituent that corresponds to the gap in *yang*-support is sentence-external (somewhere in the discourse).

| | | |

Context: *Ikan itu belum dimasak.* (That fish has not been cooked.)

(362) ~~*Ikan*~~ *yang ini telah di-masak.*
fish LIG PROX PRF PASS-cook
'This one has been cooked.'

Anything intervening *yang* and the contrastive element would cause ungrammaticality. In (363a), *yang* and the contrastive element are separated from each other by *berduri* (thorny), which should rightfully be elided as it is given. According to Merchant (2001), MaxElide demands that ellipsis be applied to as much material as possible. Examples (363b-c) are equally ungrammatical as they allow the given material *berduri* to remain overt. Only (363d) is grammatical as no material intervenes between *yang* and the contrastive element, and ellipsis has applied maximally.

Context: *Ikan berduri itu masih hidup.* (That thorny fish is still alive.)

(363) a. *~~*Ikan*~~ *yang ber-duri ini pun masih hidup.*
fish LIG INTR-thorn PROX also still alive
(This thorny one too is still alive.)

b. *~~*Ikan*~~ *ber-duri yang ini pun masih hidup.*
fish INTR-thorn LIG PROX also still alive
(This thorny one too is still alive.)

c. *~~*Ikan*~~ *yang ber-duri yang ini pun masih hidup.*
fish LIG INTR-thorn LIG PROX also still alive
(This thorny one too is still alive.)

d. ~~*Ikan*~~ ~~*ber-duri*~~ *yang ini pun masih hidup.*
fish INTR-thorn LIG PROX also still alive
'This one too is still alive.'

Consider the following examples, which correspond to the structures proposed by Kader (1981) and Cole and Hermon (2000). The contrastive element similarly corresponds to the material following *yang* in *yang* support, regardless of the word order.

Context: *Ikan yang ada duri ini ialah buntal.* (This fish that has thorns is the pufferfish.)

(364) a. *Bukan, buntal*$_1$ ~~*ikan*~~ *yang ada gigi arnab itu* t_1.
CNTR pufferfish fish REL have tooth rabbit DIST
'No, the pufferfish, the one that has rabbit teeth is.'

b. *Bukan, ~~ikan~~ yang ada gigi arnab itu ialah buntal.*
CNTR fish REL have tooth rabbit DIST COP pufferfish
'No, the one that has rabbit teeth is the pufferfish.'

The examples above differ from cleft constructions with respect to the placement of the contrastive material. The contrast in a cleft construction corresponds to the focus – the clefted constituent in a cleft and the post-copular constituent in a pseudocleft – as shown below. Thus, that the contrast in examples (364) pattern with the contrast in DPs that have undergone NP-ellipsis and *yang*-support and to the contrary of cleft constructions entails that the constituent headed by *yang* is also part of a complex DP.

Context: *Zul suka makan fugu.* (Zul likes to eat fugu.)

(365) a. *Bukan, Ali yang suka makan fugu.* (Cleft)
CNTR A. COMP like eat fugu
'No, it is ALI who likes to eat fugu.'

b. *Bukan, yang suka makan fugu ialah Ali.* (Pseudocleft)
CNTR COMP like eat fugu COP A.
'No, who likes to eat fugu is ALI.'

Additionally, although the use of supportive *yang* in the context evokes alternatives, its following statement does not necessarily need to be contrastive. For example, in (366), an alternative to *"*that has thorns*"* is evoked, e.g. "that has stripes/rabbit teeth/etc." but the statement does not convey contrastive information.

Context: *Ikan yang ada duri ini ialah buntal.* (This fish that has thorns is the pufferfish.)

(366) *Kerana ke-comel-an=nya, ~~ikan~~ yang ada duri ini ialah ke-gemar-an=ku.*
because NMZ-cute-NMZ=3 fish REL have thorn PROX COP NMZ-like-NMZ=3
'Because of its cuteness, this one that has thorns is my favourite.'

Similar to *yang*-support in DPs, free relative clauses may or may not carry contrastive information. However, unlike the optional nature of the contrastivity in *yang*-support, cleft constructions in Malay are inherently and obligatorily contrastive. To illustrate, the use of the pseudocleft in (367a) is infelicitous due to the lack of an appropriate contrastive context. The infelicity is caused by the lack of any alternatives

that could be contrasted by the pseudocleft in the context.[80] Conversely, the use of a free relative clause as the subject of the copular clause in (367b) is perfectly fine regardless of the context because it does not carry any contrastive information.

Context: *Lembu adalah sejenis mamalia.* (Cows are a type of mammal.)

(367) a. *#Yang lembu makan ialah rumput.* (Pseudocleft)
COMP cow eat COP grass
(What cows eat is grass.)

b. *Apa yang lembu makan ialah rumput.* (Canonical Copular Clause)
what COMP cow eat COP grass
'What cows eat is grass.'

This finding relates to the contrastivity of *wh*-ex-situ in Malay, whose *wh*-phrase is obligatorily clefted, as signalled by the use of *yang*. A *wh*-in-situ question may not be answered by a cleft or a pseudocleft, without the right context, which indicates that the two constructions are contrastive.[81] The context in the examples below includes a *wh*-in-situ question – *siapa* (who) is not clefted from the rest of the clause by *yang*. Therefore, the use of a cleft or a pseudocleft to answer the question is infelicitous.[82]

Context: *Siapa suka gula-gula?* (Who likes sweets?)

(368) a. *Saya suka gula-gula.* (Simplex Answer)
1.SG like sugar-RED
'I like sweets.'

[80] Also, notice that there is no difference in the form and meaning in the English translations of the examples above. Apparently, the two constructions are the same in English, which is why the same construction can be used in different contexts in English. However, they are separate constructions in Malay, and they have different interpretations.

[81] A *wh*-in-situ question may be answered by a clefted question or pseudocleft when the questioner is unaware of any alternatives to the focus, hence the use of *wh*-in-situ, but the questionee insists on an exhaustive interpretation, hence the use of a cleft construction. For example, a coherent clefted answer to the *wh*-in-situ question "*Siapa suka gula-gula*" (Who likes sweets?) could be "*Tiada siapa. Hanya X yang suka gula-gula*" (Nobody. It is only X who likes sweets).

[82] That is, unless the context includes a clefted question with *siapa* clefted by *yang*, i.e. *siapa yang suka gula-gula*? (who is it that likes sweets?), or the person answering the *wh*-in-situ question is admitting that he is the only one in a group of answerers that likes sweets, in which case he is contrasting himself out of the other people.

b. **Saya yang suka gula-gula.* (Clefted Answer)
1.SG COMP like sugar-RED
(It is ME who likes sweets.)

c. **Yang suka gula-gula ialah saya.* (Pseudoclefted Answer)
COMP like sugar-RED COP 1.SG
(Who likes sweets is ME.)

This difference in contrastivity sets cleft constructions apart from canonical copular clauses, of which the constituent immediately following supportive *yang* conveys contrastive information, as opposed to the contrast being the post-copular constituent in a pseudocleft or the clefted constituent in a cleft.

5.4 Summary

This chapter has provided an exposition of cleft constructions in Malay. The analysis proposed in this chapter is that clefts are derived via focalisation of the clefted XP, which spells out the characteristic *yang* element, which typifies clefts in Malay. This conclusion follows from the discovery of a covert copular clause above the visible part of the cleft, and the finding that the constituent comprising the clefted XP and the cleft clause occur low in the structure of the overall cleft construction. Further, clefts share a common derivation with pseudoclefts, which are derived in a fashion that is reminiscent of intraposition, in which the clausal constituent is moved to the subject position of the matrix copular clause.

I have argued that the derivation of clefts by Kader (1981) and Cole and Hermon (2000), who claim that clefts are derived from canonical copular clauses whose subject is underlyingly a complex DP that has undergone NP-ellipsis, differs from genuine cleft constructions in nontrivial ways. Based on the non-DP behaviour of the cleft clause, canonical copular clauses and cleft constructions are dissimilar in terms of definiteness, optionality of the NP, licensing, information structure and islandhood. It is only via NP-ellipsis of the complex DP subject that the canonical copular clause takes a deceivingly similar form as a cleft construction. The difference in semantics and syntactic behaviour between the two types cannot possibly be attributed to NP-ellipsis. On the contrary, it should be a reflection of different syntactic structures altogether. The cleft clause does not behave the same way a DP does, which allows us to arrive at the conclusion that the cleft clause is not a DP but a bare CP.

Chapter 6: The Diachrony of the Copulas

Our current understanding of the copulas mostly pertains to their synchrony. What we assume and believe about the history of the copulas is that *ialah* and *adalah* had evolved from their discernible morphological components, namely 3rd person *ia*, existential verb *ada*, and focus marker *lah*. This chapter addresses the gap in the literature about the historical side of nonverbal predication in Malay by examining the linguistic changes that have occurred in the history of the Malay language and how they have affected nonverbal predication to ultimately arrive at an understanding of how it was possible for the morphemes *ia*, *ada*, and *lah* to have undergone grammaticalisation and what linguistic circumstances during which historical stage of the language led to their grammaticalisation into copulas *ialah* and *adalah*.

I argue that *ia-lah* and *ada-lah* – which are actually combinations of 3rd person *ia*, dummy auxiliary verb *ada*, and comment marker *lah* (3.SG-COM and AUX-COM), as opposed to focus marker *lah*, as per Ajamiseba (1983) and Müller-Gotama (1995) – grammaticalised into copulas *ialah* and *adalah* via different pathways, which explains several facts about copular clauses in Modern Malay, such as why both copulas cannot host verum focus, why there are two different copulas, why copula *adalah* exhibits its atemporal behaviour, and why copula *ialah* selects 3rd person subjects. The relevant period for this development is the Classical Malay period (14th to 19th century), prior to which the *lah* morpheme is absent.

The structure of this chapter is as follows. First, I will discuss change in word order in the history of the Malay language from VSO to SVO, which played an essential role in the diachrony of the Malay copulas, as they grammaticalised during the transition from Classical Malay to Modern Malay. The transition from VSO to SVO was triggered by three major occurrences in the history of the language: first, an increase in the markedness of movement of the verb to the left periphery effected a decrease in sentences with verb-initial word order; next, an increase in the frequency of sentences in which the subject is topicalised with topic marker *pun* effected a rise in sentences with topic-initial word order; finally, the loss of topic marker *pun* allowed topic-initial sentences to be reanalysed as subject-initial.

Next, the grammaticalisation of the copulas is examined. It is shown that copula *adalah* was previously used to mark new information in both verbal and nonverbal clauses. Towards the completion of the change in word order from VSO to SVO during the transition from Classical Malay to Modern Malay circa the 19th century, the use of *ada-lah* as a marker of new information in non-copular clauses declined due to the decline of V-T-C movement, which was necessary for auxiliary *ada* to obtain the comment marker in the left periphery. In spite of that, the use of *ada-lah* in copular clauses continued to rise, which constitutes evidence for the specialisation of *ada-lah* in copular clauses and its subsequent grammaticalisation into a copula. As for copula *ialah*, its use in copular clauses only sparked towards the end of the Classical Malay period, during which the change in word order was nearing completion. *Ia-lah* in left-dislocation constructions underwent spec-to-head reanalysis in which the former pronominal element was reanalysed as the head of TP.

Finally, several properties of the copulas and of copular constructions in Modern Malay are re-examined from a diachronic perspective. The ways in which the copulas grammaticalised account for the phenomena observed of the copulas in the current stage of the language, namely the lack of verum focus on the copulas, the 3rd person subject restriction on *ialah*, the use of either copula in predicational vs. specificational copular clauses, and the atemporality of copula *adalah*, whilst historical evidence sheds light on the absence of relative clauses in clefts.

6.1 Change in Word Order

In the history of the Malay language, the change in word order from VSO to SVO can be said to be caused by three main occurrences. First, verb-fronting became less frequent and was dependent on the verb (or an auxiliary) being affixed by *lah*; second, topicalisation of the trigger with topic marker *pun* became very frequent; finally, the loss of *pun* led to the reanalysis of the topic as the subject.

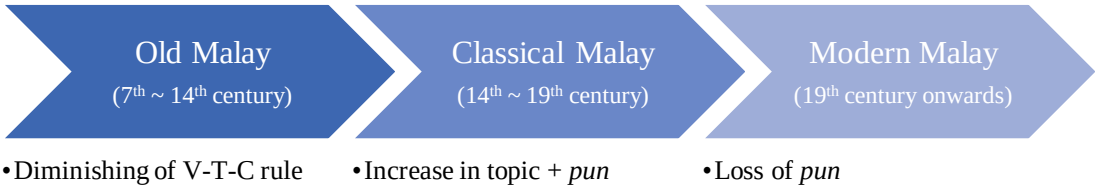


Figure 34: The Factors that Changed the Word Order of Malay

Old Malay was a mainly VSO language. Main clauses were most often verb-initial; however, conditional clauses always had the SVO order. As illustrated in the examples below, the conditional clause is SVO, whereas the main clause is VSO.

(369) a. *Kadāci kāmu tīda bhakti dy=āku ni-vunuḥ kāmu sumpaḥ.*
COND 2.PL NEG faithful PREP=1 PASS-kill 2.PL curse
'If you are not faithful to me, you will be killed (by) the curse.'
(Telaga Batu Inscription – 700 AD)

b. *Kadāci iya bhakti tatvārjjava diy āku… śanti muaḥ ka-vuatā=ña…*
COND 3.SG faithful loyal PREP 1.SG… bless OPT NMZ-do=3
'If he is faithful and loyal to me…, blessed be his efforts…'
(Kota Kapur Inscription – 686 AD)

The position of the verb and other auxiliaries or negators at the front of the clause can be analysed as V-T-C movement. The reason why conditional clauses had the SVO order was due to *kadāci* having filled C^0 in the clause periphery, blocking head movement of the verb or auxiliaries to C^0. Main clauses, on the other hand, had a VSO order as there was no element occupying C^0 to block movement. In the following examples, the verb has undergone inversion with the subject of the main clause, yielding the VSO order.

(370) a. *Tālu muaḥ ya dṅan gotra-santānā=ña.*
chastise OPT 3.SG with clan-family=3
'May he be chastised with his clan and family.'
(Kota Kapur inscription – 686 AD)

b. *Tmu muaḥ ya āhāra dṅan āir ni-minuṃ-ña.*
find OPT 3.SG food and water PASS-drink=3
'May he find food and water to be drunk by him.'
(Talang Tuwo Inscription – 684 AD)

c. *Tida mar-vuat kāmu doṣa ini.*
NEG INTR-do 2.PL sin PROX
'You do not commit this sin.'
(Telaga Batu Inscription – 700 AD)

Nevertheless, by the 9th century, the V-T-C rule had already shown signs of becoming diminished, as independent clauses with SVO word order are attested. As shown below, the subject of the clause is preverbal:

(371) a. *Iya maka-jādi pratiṣṭa di hyang haji tarkalaut sang hyang Wintang prasāda.*
3.SG CAUS-become statue LOC HON king north S. palace
'He made a statue of the king north of the palace of Sang Hyang Wintang.''
(Gandasuli inscription – 832 AD)

b. *Dayang Aṅkatan… di-bari waradāna wi śuddhapātra ulih Sang Pamgat Senāpati.*
D. PASS-give favour PREP acquittal by S.
'Dayang Angkatan... was given a favour of acquittal by Sang Pamgat Senāpati.'
(Laguna Copper-Plate inscription – 900 AD)

In Classical Malay, the frequency of subject-initial word order had increased, especially in subordinate clauses, but the basic word order of the language was VSO. It is reported by Cumming (1988) that verb-initial word order was the majority in her sample of 273 clauses by just a slight margin (51.28%), which entails that SVO order was already quite established at the time.[83]

	ST	AT	PT	All
P > T	72	0	68	140
T > P	86	31	16	133
Total	158	31	84	273
%P > T	45.57%	0%	80.95%	51.28%

Table 31: Classical Malay Constituent Order (Cumming, 1988)[84]

Even though Classical Malay had the same verb-fronting rule as Old Malay, it more visibly marked movement of the verb by the affixation of *lah*, unlike in Old Malay. This means that verb-fronting became marked in Classical Malay such that, whenever a verb became the host of *lah*, it must be fronted. This is attested by Cumming (1988) as so called *lah*-clauses are always predicate-initial, unless the trigger has been topicalised with topic marker *pun*, as illustrated, in (372) with the verb *pergi* (go).

[83] The 273 clauses are *pun*-less clauses, i.e. clauses that do not include a trigger that has been topicalised with topic marker *pun*.

[84] Abbreviations:
ST: Subject Trigger (subject of intransitive verb)
AT: Actor Trigger (subject of active transitive clauses with or without active verbal prefix *meŋ-*)
PT: Patient Trigger (subject of passive clauses with passive verbal prefix *di-*)
P > T: Predicate > Trigger (predicate-initial)
T > P: Trigger > Predicate (trigger-initial)

(372) a. *Maka Fatimah pergi ke pintu.* (SV)
DISC F. go to door
'Fatimah went to the door.'

b. *Maka pergi-lah orang ke-dua=nya itu.* (VS)
DISC go-COM person NMZ-two=3 DIST
'The two people went.'

c. *Maka Abu Hurairah pun pergi-lah kepada Zainab.* (TopV)
DISC A. TOP go-COM to Z.
Abu Hurarirah went to Zainab.'
(Hikayat Muhammad Hanafiah – 14th century AD)

Note that *lah* in examples (372b-c) does not carry a verum focal interpretation. It attaches to the verb and commonly co-occurs with elements that indicate temporal relations and sequencing of events, such as *maka* and adverbial clauses introduced by *setelah* (after), as stated by Cumming (1988).[85] According to Hopper (1979) and Lewis (1947), "-*lah* on the verb highlights and foregrounds the event, gives it special prominence in the narrative, and announces it as one of a series of actions" (Hopper, 1979, p. 227). Therefore, I follow the characterisation of *lah* as a comment marker, as is also made by Ajamiseba (1983) and Müller-Gotama (1995), who argue that it marks new information. Additionally, the status of *lah* as a marker of new information is confirmed by Ajamiseba (1983) as *ada-lah* and *ada-pun* differ in being able to introduce a new or given DP in existential sentences. The use of definite *itu* is not compatible with *ada-lah*; rather, indefinite *se-orang* (one-CLF) must be used, as shown below:

(373) a. *Ada-pun raja di Kota Maligai itu nama=nya Paya Tu Kerub Mahajana.*
AUX-TOP king in town M. DIST name=3 P.
'As for the king in Kota Maligai (and) his name was Paya Tu Kerub Mahajana.'

b. **Ada-lah raja di Kota Maligai itu nama=nya Paya Tu Kerub Mahajana.*
EXIST-COM king in town M. DIST name=3 P.
(There was a king in Kota Maligai (and) his name was Paya Tu Kerub Mahajana.)

[85] According to Müller-Gotama (1996), *maka* is a so-called "punctuation word" that marks the boundary of a clause or sentence. Due to the oral tradition of reciting the *hikayat* (story) to an audience, different punctuation words are used. The convention is to gloss *maka* as *then* but not to include its meaning in the translation; however, I will gloss it as a discourse marker (DISC). The use of *maka* to sequence events has survived in Modern Malay. It is used as a discourse marker, mainly in formal Malay, akin to English *therefore.*

c. *Ada-lah se-orang raja di Kota Maligai nama=nya Paya Tu Kerub Mahajana.*
EXIST-COM one-CLF king in town M. name=3 P.
'There was a king in Kota Maligai (and) his name was Paya Tu Kerub Mahajana.'
(Ajamiseba, 1983, p. 59)

Given that verb-initial order in both Old Malay and Classical Malay was obtained via V-T-C movement, the change to SVO order in Modern Malay was promoted by this rule becoming marked in Classical Malay.

Whilst V-T-C movement was diminishing, the use of topical structures in which the trigger was fronted with topic marker *pun* was on the rise. The frequent occurrence of *pun* is reported by Cumming (1988) with 42% of clauses with overt triggers sporting *pun* in her sample of 250 clauses with triggers. The markedness of V-T-C movement sparked a decline in verb-initial sentences, whereas frequent topicalisation with *pun* caused an increase in the frequency of SVO sentences. Although these two phenomena are unrelated, they both effected a rise in the frequency of clauses with preverbal subjects. Together they form the *pun-lah* construction, in which the constituent modified by *pun* is the topic, and the one modified by *lah* is the comment: "Classical Malay sentences are organized around a theme constituent marked by *pun*, which expresses old or given information, and a rhematic constituent marked by *lah*, which represents new information" (Müller-Gotama, 1996, p. 90).[86] However, given the existence of unmarked SVO sentences, neither *pun* nor *lah* was obligatory. Additionally, the use of either element was independent of each other, as shown below:

[86] Several authors have claimed that *lah* has other functions as well, as shown in the table below:

Function	Reference
Eventive marker; Inchoative aspect	(Cumming, 1995)
Perfective aspect	(Hopper, 1979)
Comment marker	(Müller-Gotama, 1996)

Hopper (1979) states "when the particle *-lah* is affixed to the verb, it denotes that the action or event of the clause is one of the main points of the narratives. Such verbs are almost always perfective." (p. 227).

(xxv) *…di-suroh-kan=nya bongkar sauh lalu be-layar. Maka di-layar-kan=nya-lah kechi itu…*
…PT-order-TRANS=3 raise anchor then sail and PT-sail-TRANS=3-LAH ketch DIST
'…he ordered the anchor to be weighed, and we sailed. He sailed the ketch…'
(Hopper, 1979, p. 227)

(374) a. *Kuda Umar Sa'd Maisum pantas ber-jalan.* (Unmarked)
horse U. fast INTR-walk
'Umar Sa'd Maisum's horse walked fast.'

b. *Maka Ibrahim Astar pun ber-jalan dengan anak=nya.* (*Pun*)
DISC I. TOP INTR-walk WITH child=3
'Ibrahim Astar walked with his child.'

c. *Maka ber-jalan-lah ia ke benua Habsah.* (*Lah*)
DISC INTR-walk-COM 3.SG to country H.
'He walked to the country of Habsah.'

d. *Maka ia pun ber-jalan-lah kepada Marwan.* (*Pun-Lah*)
DISC 3.SG TOP INTR-walk-COM to M.
'He walked to Marwan.'
(Hikayat Muhammad Hanafiah – 14th century AD)

In the *pun-lah* construction, although the verb has inverted with the subject, subsequent topicalisation of the subject makes it seem as though it had not moved at all. For example, (374c) shows a clause in which the verb is affixed by *lah* and moved to the front of the clause, whereas (374d) shows the same verb occurring in second position, due to the topicalisation of the subject.

This combination of verb-fronting and topicalisation is reminiscent of the V2 order that is common in many Germanic languages. Although Classical Malay was not a V2 language, the *pun-lah* construction produced this word order. Apparently, the difference between the Germanic languages and Classical Malay is that, whilst any phrasal constituent could in principle prepose in the Germanic languages, the topicalised constituent in Classical Malay most usually corresponded to the trigger of the clause.

The structure of (374d) is given below. I assume that *lah* is a single head in FocP that gets interpreted as either the focus or comment marker, depending on word order and the presence of topical material preceding it (see Section 6.2.2 for a discussion on the non-focal, comment-marking property of *ia-lah*). Also, as an optional discourse marker, *maka* is analysed to be an adverbial that adjoins to the clause, as opposed to one that is merged low in the clause and moved to the specifier of a left-peripheral projection. It corresponds to other clause-initial discourse markers, such as *manakala* (whereas), *walaupun* (although), *namun* (however), etc.

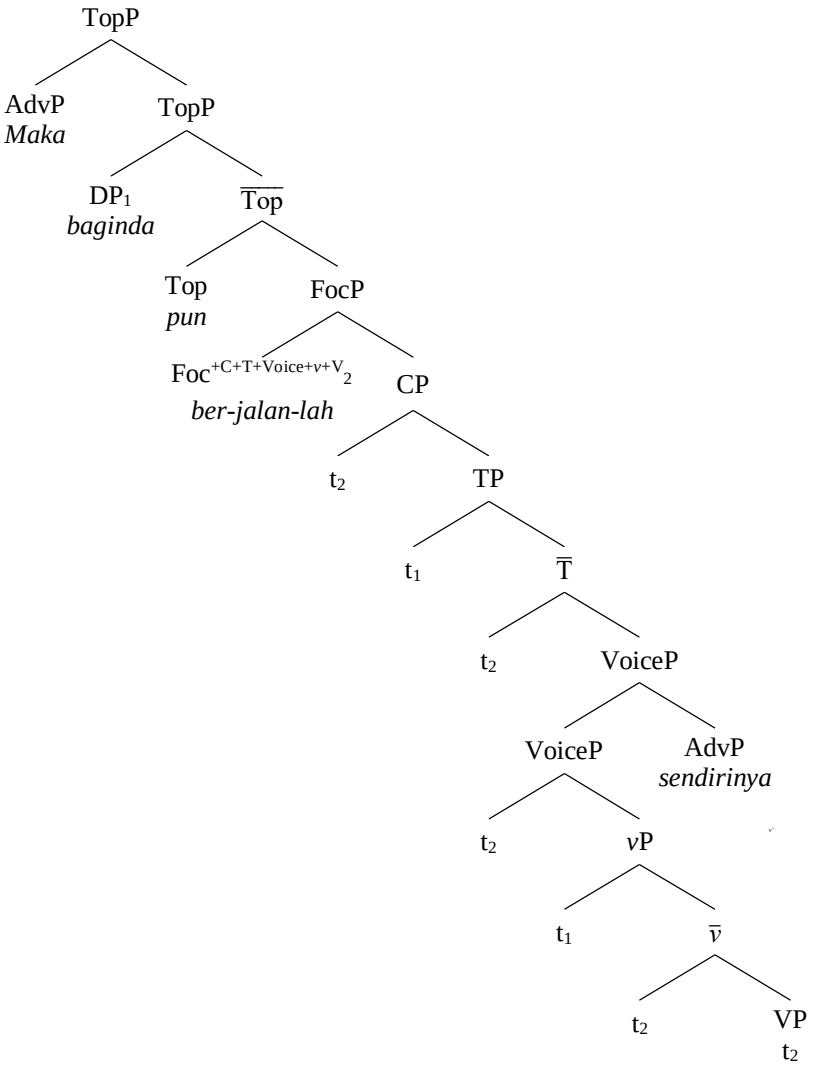


Figure 35: The Structure of the 'Pun-Lah' Construction

Finally, the loss of *pun* in Modern Malay (circa the 19th century), which previously marked preposed topics, also paved the way for the reanalysis of the clausal structure of Malay, changing the basic word order from VSO to SVO, as argued by Cumming (1988). This trajectory is in line with Bickerton and Givón (1976), who argue that "all a language has to do in order to make the VS-SV change is to reanalyse the marked topicality order as unmarked" (p. 30). Furthermore, according to Cumming (1988), the loss of *pun* in Modern Indonesian "effects the markedness reversal by which T>P clauses become (morphosyntactically) unmarked and P>T clauses (by virtue of their relative rarity) become marked" (p. 199). The decline in verb-initial clauses naturally had the effect of giving way to subject-initial clauses to become more common in the language.

In unmarked SVO sentences at the time, there is reason to believe that the preverbal subject is a topicalised element, at least in some cases, despite the absence of topic marker *pun*. There are found to be unmarked SVO clauses in which what appears as a resumptive clitic is coreferential with the preverbal subject. (375a) shows that the preverbal topic has moved from within the DP and is resumed by the enclitic to repair the violation caused by the failure of pied-piping the whole possessive DP. The same phenomenon is observed in (375b), but in an unmarked SVO clause. This finding ultimately suggests that the loss of *pun* had already begun during the Classical Malay period, starting with *pun* becoming optional in topical structures.

(375) a. *Maka ia*$_1$ *pun ada-lah suka sedikit hati=nya*$_1$.
DISC 3.SG TOP AUX-COM happy a.bit heart=3
'As for him, his heart was slightly happy.'

b. *Ia*$_1$ *terlalu sukacita hati=nya*$_1$ *akan suami=nya itu datang.*
3.SG very happy heart=3 PREP husband=3 DIST come
'She was very happy of her husband's coming.'
(Hikayat Sang Boma – 17th century AD)

Nowadays, *pun* is seldom used as a topic marker. It is now commonly used as an adverb with several different meanings, as shown in example (376).

(376) *Ali suka Ani. Abu pun suka Ani. Tetapi Ani tak suka mereka pun.*
A. like A. A. also like A. but A. NEG like 3.PL even
'Ali likes Ani. Abu also likes Ani. But Ani doesn't like them even.'

Meanwhile, *lah* is still used as a focus and comment marker on both heads and phrasal constituents of all categories to assign presentational, verum, and contrastive focus. When it attaches to a head, it may encode presentational or verum focus, whereas when it attaches to a phrase in clause-initial position, it only encodes contrastive focus.

(377) a. *Ali suka-lah Ani.*
A. like-COM/FOC A.
i. 'Ali likes Ani. (Presentational Focus)
ii. 'Ali DOES like Ani.' (Verum Focus)

b. *Ali-lah yang suka Ani.*
A.-FOC COMP like A.
'It is ALI who likes Ani.' (Contrastive Focus)

To summarise, the factors that led to the change in word order in Malay were observable since the period of Old Malay, circa 900 AD. The diminishing of the V-T-C rule and its increasing markedness caused verb-initial sentences to decline, whereas the increase in topicalisation with *pun* caused topic-initial sentences to rise. After that, the subsequent loss of *pun* neutralised the topic-subject distinction and caused sentence-initial topics to be reanalysed as the canonical subject. They all conspired to ultimately bring about change in the word order of the language from VSO to SVO.

6.2 Copular Clauses throughout the History of Malay

By surveying copular clauses at different stages of Malay, it is apparent that copulas *ialah* and *adalah* emerged quite late in the history of the language, around the time of Classical Malay. This is seen in the absence of the two items in copular clauses in Old Malay. As shown in examples (378a-d), copular clauses in Old Malay did not make use of overt copulas, regardless of whether they were predicational or otherwise. Presumably, a zero copula was used in Old Malay, which continues to date in Modern Malay. If translated into Modern Malay, as in (378e-h), overt copulas would be used optionally, which indicates the choice between overt and zero copulas.

(378) a. *Bhadravatī namā=ṇḍa aya=ṇḍa.* (Old Malay)
B. name=3 mother=3
'Bhadravati is the name of his mother.'
(Sojomerto inscription – 7th century AD)

b. *Svāmi-kāryya ka-dakṣā=ku.*
master-task NMZ-skilled=3
'The master's task is my expertise.'
(Mañjuśrīgṛha inscription – 793 AD)

c. *Adi=ṇḍa Ḍang Karayān laki Busu Tarbba nama=ṇḍa.*
sibling=3 HON HON male B. name=3
'The younger sibling of the prince, Busu Tarbba is his name.'
(Gandasuli inscription – 832 AD)

d. *Ini pa-ḍehā=nda Hawang Payangṅān.*
PROX NMZ-body=3 H.
'These are the corporeal remains of Hawang Payangnan.'
(Bukateja inscription – 840 AD)

e. *Bhadravati* (*adalah*) *nama ibunda=nya.* (Modern Malay)
B. COP name mother=3
'Bhadravati is the name of his mother.'

f. *Karya tuan* (*adalah*) *ke-pakar-an=ku.*
work master COP NMZ-expert-NMZ=3
'The master's work is my expertise.'

g. *Adiṇḍa Putera Busu Tarbba* (*adalah*) *nama=nya.*
sibling prince B. COP name=3
'The younger sibling of the prince, Busu Tarbba is his name.'

h. *Ini* (*adalah*) *jenazah Hawang Payangnan.*
PROX COP remains H.
'These are the remains of Hawang Payangnan.'

Although *ia* and *ada* are attested in Old Malay inscriptions, e.g. the Telaga Batu and Laguna Copperplate inscriptions, *lah* had not yet entered the language. Evidently, it is absent from the Tanjung Tanah manuscript (Old Malay) and the Terengganu inscription stone (Classical Malay), both from the 14th century. However, it is found in Hikayat Raja Pasai (Classical Malay), also from the 14th century. The absence of this crucial morpheme in Old Malay entails that the copulas could not have developed yet, since *lah* is one of the components in the etymology of both copulas.

The 13th or 14th century saw the advent of Islam in Malaya, as documented on the Terengganu inscription stone, and the introduction of Arabic influences such as the Jawi script. It can be considered the transition period between Old Malay and Classical Malay. Although the language at this stage was amid an influx of Arabic and Persian vocabulary, which presumably could have sparked the emergence or borrowing of some sort of copula, copular clauses in written materials at the time lacked any sort of visible copula.

(379) a. *Jika tetua bujang danda=nya lima tahil.*
COND mature single punish=3 five tael
'As for mature singles, the punishment is five taels.'
(Terengganu inscription stone – 1303 AD)

b. *Danda=nya sa-tahil sa-paha.*
punish=3 one-tael one-paha
'The punishment is one tael (and) one *paha*.'
(Tanjung Tanah manuscript – 14th century AD)

Towards the end of the 14th century, the morpheme *lah* had entered the language and it is observed to have been in frequent use on both heads and phrasal constituents. However, what was to become copulas *ialah* and *adalah* in Modern Malay did not serve the purpose of a copula. *Ada-lah* simply functioned as a marker of the topic-comment boundary of the clause as it was an auxiliary affixed by comment marker *lah*. Recall from examples (377) that *lah* encodes presentational focus and verum focus when it attaches to a head. According to Omar (2008), *ada-lah* was not used as a verb in the predicate of a sentence in Classical Malay but as a discursive element that "opens" a statement. Although *ada-lah* may be used in copular clauses, it is more accurately characterised as an auxiliary marking new information within a clause (Ajamiseba, 1983; Cumming, 1988; Müller-Gotama, 1996), as *ada-lah* was commonly used in verbal clauses as well.

(380) a. *Ada-lah emas itu di-per-oleh=nya terlalu banyak.* (Verbal)
AUX-COM gold DIST PASS-CAUS-obtain=3 SUP much
Too much gold was obtained by him.'

b. *Ada-lah dalam lubuk ini merah seperti api.* (Nonverbal)
AUX-COM in depth PROX red like fire
'The inside of the depths was red like fire.'
(Hikayat Raja Pasai – 14th century AD)

As for *ia-lah*, its function was twofold. It mostly served as a focal referential pronoun, given that it was a phrasal constituent. For example, the excerpt in (381) shows the use of *ia-lah* as a focused pronoun in both verbal and nonverbal clauses – the first clause is a nonverbal clause, whereas the second one is a verbal clause. When in a verbal clause, *ia-lah* may be clefted but, when in a nonverbal clause, it may not be clefted despite being a DP argument affixed by focus *lah*. *Ia-lah* being the only referential constituent in (381) is evidence that it functions as the subject. Additionally, the fact that *ia-lah* is clefted with *yang* in the verbal clause indicates that *ia-lah* is an argument of the verb, since only DP arguments may be clefted, as argued in Section 1.5.13.

(381) *Ia-lah raja yang pertama… Ia-lah yang ber-gelar Sultan Mahmud Syah…*
3.SG-FOC king REL first 3.SG-FOC COMP INTR-title sultan M.
'HE is the first king… It is HE who is entitled Sultan Mahmud Syah…'
(Undang-Undang Melaka – 15th Century AD)

Interestingly, under certain circumstances, it also marked the comment boundary of the clause. The *lah* morpheme signalled new information, as opposed to contrastive focus, despite the status of the host as a referential phrasal constituent. The condition for it to be used as such was such that it had to be coreferential with a left-dislocated topic. To illustrate, the following example shows the use of *ia-lah* in a topic-comment construction. Notice that it lacks a focal interpretation.

(382) *Syahdan akan baginda*$_1$ *ia*$_1$*-lah se-orang raja per-tapa-an.*
DISC PREP 3.SG.HON 3.SG-COM one-CLF king NMZ-ascetic-NMZ
'Regarding his Majesty, he was a king of asceticism.'
(Hikayat Sang Boma – 17th Century AD)

It is also found in verbal clauses, as shown in the following example. The lack of a focal interpretation and especially the absence of cleft *yang* is indicative of the non-focal property of *lah* in such sentences. It is this comment-marking form of *ia-lah* that developed into copula *ialah*, as opposed to the focused pronoun.

(383) *Ada-pun akan Batara Mahawisnu itu*$_1$ *ia*$_1$*-lah meŋ-jadi Batara Krisna.*
EXIST-TOP PREP B. DIST 3.SG-COM ACT-become B.
'Regarding Batara Mahawisnu, he became Batara Krisna.'
(Hikayat Sang Boma – 17th Century AD)

Seeing that the morpheme *lah* had only emerged in Classical Malay, it is not surprising that *ia-lah* and *ada-lah* were interpreted compositionally, rather than as abstract grammatical linking morphemes. The non-copular use of the two items continued in materials throughout the Classical Malay period.

(384) a. *Ada-lah ia ber-main di tepi laut itu.*
AUX-COM 3.SG INTR-play LOC by sea DIST
'He was playing by the sea.'

b. *Ia-lah yang meŋ-panah musuh.*
3.SG-FOC COMP ACT-shoot enemy
'It is HE who shoots the enemies.'
(Hikayat Pandawa Lima – 16th century AD)

c. *Ada-lah aku dengar ada se-orang tuan puteri.*
AUX-COM 1.SG hear EXIST one-CLF HON princess
'I heard that there was a princess.'

d. *Ia-lah yang meŋ-perintah negeri Pattani pada waktu itu.*
3.SG-FOC COMP ACT-govern state P. at time DIST
'It was HE who governed the state of Pattani at the time.'
(Hikayat Seri Kelantan – 20th century AD)

As for the copular usage of *ialah* and *adalah*, it is towards the end of the Classical Malay period that they could clearly be identified as copulas, as opposed to the compositional forms *ia-lah* and *ada-lah*. Following the change in word order, the two elements no longer had to move to clause-initial position despite being affixed by the *lah* morpheme. This resistance to movement denotes that they are somehow disjunct from other *lah*-marked elements, which commonly require to move to the front of the clause. Therefore, *adalah* and *ialah* in examples (385) are clearly copulas, instead of a comment-marking auxiliary and pronoun, as they occur in clause-medial position, despite having been affixed by *lah*, which should have triggered them to move to clause-initial position.

(385) a. *Kendara-an=nya Merpati Perak adalah ter-lebih tangkas.*
ride-NMZ=3 M. COP SUP-more agile
'The vehicle of Merpati Perak is more agile.'
(Hikayat Merpati Mas dan Merpati Perak – 19th century AD)

b. *Mereka rasa bahawa Kapitan ialah bapak=nya.*
3.PL feel COMP K. COP father=3
'They feel that Kapitan is his father.'
(Hikayat Kerajaan Sikka – 20th century AD)

As I will show in the following subsections, *ada-lah* and *ia-lah* followed different paths towards their development into copulas. Although the former was already very productive in copular clauses during the Classical Malay period, the diminishing of V-T-C movement cemented its role as a proper copula in clause-medial position, rather than a marker of new information in the left periphery. The use of *adalah* as a copula was highlighted by the decline in its use to mark the comment of a clause due to the completion of the change in word order from VSO to SVO. As for the latter, it only became used as a copula following the loss of topical marker *pun*, when the topic was reanalysed as the canonical subject during the same change in word order, after which *ia* stopped being used as a pronoun altogether.

Although each of the copulas followed different paths, their grammaticalisation nonetheless coincided in one way or another with change in word order. This is hinted in examples (385), in which *ialah* and *adalah* intervened between the subject and the nonverbal predicate, rather than occurring in front of the clause, which was the typical case in Classical Malay due to fronting by *lah*. To summarise, the following figure shows the stages of the Malay language in time and the functions of *ialah* and *adalah* during each stage.

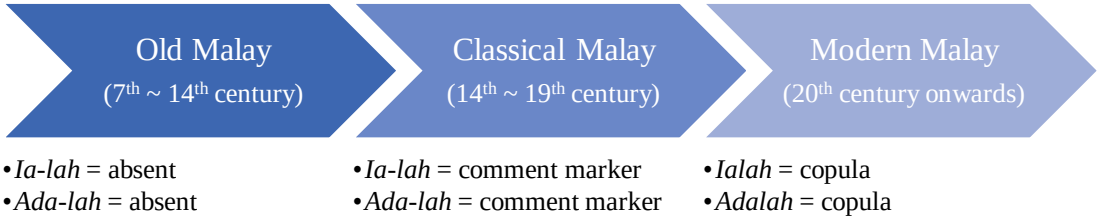


Figure 36: The Functions of 'Ialah' and 'Adalah' at Different Stages of Malay

6.2.1 The Grammaticalisation of *Adalah*

Copula *adalah* can be said to have gone through quite a simple grammaticalisation path. It evolved from dummy auxiliary *ada* affixed by comment marker *lah*, whose function was to mark new information in the front of the clause. According to Hopper (1979), *lah* was also used as a perfective aspectual marker. Therefore, verbs marked by *lah* carried an eventive interpretation. It is observed by Cumming (1988) that, as a marker of perfective aspect, *lah* induced an inchoative effect when it is affixed to stative verbs, which are interpreted to be eventive, rather than durative. This effect is also observed with predicative adjectives. To illustrate, examples (386) compare the durative reading when *ada-lah* is used and the inchoative reading when *lah* directly attaches to the predicate, which in this case is *suka* (happy).

(386) a. *Ia pun ada-lah suka sedikit hati=nya.* (Durative)
3.SG TOP AUX-COM happy a.bit heart=3
'As for him, his heart was slightly happy.'
(Hikayat Sang Boma – 17th century AD)

b. *Maka baginda pun suka-lah hati=nya.* (Inchoative)
DISC 3.SG.HON TOP happy-COM heart=3
'His Majesty's heart became happy.'
(Hikayat Putera Jaya Pati – 18th century AD)

Ada-lah was commonly used with intransitive predicates in cases in which new information was to be signalled by comment marker *lah* but the predicate itself must not move to prevent an inchoative or contrastive reading. In syntactic terms, another way of interpreting this use of *ada-lah* is such that the merging of *lah* in the left periphery left no choice for any constituent to move to it to host it. Neither any head nor any phrasal constituent could host *lah*, lest the head be interpreted as an inchoative event, or the phrase be interpreted as a contrastive focus. Therefore, as a means of repairing the stray affix violation, *ada* is merged to host it, as in other cases of the use of *ada* as a dummy auxiliary.

This repair strategy is also seen in verbal clauses in which the verb itself could not move to the left periphery to host *lah*, particularly verbs affixed by active voice marker *meŋ-*. Cumming (1988) notes "AT (Agent Trigger) clauses, which are never verb-initial, also don't occur with lah, even when there is a 'preverbal trigger' marked with pun". In order to avoid a violation of the stray affix filter, *ada* is merged to host *lah* in clauses with verbs in the active voice, as shown below:

(387) a. *Ada-lah kakanda meŋ-bawa dia…*
AUX-COM 1.SG ACT-bring 3.SG
'I brought him…"

b. **Meŋ-bawa-lah kakanda dia…*
ACT-bring 1.SG 3.SG
(I brought him…)
(Hikayat Sang Boma – 17th century AD)

Seeing that *ada-lah* was used in both verbal and nonverbal clauses, one must ask how the copula came about. Thus, a corpus analysis is done for several texts in Classical Malay spanning the 14th to the 20th century. The number of occurrences of auxiliary *ada-lah* in verbal and nonverbal clauses is presented in Table 32.

To clarify, only instances of auxiliary *ada-lah* are included in this analysis, i.e. the use of *ada-lah* as AUX-COM in nonverbal clauses (regardless of its position in the clause), as well as in verbal clauses in combination with a main verb, as shown in (388). The use of *ada-lah* in possessive/existential/locative clauses as *have*/EXIST-COM is not included in the study, the rationale being that they are not auxiliaries, as they function as contentful main verbs, which lie beyond the scope of this thesis.

(388) a. *Ada-lah ia sangat budiman.* (Nonverbal)
AUX-COM 3.SG very gentleman
'He is quite the gentleman.'

b. *Ada-lah hamba meŋ-bawa dagang-an mas dan perak.* (Verbal)
AUX-COM 1.SG ACT-bring trade-NMZ gold and silver
'I bring trade of gold and silver.'
(Hikayat Merpati Mas & Merpati Perak – 19th century AD)

Time	Text	Words	Nonverbal Clause	Verbal Clause
14th Century	Hikayat Raja Pasai	22,509	3	1
	Hikayat Muhammad Hanafiah	56,906	1	1
	Total:	79,415	4	2
15th Century	Undang-Undang Melaka	14,628	7	0
	Total:	14,628	7	0
16th Century	Hikayat Pandawa Lima	85,722	5	5
	Hikayat Indera Putera	70,954	3	1
	Total:	156,676	8	6
17th Century	Hikayat Sang Boma	92,113	14	16
	Hikayat Banjar & Kota Waringin	43,073	8	6
	Total:	135,186	22	22
18th Century	Hikayat Sang Bima	25,340	4	1
	Hikayat Patani	1,335	1	1
	Hikayat Putera Jaya Pati	26,912	1	2
	Hikayat Syah Mardan	31,081	6	5
	Hikayat Nakhoda Muda	17,760	5	0
	Hikayat Hasanuddin	7,586	5	13
	Total:	140,014	22	22
19th Century	Hikayat Pahang	46,212	2	4
	Hikayat Merong Mahawangsa	39,163	5	11
	Hikayat Merpati Mas & Perak	49,974	19	13
	Total:	135,349	26	28
20th Century	Hikayat Johor Serta Pahang	7, 537	1	3
	Hikayat Seri Kelantan	28,612	0	5
	Hikayat Kerajaan Sikka	78,397	22	2
	Total:	114,546	23	10

Table 32: The Number of Occurrences of Auxiliary 'Ada-lah'

At the beginning of the Classical Malay period in the 14th century, *ada-lah* as an auxiliary was quite scarce since the *lah* morpheme had only recently entered the language, as attested in the 14th century with only six instances of *ada-lah* as an auxiliary in the corpus of 79,415 words by the Malay Concordance Project. The number steadily increased throughout the Classical Malay period, despite the spike in the 15th century.[87]

Since auxiliary *ada-lah* was used in both verbal and nonverbal clauses, its copular and comment-marking forms coexisted. This layering made it practically impossible to tell whether it was a copula or a marker of new information in copular clauses, especially since its frequency in either type of clause was equal. However, towards the end of the 19th century, the two uses became apparent, as the latter fell out of fashion. As its use in signalling new information declined, its use as a copula remained. This decline coincided with the completion of the change in word order from VSO to SVO and suggests that the grammaticalisation of copula *adalah* occurred towards the end of the Classical Malay period.

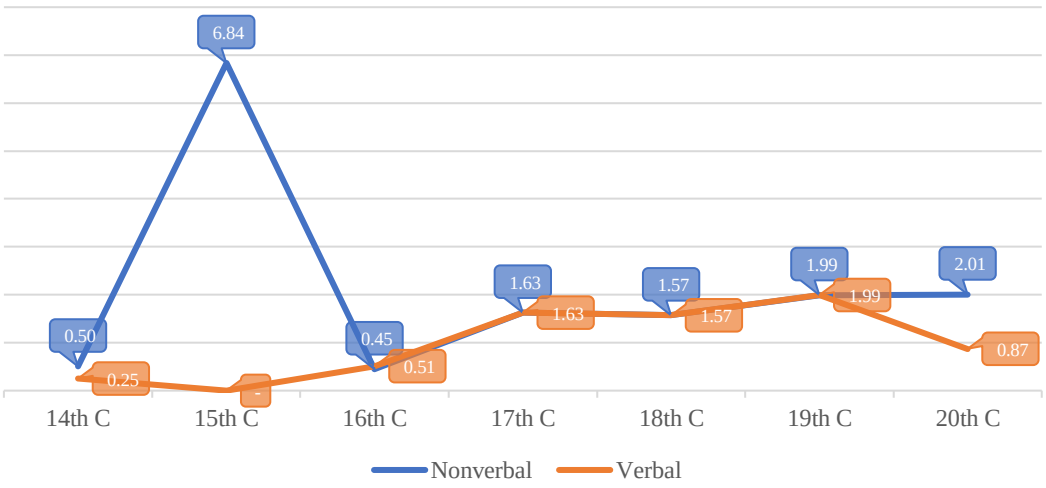


Figure 37: The Frequency of Auxiliary 'Ada-lah'[88]

[87] Due to there being only a single text from the 15th century available in the MCP corpus, as explained in Section 1.3, there is not enough data for the frequency of *ada-lah* in the chart to be representative of that period, which is the reason why there is an unexpected spike in Figure 37 in the 15th century. Most likely, there should be a steady increase.

[88] N *Ada-lah* ÷ (N words ÷ 10,000)

In spite of its decline in verbal clauses, the frequency of *ada-lah* in nonverbal clauses remained constant, which is analysed to be due to *ada-lah* having specialised in copular clauses and gaining its use as a copula before the completion of the change in word order. That the copular usage of *ada-lah* was not affected by the change in word order, as opposed to its usage as a marker of new information in verbal clauses, serves as evidence that comment marker *lah* had already fused onto the root *ada* in copular clauses. Had this grammaticalisation not happened, the copula would also have declined the same way it did in verbal clauses.

As a matter of fact, the trend observed in the chart coincides with the changing of the word order in Malay from VSO to SVO; the decline in V-T-C movement caused *ada-lah* as a marker of new information, which must occur in clause-initial position, to drop. Once the change in word order was complete, use of the comment marker declined and what was left was the copula. In fact, *ada-lah* occupied clause-medial position in all of the copular clauses in 20th century Hikayat Kerajaan Sikka. Cumming (1988) compares the use of *pun* and *lah* in Classical Malay and Contemporary Indonesian and concludes that whilst they were very frequent in the former, they have become quite rare in the latter.

In Modern Malay, although *ada-lah* is no longer used to mark the comment of a clause, vestiges, though scarce, exist in Modern Malay in fixed phrases of announcement or declaration such as *adalah + dimaklumkan* (informed), *diisytiharkan* (declared), *ditegaskan* (asserted), etc. Such a use of *ada-lah* is prescribed to be ungrammatical by DBP, but it is regularly used in formal writing, as shown below:

(389) *Ada-lah di-maklum-kan bahawa dakwa-an ter-sebut tidak benar.*
AUX-COM PASS-inform-APPL COMP claim-NMZ NVOL-mention NEG true
'It is informed that the claim mentioned is not true.'
(Dahali, 2021)

Nowadays, verbs and auxiliaries may only be affixed by *lah* by moving to the left periphery. However, movement is not required for copula *adalah* as *lah* has fused onto the root *ada,* which means that the two morphemes can no longer be separated. Unlike all other verbs and auxiliaries, copula *adalah* is the only auxiliary that need not move to clause-initial position to obtain the *lah* affix because *lah* is already fossilised on *ada,* owing to the grammaticalisation of the copula.

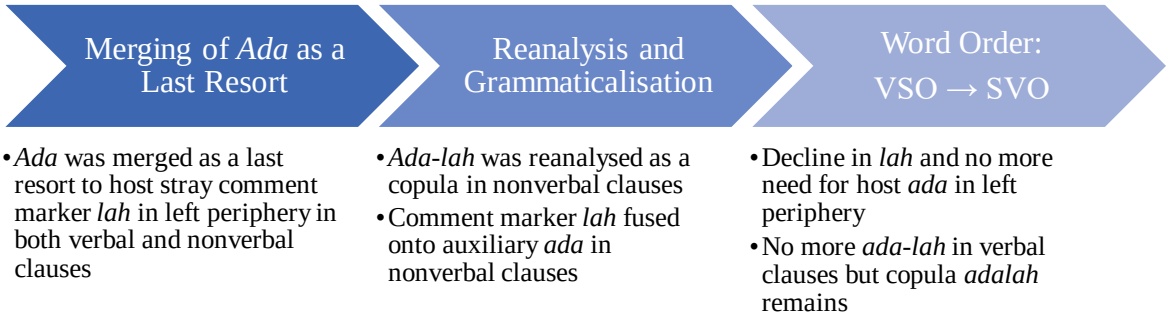


Figure 38: The Grammaticalisation of Copula 'Adalah'

6.2.2 The Grammaticalisation of *Ialah*

Copula *ialah* followed a different and complex grammaticalisation path from 3rd person *ia*, which commonly functioned as an argument. As shown in the examples below, *ia-lah* functioned as the focal subject in both verbal and nonverbal clauses. It is safe to assume that *ia-lah* had not yet grammaticalised into a copula in the middle of the Classical Malay period as it is the only referential DP that may be interpreted as the subject of the clauses in (390). Furthermore, *ia-lah* may be clefted from the rest of the clause by *yang* in (390b), which is only possible for DP arguments, as explained in Section 1.5.13.

(390) a. *Ia-lah suami Fatimah Zahra.* (Nonverbal)
3.SG-FOC husband F.
'HE is the husband of Fatimah Zahra.'

b. *Ia-lah yang meŋ-tangkap dia.* (Verbal)
3.SG-FOC COMP ACT-catch 3.SG
'It is HE who caught him.'
(Hikayat Muhammad Hanafiah – 14th century AD)

As a phrasal constituent, *ia-lah* is interpreted as a contrastive focus in all cases in which it occurs in clause-initial position, be it in verbal or in nonverbal clauses. Even when *ia-lah* is preceded by a topic, it retains a focal interpretation as usual, as illustrated in the example below:

(391) *Maka dalam hati=nya Ken Mawar Turirah ia-lah ini Sang Rajuna.*
DISC in heart=3 K. 3.SG-FOC PROX S.
'In Ken Mawar Turirah's heart, HE is Sang Rajuna.'
(Hikayat Pandawa Lima – 16th century)

However, notice that *ia-lah* in the examples below does not carry any focal reading whatsoever. In cases in which it occurs in what appears to be a specificational copular clause, it does not get interpreted as a focus. In such cases, *lah* can be said to be a comment marker, as opposed to a focus marker, despite combining with a phrasal constituent.

(392) a. *Ada-pun raksasa itu ia-lah dewa Batara Mahawisnu.*[89]
AUX-TOP monster DIST 3.SG-COM god B.
'As for the monster, it was Dewa Batara Mahawisnu.'

b. *Ada-pun ke-datang-an hamba ini ia-lah hendak meŋ-rebut negeri tuanhamba.*
AUX-TOP NMZ-come-NMZ 1.SG PROX 3.SG-COM want ACT-seize state 2.SG
'As for my arrival, it is to seize your state.'
(Hikayat Sang Boma – 17th century AD)

In terms of information structure, these copular clauses conform to the Topic-Focus alignment by Mikkelsen (2005a) – the DP preceding *ia-lah* is the topic, and the one following *ia-lah* the focus. *Ia-lah* itself does not carry any information-structural role but marks the partition between the topic and comment. As illustrated below, this configuration is different from one in which *ia-lah* is sentence-initial.

(393) a. [Focus *Ia-lah*] [Topic *dewa Batara Mahawisnu.*]
3.SG-FOC god B.
'HE is Dewa Batara Mahawisnu.'

b. [Topic *Ada-pun raksasa itu*] [Comment *ia-lah* [Focus *dewa Batara Mahawisnu.*]]
AUX-TOP monster DIST 3.SG-COM god B.
'As for the monster, it was Dewa Batara Mahawisnu.'

Ia-lah seemingly functions as a resumptive pronoun to the topic. However, resumption was not obligatory in topical constructions in general, as shown in (394), which entails that there were special circumstances that necessitated the use of *ia*. The crucial difference between (392) and (394) is the morpheme *lah*. Although examples (394) also involve topicalisation, they neither include *lah* nor *ia*.

[89] Unlike *pun*, which often marked old topics, *adapun* was used to mark new topics. It always preceded the constituent corresponding to the topic. The left periphery in Classical Malay apparently involved multiple topic phrases, possibly contrastive topic phrase (Lipták, 2011) and topic phrase – *adapun* heads the higher CTopP, whereas the preposed topic occupies the lower TopP.

(394) a. *Ada-pun nama negeri ini Pasai.*
AUX-TOP name country PROX P.
'As for the name of this country, (it is) Pasai.'

b. *Ada-pun besar batang lembing itu dua jengkal lilit.*
AUX-TOP big stick spear DIST two jengkal circumference
'As for the size of the spear, (it is) two *jengkal* in circumference.'
(Hikayat Raja Pasai – 14th century AD)

Copular clauses such as the one below with *ia* but without *lah*, are not attested, which suggests that *ia* was not base generated as an underlying argument, but presumably as a resumptive pronoun that was spelt out to host *lah* whilst its antecedent moved to a higher peripheral projection. Also, it entails that the surfacing of *ia* was contingent on the use of the comment-marking *lah* morpheme in the copular clause.

(395) **Ada-pun nama negeri ini ia Pasai.*
AUX-TOP name country PROX 3.SG P.
(As for the name of this country, it is Pasai.)

One might argue that, instead of *ia*, *ada* should be merged as a dummy auxiliary to host the stray affix, as described in predicational copular clauses in Section 6.2.1. However, the crucial difference between copular clauses with *ia-lah* and those with *ada-lah* is that, in the former, there is a preposed topic preceding *ia-lah*, whereas in the latter, there is no preposed topic. Neither the subject nor the predicate in predicational copular clauses undergoes Ā-movement to the left periphery of the clause. Therefore, evoking the [*u*Top] feature on the copula by Mikkelsen (2005a), *ia-lah* can be analysed to be different from *ada-lah* in that it has the additional [*u*Top] feature that attracts a topic to it, which entails a different derivation from *ada-lah*, which does not have said feature and does not involve topicalisation.

Given that the precursor of both copulas *ia-lah* and *ada-lah* were merged in the left periphery, both specificational and predicational copular clauses should be derived in a similar manner prior to the merging of the CP. To elaborate, in both types, the constituent that moves to SpecTP is the DP that is closest to the T^0 head bearing the EPP feature, viz. the referential DP in SpecPredP. Therefore, the structure of the root TP in both types are identical such that SpecTP is occupied by the referential DP, whilst the predicate DP remains in the complement position of PredP.

It is only when the CP is merged that the derivation of the two types diverges. In predicational copular clauses, the comment marker *lah* is merged as the head of FocP and *ada* is merged subsequently to serve as its host, considering that nothing can move to it – the predicate may not move as it would be interpreted as a contrastive focus since it is a phrasal constituent affixed by *lah* in clause-initial position (or it would be interpreted as inchoative in the occasion that it is an adjectival predicate). As for specificational copular clauses, the [*u*Top] feature on *lah* identifies the DP predicate as a goal and triggers it to move to SpecFocP. Following that, a higher TopP attracts it to move further to it, given that the [Top] feature on the DP is interpretable and not deleted. Finally, movement of the topic leaves the resumptive pronoun *ia* in SpecFocP to host *lah* and the derivation converges, as illustrated below:

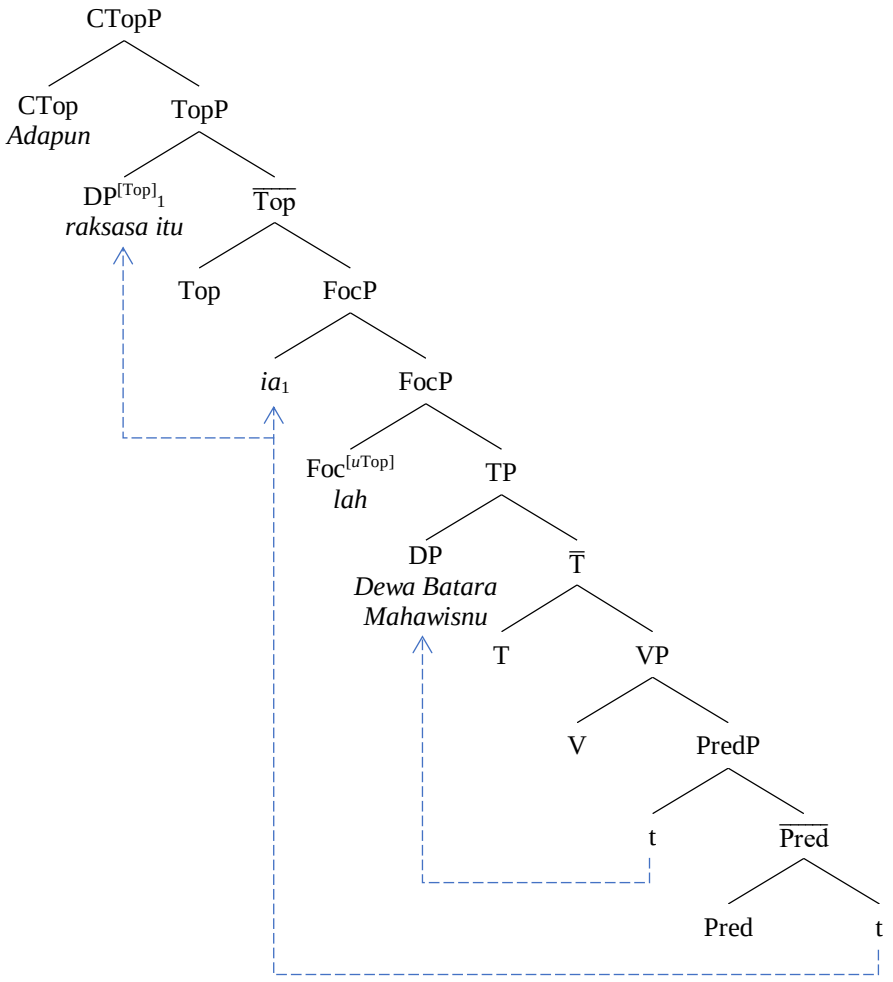


Figure 39: The Structure of a Copular Clause in Classical Malay with 'Ia-lah'

As for the actual copularisation of pronominal *ia-lah*, it can be said to have followed a common path taken by copulas that were formerly pronominal elements. Stassen (1997) states "PRO-NOMINAL COPULAS or PRO-COPULAS originated as resumptive subject pronouns in a topic-comment structure; they formed part of the sentence nucleus and were anaphorically related to the subject, which was placed outside the nucleus in the (commonly sentence-initial) topic position" (p. 77). The same position is taken by Hengeveld (1992). Principally, this is the phenomenon that occurred in the transition between Classical Malay and Modern Malay. Once the change in word order from VSO to SVO was complete, the clausal structure of copular clauses, in which left-dislocation and the spell-out of resumptive pronoun *ia-lah* had occurred, was totally reanalysed, triggering the grammaticalisation of *ialah* into a copula. What began as a marked topical construction had developed into a regular unmarked SVO clause. To be precise, the loss of *pun* (as well as *adapun*) diminished the topic-subject distinction and enabled reinterpretation of the topic as the unmarked grammatical subject in SpecTP.

Van Gelderen (2015) states that certain pronouns may grammaticalise into copulas via reanalysis from a phrasal constituent occupying SpecPredP into the Pred^{0} head, as argued of the demonstratives in Mandarin and other languages. The examples below, cited in Van Gelderen (2015), show the dual use of *shi* as a demonstrative and copula in Old Chinese and only as a copula in Modern Mandarin Chinese:[90]

(396) a. *Shi shi lie gui.* (Old Chinese)
PROX COP violent ghost
'This is a violent ghost.'
(Peyraube & Wiebusch, 1994)

b. *Zhe shi lie gui.* (Mandarin Chinese)
PROX COP violent ghost
'This is a violent ghost.'

[90] In fact, spec-to-head reanalysis is also common in other parts of the grammar crosslinguistically. In the CP domain, phrases in SpecCP commonly grammaticalise into complementisers, e.g. French *parce que* (because) from *par ce que* (by this that). Elsewhere, the development of French *pas* into a negator, as described by Jespersen's cycle, is also arguably an instance of spec-to-head reanalysis.

This spec-to-head reanalysis is the process undergone by pronominal *ia-lah*, as illustrated in Figure 40. The reanalysis of the topic as the canonical subject in SpecTP compressed, as it were, *ia-lah* into the head of TP, hence its grammaticalisation into a copula. In addition to that, the copula *ialah* has lost some of the inherent features of its previous pronominal component via semantic bleaching during the process of grammaticalisation (Lohndal, 2009; Van Gelderen, 2015). Although it retains its 3rd person feature, which explains why it only selects 3rd person subjects, it has lost its referential and number features. As for *lah*, it has retained its [*u*Top] feature. Also, the fusing of the two individual morphemes into a single one has allowed the remaining features of *ia* and *lah* to be collated onto a single head.

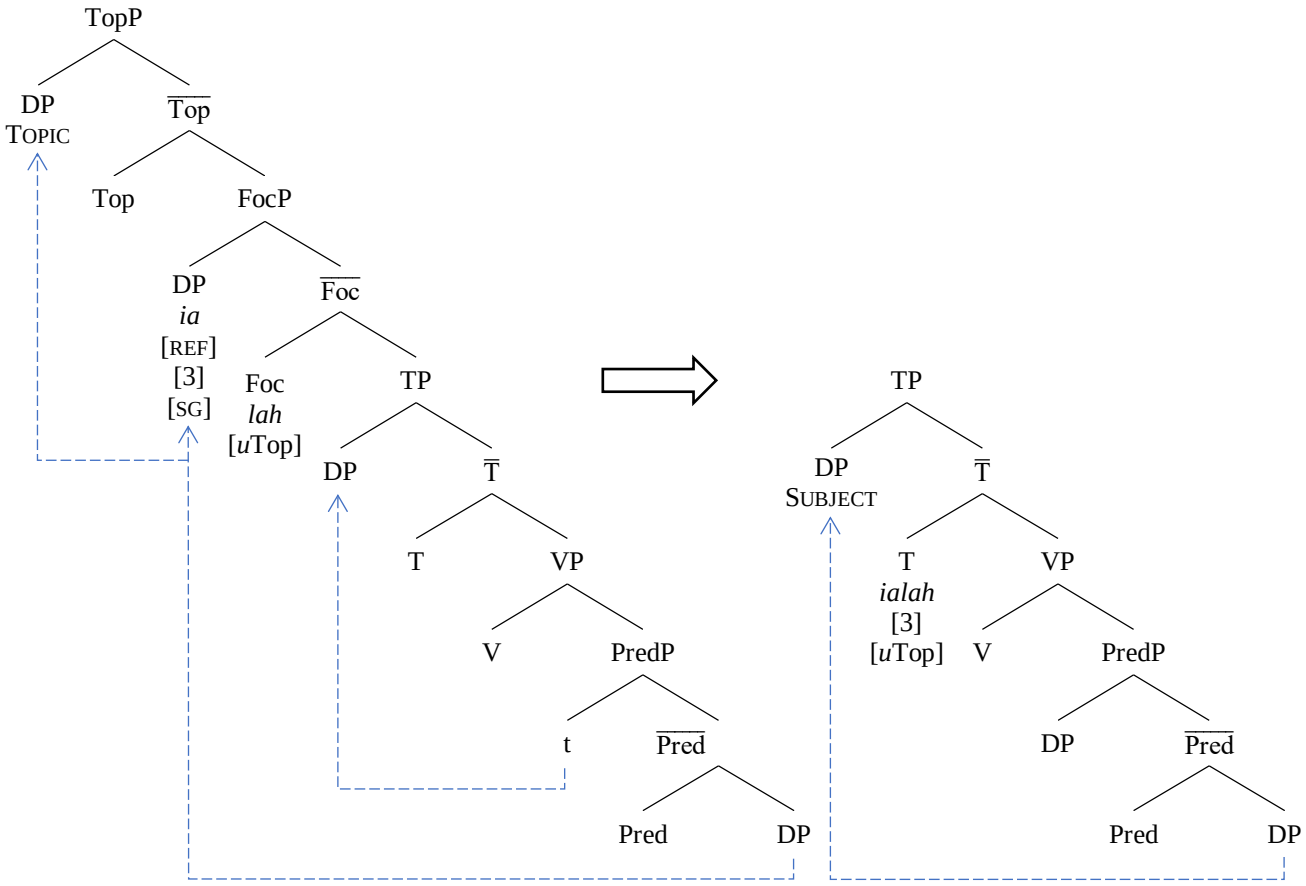


Figure 40: The Spec-to-Head Reanalysis of 'Ialah'

As hypothesised, the grammaticalisation of *ia-lah* into a copula occurred at the same time the word order of Classical Malay was changing from VSO to SVO, as shown in Figure 41. The corpus analysis data to chart Figure 41 are provided below:

Time	Text	Words	Nonverbal Clause	Verbal Clause
14th Century	Hikayat Raja Pasai	22,509	0	2
	Hikayat Muhammad Hanafiah	56,906	11	15
	Total:	79,415	11	17
15th Century	Undang-Undang Melaka	14,628	2	3
	Total:	14,628	2	3
16th Century	Hikayat Pandawa Lima	85,722	3	7
	Hikayat Indera Putera	70,954	3	31
	Total:	156,676	6	38
17th Century	Hikayat Sang Boma	92,113	9	8
	Hikayat Banjar & Kota Waringin	43,073	6	3
	Total:	135,186	15	11
18th Century	Hikayat Sang Bima	25,340	1	10
	Hikayat Patani	1,335	0	2
	Hikayat Putera Jaya Pati	26,912	1	2
	Hikayat Syah Mardan	31,081	9	4
	Hikayat Nakhoda Muda	17,760	1	0
	Hikayat Hasanuddin	7,586	5	15
	Total:	140,014	17	33
19th Century	Hikayat Pahang	46,212	27	9
	Hikayat Merong Mahawangsa	39,163	2	7
	Hikayat Merpati Mas & Perak	49,974	0	2
	Total:	135,349	29	18
20th Century	Hikayat Johor Serta Pahang	7, 537	6	6
	Hikayat Seri Kelantan	28,612	2	3
	Hikayat Kerajaan Sikka	78,397	24	7
	Total:	114,546	32	16

Table 33: The Number of Occurrences of 'Ia-lah'

Regarding the frequency of *ia-lah*, its use in nonverbal clauses rose consistently from the 16th century. As it peaked in the 20th century, its use in verbal clauses declined, consistent with the VSO to SVO change. Although the decline in the frequency of *ia-lah* in verbal clauses in the 17th century might be surprising, it is probably just a matter of choice by the author, as the usage of *ia* as a referential pronoun without *lah* was still very productive at the time.

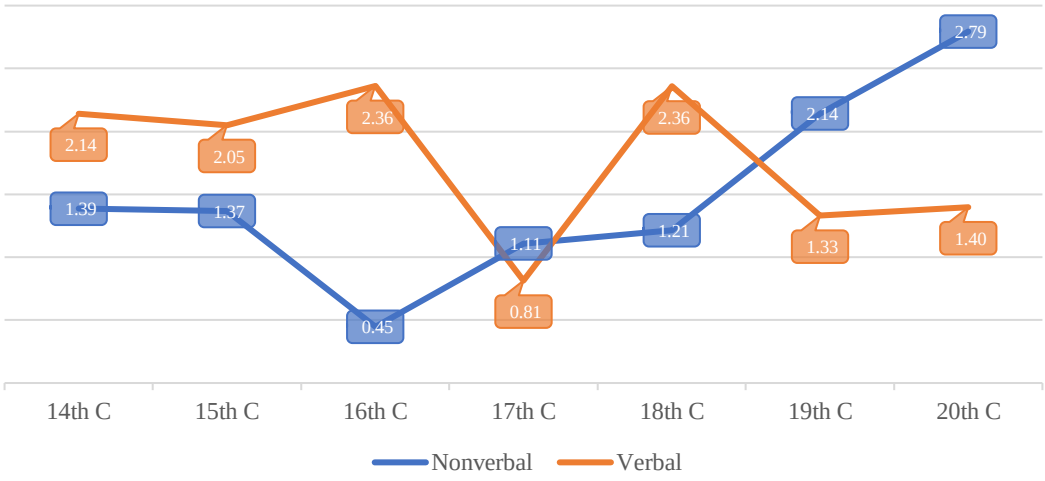


Figure 41: The Frequency of 'Ia-lah'

The increase in the usage of *ia-lah* as a copula coincided with the decline of topic marker *pun*, which drove the change in word order from VSO to SVO. Since *pun* was becoming obsolete, speakers of the language gradually lost the ability to recognise the constituent preceding *ia-lah* as a topic, hence its reanalysis as the canonical subject. This phenomenon is exemplified in cases in which *ia-lah* occurs in clause-medial position without *pun* to distinguish between topic and subject. The constituent preceding *ia-lah* can be said to occupy SpecTP in the examples below:

(397) a. *Utusan=nya ialah Daeng Masuki itu.*
messenger=3 COP D. DIST
'The messenger was Daeng Masuki.'

b. *Kapitan ialah bapak=nya.*
K. COP father=3
'Kapitan was his father.'
(Hikayat Kerajaan Sikka – 20th century AD)

Despite this, its frequency in verbal clauses in the 20th century was still quite significant. Towards the end of the 20th century, 3rd person *ia* had fallen out of fashion and was replaced by *dia*. This meant that *ia* and *ia-lah* could no longer be used as a referential pronoun, as described in Section 2.1. Despite the loss of pronominal *ia*, its *lah*-affixed form remained in copular clauses, which suggests that *ia-lah* had specialised as a copula. This occurrence appears to have happened in the 19th century when the use of *ia-lah* in verbal clauses decreased. Whilst its pronominal use declined, its copular use rose, which subsequently enabled *ia-lah* to specialise in copular clauses.

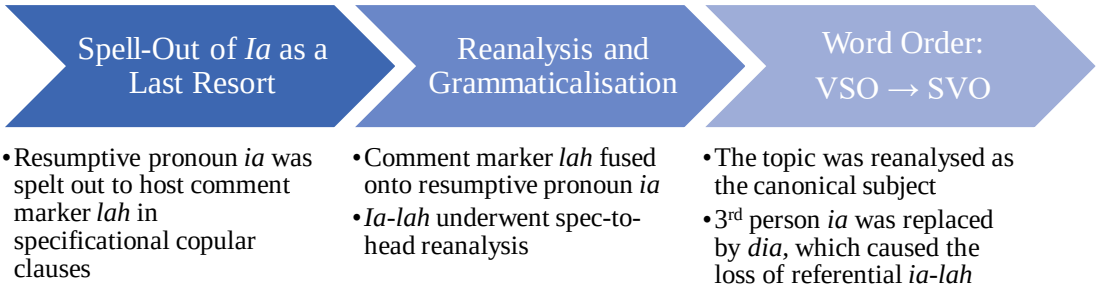


Figure 42: The Grammaticalisation of Copula 'Ialah'

6.3 The Potential Role of Language Contact

Malay has undergone extensive contact-induced change, especially with respect to the expansion of its vocabulary. According to Nathesan (2015), the Abridged Malay-English Dictionary by Wilkinson (1908) registers a total of 1524 words in Malay borrowed from foreign languages, which is quite sizeable for "a book of conventionally small size" (Wilkinson, 1908, p. v), with a significant proportion being of Sanskrit, Persian, Arabic, and English origins. Obviously, the influence of the English language on Malay has grown since then, with its special status as the medium of communication in official purposes such as proceedings in the Houses of Parliament and the Legislative Assembly, as outlined in the Federal Constitution under Article 152, with its recognition as Malaysia's second language, and as the medium of instruction in the teaching of Science and Maths in the 90s under the *Pengajaran dan Pembelajaran Sains dan Matematik dalam Bahasa Inggeris* (PPSMI) programme. Owing to its growing influence, there has been an influx of lexis that describes new concepts such as scientific and technological advancements, such as *komputer* (computer) *robotik* (robotic), etc., as also stated by Thomason and Kaufman (1992): "large numbers of English loanwords in scientific and technological areas occur in many languages and are not accompanied by structural borrowing" (p. 78).

Considering the impact of intense language contact, the copulas in Malay are claimed by Omar (2014) to have emerged in the 20th century and to be a product of translating the copula in Dutch or English to Malay. However, we are cognisant at the very least that they did not emerge in the 20th century. That the completion of the change in word order from VSO to SVO towards the 19th century left the use of *ia-lah* and *ada-lah* in nonverbal clauses unaffected, in contrast to their decline in verbal

clauses, as shown in Figure 37, is indicative of the demorphologisation and grammaticalisation of the copulas even before the 19th century. Arguably, this could have happened earlier, considering the language in the Malay dictionary by Bowrey (1701) is more characteristic of Modern Malay than Classical Malay. Therefore, the assumption that the copulas emerged in the 20th (or even the 19th) century is inaccurate.

Presumably, other languages that have had significant contact with Malay throughout history could have triggered the emergence of the copulas. Consider Sanskrit, one of the first languages known to have had a great impact on Old Malay, the earliest records of which date to the 4th century on the Muarakaman inscriptions from Kutai, Indonesia (Baskoro, 2016). According to Deshpande (1979), despite its massive influence on Southeast Asian languages for centuries, Indic, as well as the Hindu-Buddhist beliefs, traditions, and cultures associated with it, has caused no structural transfer (Thomason & Kaufman, 1992). Old Malay did not only allow the borrowing of a great number of lexical words from Sanskrit, but also various functional words, e.g. *saya* (1.SG; from *sahāya*), *antara* (between; from *anantara*), *kerana* (because; from *kāraṇa*), *tetapi* (but; from *tathāpi*), etc. Hypothetically, Sanskrit *as* and *bhū* as free-standing copulas could have entered Malay, or Malay could have *replicated* (Heine & Kuteva, 2005) Sanskrit copular clauses. However, this did not occur as they remained without overt copulas at the time. Although *svayaṃ-bhu* (self-COP; self-existing) is attested in the 684 AD Talang Tuwo inscription, the *bhu* morpheme was fixed on *svayaṃ* and not copular. It is not found elsewhere.

As shown in Section 6.2, the *lah* morpheme, had only entered the Malay language in the 14th century, and with it, the bimorphemic precursors of the Malay copulas *ada-lah* (AUX-COM) and *ia-lah* (3.SG-COM) had only gained usage in the Classical Malay era. By then, the influence of Sanskrit had diminished considerably, due to the advent of Islam in the Malay Archipelago and Arabic having replaced Sanskrit as the most dominant contact language, as evidenced by the development of the Jawi script based on the Arabic abjad, the use of Arabic in religious practice, etc. Coupled with the fact that the demorphologisation and grammaticalisation of *ada-lah* and *ia-lah* occurred sometime slightly prior to the 19th century, we can conclude that Sanskrit assumed no role in the emergence of the copulas in Malay.

Aside from the chronological discrepancies described above, I provide further evidence against the view that the copulas emerged out of some pressure to align with a foreign language or having been triggered by language-external factors. Specifically, the development of the copulas in Malay could not have been a case of structural borrowing or replication as the many morphological, semantic, and syntactic properties, behaviours, and phenomena of *ialah* and *adalah* are unique.

Structural borrowing or replication commonly allows the inheritance of certain properties associated with the structures borrowed from the source language. For example, the copulas *si* in Tsat and *pɪn* in Moklen exhibit properties that are also present in copular clauses in the languages from which the copulas were borrowed, Mandarin and Thai respectively. In both languages, the copula is only used with nominal predicates but not predicates of other syntactic categories, as shown below, likewise the use of the copulas in Mandarin and Thai. Apparently, this is not just a case of lexical borrowing, but replication of the copular construction including the syntactic selectional properties of the copula that is borrowed.

Tsat (Austronesian – Chamic)

(398) a. *Di*^{55} *nan*^{33} *sa*^{33} *mo* *si*^{11} *mai*^{33} *sa*^{33}.
lie.down DIST GEN cow COP female GEN
'The cow lying down is a female.'

b. *ʔai*^{33} *ni*^{33} *sat*^{24} *ʔan*^{33}.
water PROX truly cold
'This water is very cold.'
(Zheng, 1997)

Moklen (Austronesian – Other Malayo-Sumbawan)

(399) a. *Cəy* *pɪn* *məklɛːn*.
1.SG COP Moklen
'I am a Moklen.'

b. *Cʰiŋɔʔ* *lahan* *kutəy*.
rubber.tree many very
'The rubber trees are very many.'
(Larish, 2005)

As I have shown in Chapter 3, the Malay copulas *ialah* and *adalah* are used differently according to the relation between subject and predicate being predicational, specificational, or equative. Given that none of the languages with which Malay has had contact exhibit this distinction, it is most unlikely that replication had taken place. Also, given that the two copulas developed around the same time, had structural borrowing truly been the case, it should have resulted in the formation of a single copula in Malay to be used in all types of copular clauses, regardless of relation. On the contrary, the use of two different copulas in different environments entails different paths of grammaticalisation.

Further, the atemporal property that disallows overt encoding of the copulas in temporally bound environments in Malay is a reflection of how the copula *adalah* developed from the dummy auxiliary *ada* on a last-resort basis that would have induced an incoherent inchoative interpretation of the nonverbal predicate, had the dummy auxiliary not been merged (more on this in Section 6.4.4). Unlike the Malay copulas, which cannot be used in temporally bound contexts such as clauses with past and future time references, the Arabic copula KWN only occurs in the past and future tenses, and it cannot be used in the present tense (Alharbi, 2017). This complete opposite of the behaviour of the copulas in the two languages makes it clear that Arabic was not a model for the development of the Malay copulas or copular clauses.

Perhaps Portuguese might have had a facilitating effect on the grammaticalisation of *adalah,* seeing that it exhibits the same *ser* vs. *estar* distinction observed in Spanish. However, it certainly was not the root or cause, strictly speaking, as the aspectual factor carried by the *lah* morpheme, first attested in the 14th century, which necessitated the use of non-aspectual *ada* in nonverbal clauses and conditioned its grammaticalisation into the atemporal copula *adalah,* had already been present in the language centuries before the arrival of the Portuguese in the 16th century.

Also, the requirement that *ialah* combine with 3rd person subjects is a reflection of the grammaticalisation of *ialah* from a 3rd person pronoun. Although copular clauses in the contact languages exhibit agreement patterns with the φ-features of the subject, there is no restriction on the copulas that dictates what can and cannot be the subject of the copular clause.

In conclusion, the development of the Malay copulas was circumstantial and not a contact-induced product of borrowing or replication. In several cases, the chronology of the development of the copulas does not align well with that of some of the contact languages that presumably could have affected copularisation. Besides, replication is unable to account for the various unique properties of the copulas. Conversely, language-internal factors led to the grammaticalisation of the copulas *ialah* and *adalah*. It was specifically due to the morphosyntactic requirement that there be a host for *lah* (which very importantly predates the arrival of all the European colonial languages) that the copulas emerged.

6.4 From Diachrony to Synchrony

Some of the phenomena observed of copular clauses in Modern Malay can be attributed to the way in which the copulas grammaticalised. They have been described in the previous sections in the thesis: the impossibility for the copulas to host verum focus via pitch accent (Sections 2.1.2 & 2.2.2); the selectional properties exhibited by copula *ialah* on 3rd person subjects (Section 3.3.1); the exclusive use of *adalah* in predicational copular clauses and *ialah* in specificational copular clauses (Section 3.3.2); the atemporal property of the copula *adalah* (Section 4.4); and the lack of a relative clause in the structure of a cleft construction (Section 5.3). This section provides a description of those phenomena from a diachronic perspective.

6.4.1 Verum Focus

Copulas *ialah* and *adalah* may not express verum focus, as noted by Mustaffa (2018). For instance, the following example can only have a non-focal interpretation as it is found that the copula cannot be pitch-accented, unlike other auxiliaries affixed by *lah*.

(400) a. *Perkara itu adalah benar.*
matter DIST COP true
i. #'That matter IS true.' (Verum Focus)
ii. 'That matter is true.' (No Focus)

b. *Orang yang dedah-kan perkara itu ialah Zul.*
person REL reveal-APPL matter DIST COP Z.
i. #'The person who revealed that matter WAS Zul.' (Verum Focus)
ii. 'The person who revealed that matter was Zul.' (No Focus)

In copular clauses in Modern Malay, verum focus requires the use of some other linguistic material, e.g. an adverb. Without the adverb in the following example, the truth value of the proposition cannot be asserted.

(401) a. *Perkara itu sememangnya adalah benar.*
matter DIST surely COP true
'That matter surely is true.'

b. *Orang yang dedah-kan perkara itu sememangnya ialah Zul.*
person REL reveal-APPL matter DIST surely COP Z.
'The person who revealed that matter surely was Zul.'

This is especially surprising for *adalah,* as aspectual auxiliary *ada-lah* may host verum focus in verbal clauses, as do other auxiliaries, as illustrated below:

(402) a. *Ada-lah Zul dedah-kan perkara itu.*
AUX-FOC Z. reveal-APPL matter DIST
'Zul DID reveal that matter.'

b. *Mesti-lah Zul dedah-kan perkara itu.*
Must-FOC Z. reveal-APPL matter DIST
'Zul MUST reveal that matter.'

The etymology of the copulas sheds light on their inability to carry verum focus. Despite being what appears to be combinations of their roots with focus marker *lah,* as speculated by Yap (2007), the copulas cannot be focused because the *lah* morpheme in their precursors was in fact the comment marker, not the focus marker. The *lah* on the copulas was used to mark the partition between topic and comment, and not to contrast or assert the truth value of a proposition. Therefore, it is no surprise that the copulas may not be focused, either by pitch accent or movement, given that focus marking and comment marking are two different functions.

Recall from Chapter 2 that the copulas may co-occur with in-situ post-copular foci, which carry a contrastive interpretation, as in the examples below. [91] The possibility for focus marker *lah* to co-occur with the copula entails two things about *lah*: it was used to mark the comment portion of the clause, which included the focus, but did not precisely correspond to the focus; it is no longer associated with FocP.

[91] The *lah* morpheme seemingly combines with the focus in-situ, but since it is located in FocP, its apparent affixation to the focus is achieved by movement of the whole clause to SpecFocP.

(403) a. [Topic *Perkara itu*] [Comment *adalah* [Focus *benar-lah.*]]
matter DIST COP true-FOC
'That matter is TRUE.'

b. [Topic *Orang yang dedah-kan perkara itu*] [Comment *ialah* [Focus *Zul-lah.*]]
person REL reveal-APPL matter DIST COP Z.-FOC
'The person who revealed that matter was ZUL.'

That the copulas are no longer associated with FocP is supported by the impossibility for them to occur in the periphery. They do not move to FocP to obtain *lah*, as their precursors, *ia-lah* and *ada-lah*, did in Classical Malay, and as other auxiliaries still do, as in (402).

(404) *[FocP *Adalah* [TP *perkara itu benar.*]]
COP matter DIST true
(That matter IS true.)

Notice that this constituent order resembles a copular clause in Classical Malay, e.g. (405a). It was possible for the precursor of the copula to occur in clause-initial position because FocP was where dummy auxiliary *ada* was merged to rescue the stranded comment marker *lah*. Now in Modern Malay, such a derivation is no longer possible since the copula has grammaticalised into a single non-decomposable lexeme and no longer reflects its past AUX-COM form. The contrast between (405) and (406) below shows how the clause-initial auxiliary is possible in Classical Malay, but not possible in Modern Malay, despite the same array of constituents and meaning:

(405) a. *Ada-lah ia anakanda tuanku.* (Classical Malay)
AUX-COM 3.SG child Majesty
'He is your Majesty's child.'

b. *Tuanhamba ini ada-lah seperti orang meŋ-ambil bunga di dalam duri.*
2.SG PROX AUX-COM like person ACT-take flower LOC in thorn
'You are like a person taking a flower in a thorny bush.'
(Hikayat Sang Boma – 17th century AD)

(406) a. **Adalah dia anakanda tuanku.* (Modern Malay)
COP 3.SG child Majesty
(He is your Majesty's child.)

b. *Tuanhamba ini adalah seperti orang meŋ-ambil bunga di dalam duri.*
2.SG PROX COP like person ACT-take flower LOC in thorn
'You are like a person taking a flower in a thorny bush.'

6.4.2 3rd Person Subject Selection

Agreement markers often have pronominal origins (Ariel, 2000; Hopper & Traugott, 2003). Additionally, Givón (2015) argues that pronominal agreement starts out as *topic* agreement in topical constructions with left- or right-dislocation.

Copula *ialah* certainly does fit the bill for the characterisations mentioned above; selection of 3rd person subjects by copula *ialah* very clearly echoes its past as a pronoun. As something that grammaticalised from a 3rd person resumptive pronoun, it should naturally have a tendency to combine with constituents with which it shares certain features. Obviously, it retains its 3rd person feature and has lost its referential feature, but more subtly, it has developed the feature [–Pronominal] that makes it combine with R-expressions, as opposed to pronouns. Its requirement for non-pronominal DPs is attributed to its development from a resumptive pronoun. Consider the following data from Modern Malay:

(407) a. *Ali*$_1$ *kan, dia*$_1$ *suka Zainab.* (Full DP Subject: ✓)
A TOP 3.SG like Z.
'Ali, he likes Zainab.'

b. *Zainab*$_1$ *kan, Ali suka dia*$_1$. (Full DP Object: ✓)
Z. TOP A. like 3.SG
'Zainab, Ali likes her.'

c. *Dia*$_1$ *kan,* (**dia*$_1$) *suka Zainab.* (Pronominal Subject: ✗)
3.SG TOP 3.SG like Z.
(He, he likes Zainab.)

d. *Dia*$_1$ *kan, Ali suka dia*$_1$. (Pronominal Object: ✓)
3.SG TOP A. like 3.SG
'She, Ali likes her.'

Resumption mostly occurs when a non-pronominal DP is dislocated. When pronouns are involved, it is only when the dislocated pronoun functions as an object that resumption is possible. This pattern perfectly describes the requirement for the copula to combine with subjects that are R-expressions. Judging from the absence of sentences with resumptive pronouns whose antecedents are dislocated subject pronouns in the historical texts examined, the pattern observed in examples (407) presumably holds true in Classical Malay as well.

6.4.3 Predication vs. Specification

The use of either *adalah* or *ialah* in predicational and specificational copular clauses in modern Malay is a consequence of how the copulas developed in the history of Malay. *Adalah* is only used in predicational copular clauses, whereas *ialah* is used in specificational and equative copular clauses.

The occurrence of *ialah* in specificational copular clauses in Modern Malay is a result of different ways in which the copular clauses were derived in Classical Malay, which was in turn a reflection of different features on the head of FocP. The *lah* morpheme in specificational copular clauses specifically had the [*u*Top] feature that triggered the movement of the predicate cum topic. Subsequent movement of the topic to a higher topical projection then spelt out a resumptive pronoun that had to host the *lah* morpheme in FocP.

Conversely, *ada-lah* lacked the [*u*Top] feature, so the derivation of a predicational copular clause did not involve topicalisation. It essentially did not have any specific features or selectional properties that could have restricted the syntactic category of the constituent that occurred in the copular clause, which also entails that it could also occur in the left periphery of verbal clauses. Once it had grammaticalised, this freedom allowed copula *adalah* to combine with all sorts of nonverbal predicates.

Suppose that above *ada-lah* was a TopP, the resulting copular clause would still have *ada-lah* as the copula, as the *lah* morpheme did not have the necessary [*u*Top] feature to attract the topic to host it. Therefore, dummy auxiliary *ada* would have been merged to host *lah* nonetheless, prior to movement of the topic to TopP.

6.4.4 Atemporality

Recall from Chapter 4 that, in Modern Malay, copula *adalah* has atemporal properties such that it may not combine with predicates that denote a temporary state or occur in contexts that carry a change-of-state interpretation. This property developed during its grammaticalisation in Classical Malay, due to its use to prevent an inchoative interpretation of the predicate.

As argued in Section 6.2.1, copula *adalah* emerged from the necessity for focus marker *lah* to attach to a host in the event that the nonverbal predicate itself could not move, due to the inchoative effect that *lah* had on the predicate, as argued by Cumming (1988). The copula can be said to have gained an atemporal connotation that developed from its use to prevent an inchoative reading of the predicate. Even in Modern Malay, this inchoative effect is still seen, as shown in the contrast below.

(408) a. *Pokok itu adalah besar.* (Atemporal)
tree DIST COP big
'That tree is big.'

b. *Besar-lah pokok itu.* (Inchoative)
big-FOC tree DIST
'That tree became big.'

Dummy auxiliary *ada* was merged in FocP as a semantically vacuous element to host the stray *lah* morpheme in Classical Malay, as it still does in Modern Malay. To illustrate, when *ada* is merged within TP as a meaningful auxiliary, it carries some aspectual information. For example, the use of *ada* in the inflectional layer promotes an episodic reading, as opposed to a habitual reading without *ada*.[92]

(409) a. *Dia mesti belajar bahasa Latin.* (Habitual)
3.SG must study language Latin
'He must study Latin.'

b. *Dia mesti ada belajar bahasa Latin.* (Episodic)
3.SG must AUX study language Latin
'He must have studied Latin.'

When the examples above are transformed into polar questions, the same habitual and episodic interpretations are retained. With *ada*, the question carries an episodic reading, whereas without it, the question carries a habitual reading.

[92] The following example is ambiguous between a habitual and an episodic reading, as *ada* could either have been merged in TP as a meaningful auxiliary and moved to FocP or directly merged in FocP as a dummy auxiliary:

(xxvi) *Ada-kah dia belajar bahasa Latin?*
AUX-Q 3.SG study language Latin
i. 'Does he study Latin?' (Habitual)
ii. 'Did he study Latin?' (Episodic)

(410) a. *Mesti-kah dia belajar bahasa Latin?* (Habitual)
must-Q 3.SG study language Latin
'Must he study Latin?'

b. *Mesti-kah dia ada belajar bahasa Latin?* (Episodic)
must-Q 3.SG AUX study language Latin
'Must he have studied Latin?'

However, when merged as a supporting dummy auxiliary, the episodic reading is impossible. For instance, *ada*-support is applied in the following example to host interrogative marker *kah* in the left periphery (recall from Section 1.5.4 that moving an existing auxiliary is optional), and the resulting question is habitual, which is in contrast to the episodic reading of example (410b) in which *ada* is merged within TP. *Ada* could not have been merged in TP and moved to the left periphery, as it would have skipped *mesti* since it is merged in a lower position, which remains in its base-generated position.

(411) *Ada-kah dia mesti belajar bahasa Latin.*
AUX-Q 3.SG must study language Latin
i. 'Must he study Latin?' (Habitual)
ii. #'Must he have studied Latin?' (Episodic)

The copula grammaticalised from the same dummy auxiliary that forms the root of *ada-kah* in the example above.

6.4.5 The Structure of Clefts

As argued in Chapter 5 for clefts in Modern Malay, the cleft clause is not a complex DP whose head is null. This was also not the case in Classical Malay, as confirmed by several facts about focalisation in this stage of the language.

Unlike in Modern Malay, the use of *yang* seemed optional when a DP argument was focalised and/or affixed by *lah* in Classical Malay. There are many examples in which *yang* was not used with focalised DP arguments in verbal clauses, as illustrated below:

(412) a. *Ia-lah meŋ-bunuh cucu=ku Hasan dan Husain!*
3.SG-FOC ACT-kill grandchild=1 H and H.
'HE killed my grandchildren, Hasan and Husain!'
(Hikayat Muhammad Hanafiah – 14th century)

b. *Ia-lah meŋ-bunuh Gur Akas…*
3.SG-FOC ACT-kill G.
'HE killed Gur Akas…'
(Hikayat Indera Putera – 16th century)

c. *Ia-lah di-kehendak-i oleh baginda.*
3.SG-FOC PASS-want-APPL by HON
'HE is wanted by His Majesty.'
(Hikayat Syah Mardan – 18th century)

Given that *yang* was obligatory in headless relative clauses, as shown below, this finding indicates that there was no relative clause in cleft constructions.

(413) a. *Maka *(yang) tinggal habis lari.*
DISC REL remain finish flee
'(Those) who remained all fled.'
(Hikayat Sang Boma – 17th century AD)

b. *Kakanda ini hendak meŋ-cari *(yang) belum kakanda ketahu-i.*
1.SG PROX want ACT-find REL IMPRF 1.SG know-APPL
'I want to find (that) which I do not yet know.'
(Hikayat Syah Mardan – 18th century AD)

c. *Maka di-ceritera-kan oleh *(yang) ber-hikayat…*
DISC PASS-tell-APPL by REL INTR-story
'As told by (he) who tells stories…'
(Hikayat Merpati Mas dan Merpati Perak – 19th century AD)

There is also evidence of a matrix clause in clefts in Classical Malay. Given that there can only be one *lah*-marked constituent in a single clause, the two *lah*-marked constituents in the examples below should occur in different clauses.

(414) a. *Ada-lah ia-lah atas=nya meŋ-kuat-kan…*
AUX-COM 3.SG-FOC on=3 ACT-strong-APPL
'It is HE (who), on top of it, strengthened…'
(Hikayat Hasanuddin – 18th century)

b. *Maka ada-lah segala benda yang ter-sebut itu-lah yang di-kasihi=nya.*
DISC AUX-COM all thing REL NVOL-mention DIST-FOC COMP PASS-love=3
'It was ALL THOSE THINGS THAT ARE MENTIONED that are loved by him.'
(Hikayat Abdullah bin 'Abdul Kadir – 19th century)

6.5 Summary

This diachronic investigation has revealed that the grammaticalisation of the copulas was greatly influenced by the VSO-SVO change during the transition between Classical Malay and Modern Malay, which was galvanised by three events: the diminishing of the V-T-C movement rule, the rise in topical constructions with preposed triggers, and the loss of *pun* in such topical constructions.

The first event allowed *ada-lah* (AUX-COM) as a former marker of new information to specialise in copular clauses. Movement of the verb and other inflectional elements had become marked in Classical Malay, especially due to the emergence of *lah* at the time, which affixed to the moved elements. Due to the increasing markedness of V-T-C movement, verb-initial sentences with *lah* declined and, as a result, the use of *ada-lah* also dropped. Remarkably, the frequency of copula *adalah* saw a rise, which suggests that it had grammaticalised prior to the decline of verb initial sentences with *lah*.

The second event caused an increase in sentences with preverbal subjects. This event occurred in tandem with the diminishing of the V-T-C rule to promote verb-medial word order: SVO and TopVO. The increase in topical constructions also allowed the generation of what was to become specificational copular clauses in Modern Malay with *ialah*. Particularly, topicalisation forced the constituent that hosted the *lah* morpheme to move to SpecTopP and leave the resumptive pronoun *ia*, forming *ia-lah*, which marked the comment portion of the sentence, rather than being a contrastive or focal argument.

The third event allowed *ia-lah* (3.SG-COM) to undergo spec-to-head reanalysis from a resumptive pronoun occupying SpecFocP to a copula heading TP. Paired with the markedness of the V-T-C movement, the rise in topical structures in which the subject was preposed with topic marker *pun* prompted the growing frequency of sentences with preverbal subjects. In copular clauses, *ia-lah* functioned as the resumptive pronoun to left-dislocated subjects. The loss of *pun* neutralised the markedness of topical structures, which enabled topicalised subjects to be reanalysed as the unmarked canonical subject. This then fed the reanalysis of *ia-lah* into a copula via compression of the phrase into a head.

Both copulas emerged out of the need to provide a host for the *lah* morpheme, which was a language-internal development, as opposed to a contact-induced, language-external development such as structural borrowing or replication. Most importantly, the unique properties and behaviours of the Malay copulas were not inherited from other languages but were the idiosyncratic result of their grammaticalisation from dummy auxiliary *ada*, 3rd person *ia*, and comment marker *lah*.

The different pathways through which the Malay copulas grammaticalised in the past provide insight into their behaviour in the current stage of the language. For instance, the language employs two different copulas due to the specific environments in which each copula grammaticalised, which enabled their specialised function with respect to the predicates or arguments with which they combine. Copula *adalah* is strictly predicational, whereas copula *ialah* may form specificational or equative copular clauses. The data seen regarding the inability for the copulas to host verum focus, the atemporality of copula *adalah*, and the 3rd person agreement by copula *ialah* all fall back to how the two copulas developed from their morphological components and the circumstances that made their use in the copular clause necessary. Ultimately, the behaviours of the copulas in Modern Malay are vestiges of their past form as 3rd person *ia*, dummy auxiliary *ada*, and their combinations with comment marker *lah* in the copular clauses in which they occurred.

Chapter 7: The Copulas in Typological Perspective

The diachronic perspective in the previous chapter leads to the broader typological study of copular clauses within Austronesian. A typological survey of 40 Austronesian languages reveals that the emergence of overt copulas in Austronesian languages is related to word order. Apparently, verb-medial languages are statistically more likely to have overt copulas than verb-initial languages.

The absence of overt pronominal copulas in the Philippine-type languages is hypothesised in this chapter to be more profoundly related to the notion of *subject*, which such languages lack. In consideration of the grammaticalisation of pronominal copulas via spec-to-head reanalysis of a resumptive pronoun (e.g. Malay *ialah*), the lack of the notion of subject and the absence of the privileged subject position, viz. SpecTP, bleeds reanalysis of topic as subject, which *eo ipso* precludes spec-to-head reanalysis of resumptive pronouns (spelt out following dislocation of a topic) into copulas. As for verbal copulas, their absence from the Philippine-type languages concerns the encoding of thematic relations on the verb. Given that verbal copulas often develop from unergative posture verbs, copularisation is not possible in languages that observe Austronesian alignment due to the voice marking. The thematic role of the agent of an unergative verb is reflected on the verb, which is incompatible with the unaccusative nature of copular clauses. Furthermore, the trigger in voice alternations such as the benefactive voice and the locative voice is encoded as an applied argument. This makes the clause transitive, which is in opposition with the strictly intransitive copular clause.

Based on the failure to trace the copulas in the individual languages examined in this chapter to a common ancestral root, no copula can be reconstructed for Proto-Austronesian. This is because most of the Formosan languages (which make up the branches of the highest order of the Austronesian language tree) lacked overt copulas. In fact, it is not possible to reconstruct a copula beyond very closely related languages of a subgroup, such as Malay and Iban, let alone different Austronesian languages from different subgroups, which suggests that the copulas in the Austronesian languages are relatively recent innovations.

7.1 Alignment, Word Order, and the Emergence of Copulas

Austronesian languages are mainly verb-initial or verb-medial to the extent that there are no verb-final languages in the family other than a handful of Western Melanesian languages, which are generally agreed to have developed verb-final word order due to contact with Papuan languages (Polinsky & Potsdam, Forthcoming). As for SVO word order, Donohue (2007) states that the order is found in many southern Austronesian languages (between mainland Southeast Asia and New Guinea).

This group of languages is renowned for zero encoding of the copula; however, there remain many Austronesian languages, such as Malay, Daakaka, Biak, etc., that allow the use of an overt copula in nonverbal clauses, whose copular status is determined using the same criteria in Section 1.1.2 – it accompanies a nonverbal predicate in a non-modifying manner and has different semantics in clauses with verbal predicates, if possible in verbal clauses at all. However, data from some of the languages in this typological study are scarce, which makes it difficult to ascertain the copular status of an item. For instance, Jarai, Rejang, Irarutu, etc. are understudied languages that require more attention by linguists, making linguistic data hard to come by. Without the requisite linguistic data, I have relied on descriptive grammars to identify copulas.

Nevertheless, the items construed as copulas in this study accompany nonverbal predicates, as opposed to verbal predicates, which satisfies the first criterion. Optionality of the item to be examined as a copula is used as a basic test for the predicatehood of the nonverbal constituent. The fact that it is omissible is enough to signal that the nonverbal constituent corresponds to the predicate. Additionally, the fact that most of the items examined (especially all of the pronominal copulas) are semantically vacuous in their use with nonverbal predicates, as opposed to their use only as meaningful elements in verbal clauses, satisfies the second criterion and allows us to provisionally identify them as copulas in nonverbal clauses. For example, similar to *itu* in Malay (especially the variety spoken in Indonesia), the distal demonstrative *iku* in Javanese may be used optionally in copular clauses with the proximal demonstrative *iki* as the subject. Clearly, it does not carry any semantic content. Therefore, the predicate of the clause is unambiguously the nonverbal constituent

"jenis woh". Conversely, this meaningless use of *iku* in verbal clauses is not possible. Given that *iku* is not a predicate and is used as a meaningful distal demonstrative in verbal clauses, it can be identified as a copula in nonverbal clauses.

Javanese (Austronesian – Other Malayo-Sumbawan)

(415) a. *Iki* (*iku*) *jenis woh.* (Nonverbal)
PROX DIST type fruit
'This is a type of fruit.'

b. *Iki* (**iku*) *wis di-pangan.* (Verbal)
PROX DIST PRF PASS-eat
'This has been eaten.'

Perhaps in further research, the copular clauses in the languages examined can be revisited to confirm the status of the items in question as copulas. Until more data from the understudied languages become available, one might consider this typological survey a preliminary study.

The question that arises from the observation of some languages allowing overt encoding of the copula(s) despite the predominance of the use of zero copulas is what features condition or prevent the emergence or development of overt copulas, also assuming the presence of zero copulas. In Table 34, a survey of forty Austronesian languages reveals that there is not much correlation (or a very weak one at best) between inflectional morphological marking (e.g. person, number, gender, case, subject-verb agreement, tense, aspect, and mood) and overt encoding of the copula.[93] The finding that the analytic languages allow overt encoding of the copula constitutes further evidence against the dummy hypothesis by Dik (1997): copulas "carry those operators of Tense, Aspect, and Mood which require a verbal form if they are to be expressed" (p. 199); the copulas are not simply morphemes employed to host grammatical relations that are usually encoded on the verb, since the verb in the languages examined do not exhibit such grammatical distinctions in the normal case.[94]

[93] The CASE, AGR, and TAM columns are for overt morphological marking of case, agreement, and TAM features on the verb or related DPs, e.g. Malay does not have case distinctions and agreement marking, whilst TAM is marked periphrastically.

[94] This does not include the morphological marking of the trigger on the verb, which is arguably an instance of voice, as opposed to agreement in the traditional sense.

Subgroup	Language	Align.	Order	CASE	AGR	TAM	COP
Malayic	Malay	ACC	SVO	-	-	-	✓
	Iban	ACC	SVO	-	-	-	✓
	Minangkabau	ACC	SVO	-	-	-	✓
	Urak Lawoi’	ACC	SVO	-	-	-	✓
Chamic	Cham	ACC	SVO	-	-	-	✓
	Jarai	ACC	SVO	-	-	-	✓
	Tsat	ACC	SVO	-	-	-	✓
	Acehnese	SPLIT	SVO	-	✓	-	-
Other Malayo-Sumbawan	Javanese	ACC	SVO	-	-	-	✓
	Moklen	ACC	SVO	-	-	-	✓
	Rejang	ACC	SVO	-	-	-	✓
	Batak	ACC	VOS	-	-	-	✓
South Halmahera-West New Guinea	Ambel	ACC	SVO	-	✓	-	✓
	Matbat	ACC	SVO	-	✓	-	-
	Biak	ACC	SVO	-	✓	-	✓
	Irarutu	ACC	SVO	-	✓	✓	✓
Southern Oceanic	Lelepa	ACC	SVO	-	✓	✓	✓
	Daakaka	ACC	SVO	-	✓	✓	✓
	Bierebo	ACC	SVO	-	✓	✓	✓
	Erromangan	ACC	SVO	-	✓	✓	✓
Meso-Melanesian	Teop	ACC	SVO	-	✓	-	-
	Papapana	ACC	SOV	✓	✓	✓	-
	Nakanai	ACC	SVO	✓	✓	✓	-
	Roviana	ERG	VSO	✓	✓	-	-
Formosan	Rukai	ACC	VSO	✓	✓	✓	-
	Puyuma	ERG	VSO	✓	✓	✓	✓
	Seediq	SPLIT	VSO	✓	✓	✓	-
	Saisiyat	SPLIT	SVO	✓	✓	✓	-
Barito	Dusun	ERG	VSO	✓	-	✓	-
	Malagasy	SPLIT	VOS	✓	-	✓	-
	Ma’anyan	SPLIT	SVO	-	-	-	-
	Paku	SPLIT	SVO	-	-	-	-
Polynesian	Samoan	ERG	VSO	✓	-	-	-
	Tongan	ERG	VSO	✓	-	-	-
	Hawaiian	ACC	VSO	-	-	-	-
	Māori	ACC	VSO	-	-	-	-
Philippine	Tagalog	ERG	VSO	✓	-	✓	-
	Cebuano	ERG	VSO	✓	-	✓	-
	Ilocano	ERG	VSO	✓	-	✓	-
	Hiligaynon	ERG	VSO	✓	-	✓	-

Table 34: Comparison of Austronesian Languages

The features most pertinent to the copulas are those related to syntactic alignment, word order, and case marking. Meanwhile, agreement and TAM seem superfluous, as concluded by there being no pattern in the contingency tables below:

	+COP	–COP
+AGR	8	9
–AGR	11	12

Table 35: ±COP *vs.* ±AGR *Contingency*

	+COP	–COP
+TAM	6	11
–TAM	13	10

Table 36: ±COP *vs.* ±TAM *Contingency*

A quick look at the table reveals several interesting (near-)universals:

- All the ergatively aligned languages mark for case.
- All the ergatively aligned languages are VSO.
- Except Puyuma, all the languages with overt copulas are accusatively aligned.
- Except Puyuma, none of the languages with overt copulas mark for case.
- Except Puyuma, none of the languages with case marking have overt copulas.
- Except Puyuma and Batak, all the languages with overt copulas are SVO.
- Except Batak, Papapana, Rukai, Hawaiian, and Māori, all the accusatively aligned languages are SVO.

It is observed in Table 34 that the syntactic alignment of an Austronesian language correlates with its word order. All the ergatively aligned languages in the table have verb-initial order (9/9; 100%), whereas most accusatively aligned languages have SVO order (20/25; 80%), as summarised in the table below. It clearly reflects the ancestral verb-initial order of Proto-Austronesian (Donohue, 2005), which was ergatively aligned, as well as the modern Austronesian languages that are the most conservative, e.g. Tagalog. As a language becomes split-ergative and then accusatively aligned, the word order also changes, which makes apparent a one-way implicational relation. In fact, it was perhaps first noticed by Blust (2001) that this relation holds between the number of voice alternations – which implies syntactic alignment, i.e. a higher number of voice alternations signals the conservativeness of the ergative syntax of the language – and word order: "this relationship can be stated as an implication such that the presence of three or more focus possibilities implies verb-initial constituent order with almost perfect accuracy" (p. 70).

Alignment	V-Initial	V-Medial	V-Final	Total
Ergative	9	0	0	9
Split	2	4	0	6
Accusative	4	20	1	25

Table 37: Alignment vs. Word Order

With regard to case, although the loss of it is related to change in word order, it is unlikely to directly have anything to do with the emergence of copulas in a language. Following change in word order, case is lost as it becomes superfluous, as opined by Fischer (2010). This appears to be the case, as opposed to change in word order following or due to loss of case, as several of the non-accusative languages (i.e. ergative and split-ergative languages) have retained case marking despite having undergone change in word order, e.g. Papapana, Nakanai, and Saisiyat.

Most importantly, notice the predominance of languages with SVO word order to allow overt encoding of the copula. There is apparently a much higher probability for an Austronesian language to have overt copulas if the word order is SVO, though there still remain a few SVO languages without overt copulas, namely Saisiyat, Ma'anyan, Paku, Teop, Nakanai, Matbat, and Acehnese.

Copula	V-Initial	V-Medial	V-Final	Total
Overt	2	17	0	19
Null	13	7	1	21

Table 38: Word Order vs. Overt Copula

Puyuma and Toba Batak are the only verb-initial languages to allow overt copulas. An explanation for the emergence of the copula in Puyuma is currently wanting as data on the language are scarce. As for Toba Batak, the case of the clause-medial position of what Woollams (2005) calls the "copulative particle" is interesting, considering that the language is VOS. It is only when an SVO order is used that *émkap* is used, primarily because *kap* is an emphatic marker. Nonetheless, *émkap* can be identified as a copula based on its exclusive distribution in nonverbal clauses, as opposed to *kap*, which occurs in verbal clauses. This information-structural relation features prominently in both change in word order as well as the development of overt copulas, as it has in the history of Malay.

Toba Batak (Austronesian – Northwest Sumatra-Barrier Islands)

(416) *Kuta si meriahna émkap Juhar.*[95]
village REL big.NMZ COP J.
'The biggest village is Juhar.'
(Woollams, 2005, p. 89)

There being no overt copulas in the non-accusatively aligned languages (with Puyuma as an exception) makes for an even stronger suspicion that the emergence of copulas is related to the syntactic alignment of a language, and subsequently the word order of the language. Given the two statistical tendencies, i.e. for SVO languages to be accusative and for overt encoding of the copula to be in SVO languages, an implicational relation between the two is apparent. If this relation holds, it could explain why there are no overt copulas in the ergative Austronesian languages.

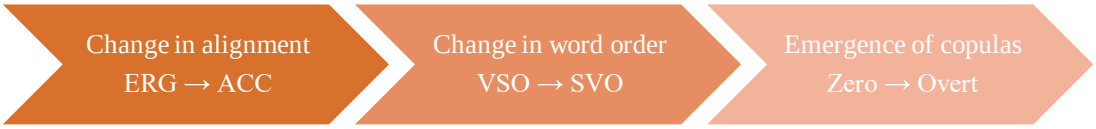


Figure 43: The Cline of the Emergence of Pronominal Copulas in Austronesian

Based on the cline above, in order for the word order of an Austronesian language to change from VSO to SVO, it must first evolve from an ergative language into an accusative language. Only after the language has developed SVO word order can overt copulas emerge. However, it must be noted that a language might not need to develop overt copulas at all, as zero encoding of the copula might suffice. It is only through certain circumstances that copulas emerge in a language, such as the information-structural changes in the history of Malay discussed in Chapter 6.

The cline predicts the following stages. Languages in stage 1 are all those that are ergatively aligned, and thus VSO, such as Tagalog; those in stage two have

[95] *Émkap* is also used as an expression used for clarification in appositives (see Section 2.3.2 on Malay *iaitu*), as illustrated below:

(xxvii) *Kerina si nggeluh erdalin arah dalin é, émkap kematén.*
all REL live walk vicinity road DIST COP death
'All who live must go down that path, that is, death.'
(Woollams, 2005, p. 267)

undergone change in alignment but have retained verb-initial word order, e.g. Rukai, Hawaiian, and Māori;[96] those in stage three have undergone both change in alignment and word order, e.g. Matbat, Teop, and Nakanai. Finally, those in stage four have developed overt copulas following change in alignment and word order, e.g. Malay.

Stage	Alignment	Word Order	Copula
1.	ERG	VSO	Zero
2.	ACC	VSO	Zero
3.	ACC	SVO	Zero
4.	ACC	SVO	Overt

Table 39: The Stages towards Copula Emergence

Therefore, the immediate question that needs answering is why it is not possible for the ergative languages to undergo change in word order from VSO to SVO without first having undergone change in alignment from ergative to accusative.

7.1.1 Change in Alignment from Ergative to Accusative

Before answering the question relating to alignment and word order, investigating the change of the alignment in the history of Malay from ergative to accusative should prove beneficial. The examination of Tagalog (ergative), Malagasy (split-ergative), and Indonesian (almost fully accusative) by Aldridge (2011a) yielded the finding that the change from ergative to accusative alignment in Austronesian languages is triggered by the reanalysis of two constructions, the antipassive construction and the ergative construction (i.e. the patient voice). Without going too deeply into the case-theoretic details, I shall provide evidence from Malay for the analysis.

First, the antipassive construction (which is semantically transitive but syntactically intransitive) evolved into the regular transitive construction, which "yields a mapping from semantic to grammatical relations which is parallel to transitive clauses in accusative languages" (Aldridge, 2011a, p. 2). This change is

[96] Stage 2 could also be appropriate for split-ergative languages. Based on their SVO order, the split-ergative languages in the table might arguably be mainly accusative, but with remnants of ergative syntax, as in the case of Malay and the patient voice that it has retained. For example, despite Acehnese, Ma'anyan and Paku being categorised as split-ergative languages, they have marked passives (Diedrich, 2018; Gudai, 1985; Legate, 2012), which are characteristic of nominative-accusative languages.

exemplified by the correspondence of the intransitive *maŋ-* prefix on the verb in the antipassive construction in Tagalog with the transitive *meŋ-* prefix in Malay and the likewise transitive *maŋ-* prefix in Malagasy.[97] As illustrated below, the most conservative of the three languages, the ergatively aligned Tagalog, has retained the antipassive construction and the *maŋ-* prefix indicates that the verb is intransitive. Conversely, the same construction has turned transitive in the split-ergative and accusative languages Malagasy and Malay, as the corresponding prefixes are used on verbs that take an internal argument and an external argument.

(417) a. *Naŋ-kagat ang lalaki ng bato.* (Tagalog – Antipassive)
INTR.PRF-bite ABS man OBL stone
'The man bit a stone.'

b. *Naŋ-kaikitra ny vato ny lehalihy.* (Malagasy – Transitive)
PST.AV-bite DET stone DET man
'The man bit the stone.'

c. *Lelaki itu meŋ-gigit batu itu.* (Malay – Transitive)
man DIST ACT-bite stone DIST
'The man bit the stone.'

In the history of Malay, the reanalysis of the antipassive into transitive is evident as recently as the Classical Malay period. Cumming (1988) reports that preposition *akan* was optionally used to mark the patient of a clause in Classical Malay. The optional nature of the preposition suggests that its use was declining, which means that the reanalysis had started much earlier than the Classical Malay period. Most importantly, its use to mark the patient as an oblique, rather than the selection of the patient directly by the verb, is characteristic of the antipassive voice. To illustrate, the patient in the examples below is introduced by preposition *akan*, which denotes that the construction is intransitive, specifically antipassive:

(418) a. *Jikalau tuan hendak akan anak…* (Verbal)
COND 2.SG want PREP child
'If you want a child…'

b. *Maka sahut=nya akan Tun Beraim Bapa…*
DISC answer=3 PREP T.
'He answers Tun Beraim Bapa…'

[97] In Tagalog and Malagasy /m/ → /n/ in the past tense or perfective aspect.

c. *Tuan-lah akan bapa hamba.* (Nonverbal)
2.SG-FOC PREP father 1.SG
'You are my father.'

d. *Di sini-lah akan tempat=ku.*
LOC here-FOC PREP place=1.SG
'Here is my place.'
(Hikayat Raja Pasai – 14th century)

In Modern Malay, the use of *akan* in such a way has largely been lost. Its use is now restricted to a small set of predicates, mostly psych verbs such as *cinta* (love), *benci* (hate), *ingat* (remember), etc. Even then, the use of *akan* is optional and does not affect the meaning of the sentence in any way.

(419) a. *Aku benci* (*akan*) *dia.*
1.SG hate PREP 3.SG
'I hate her.'

b. *Aku cinta* (*akan*) *dia.*
1.SG love PREP 3.SG
'I love her.'

With regard to case, the antipassive construction is analysed to have developed accusative case. Technically, the loss of *akan* would have left the patient caseless, but its visibility indicates the availability and assignment of structural case by the verb. Therefore, the previously oblique patient having acquired case lends support to the hypothesis by Aldridge (2011a) that the antipassive construction undergoes reanalysis into the transitive construction in Austronesian languages in the course of their change in alignment from ergative to accusative.

Second, Aldridge (2011a) argues that the ergative transitive construction (viz. the patient voice) evolved into the passive construction. This development is shown in the use of the *ni-* prefix in Old Malay, which later evolved into the passive *di-* prefix. As illustrated in (420a), *ni-* had not fully evolved into a valency-decreasing morpheme just yet, as both the DPs retain their argument status. The argument status of the agent in the *ni-* construction parallels that of the agent in the patient voice, e.g. in (420b), making the two constructions basically identical, with the exception of the *ni-* morpheme.

(420) a. *Jāṅan ya ni-knāi sa-vañak=ña yaṃ upasargga.* (*Ni-* Construction)
PROH 3.SG NI-afflict one-many=3 LIG misfortune
'May <u>all misfortunes</u> not afflict him.'
(Talang Tuwo inscription – 684 AD)

b. *Ini janma ku=minta.*[98] (Patient Voice)
PROX life 1.SG=request
'I requested <u>this life</u>.'
(Mañjuśrīgṛha inscription – 793 AD)

However, there are examples of the *ni*-construction lacking an agent, making it appear as a true passive in some instances, as shown below. This observation entails that the *ni-* morpheme was starting to gain valency-decreasing properties and innovate into a true passive, allowing optionality of the agent.

(421) *Tatkālā=ña yaṃ maṃmaṃ sumpaḥ ini ni-pāhat…*
during=3 YAṂ imprecation curse PROX PASS-engrave
'When the imprecation of this curse was engraved…'
(Kota Kapur inscription – 686 AD)

A century later, in the true passive in (371b), repeated below, the agent of the clause is very clearly seen to have been demoted to an adjunct *oleh*-phrase.[99]

(422) *Dayang Aṅkatan…di-bari waradāna ulih Sang Pamgat Senāpati.* (Passive Voice)
D. PASS-give favour by S.
'Dayang Angkatan... was given a favour… by Sang Pamgat Senāpati.'
(Laguna Copper-Plate inscription – 900 AD)

Based on the similarity in the argumenthood of the agent of both the *ni*-construction and the ergative construction, the two constructions can be said to have initially been one and the same, with merely the difference of an additional optional morpheme on the verb. Although examples (420a) and (420b) appear different in terms of word order, they are identical constructions, save for the *ni-* morpheme. The difference in the form of the agent as either a clitic or a full DP affects the word order of the sentence. In (420a), it is a free-standing DP that remains in Spec*v*P after movement of the verb to Voice, resulting in a Patient-Verb-Agent order (see Figure 45

[98] The English translation provided by Griffiths (2020) is "this life has been requested by me", which is in the passive voice; however, since it is in the patient voice (see Arka & Manning, 1998 for a discussion on the argumenthood of the pronominal patient as the grammatical subject), the translation should more accurately be in the active voice in English, with the underlining indicating the trigger.

[99] Even in Modern Malay, verbs affixed by passive *di-* may retain an agent argument, albeit rare.

for the syntactic representation of the sentence). Meanwhile, in (420b), it is a proclitic that moves together with the verb to Voice, resulting in a Patient-Agent-Verb order.

The variant of the ergative construction with the optional *ni-* prefix later evolved into a distinct construction once it acquired its valency-decreasing property. In other words, the optional *ni-* morpheme split from the main ergative construction and afterwards independently developed into a passive morpheme, with the /n/ → /d/ change first attested in the Gandasuli inscription (832 AD). Finally, the split of the two constructions allowed the ergative construction to survive unchanged and be known as the patient voice, alongside the new true passive construction with the *di-* morpheme that is known today.

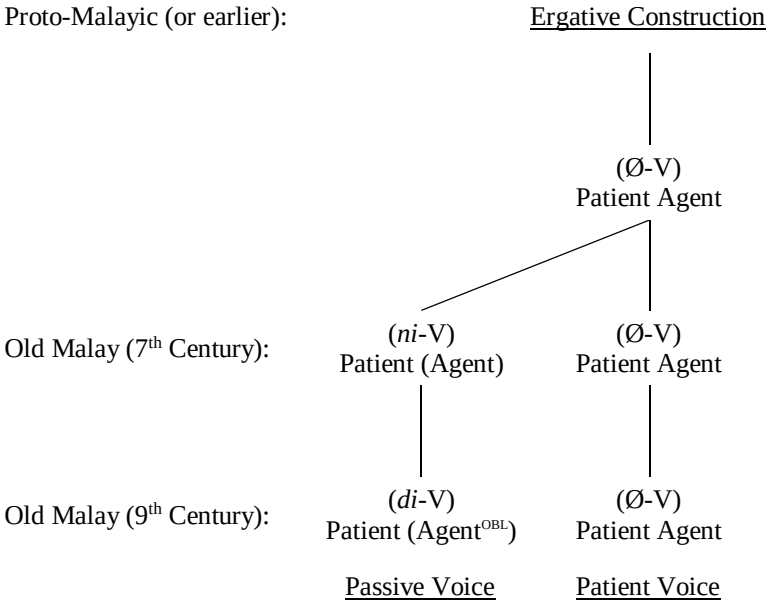


Figure 44: The Passive-Patient Split

The development of the true passive from the ergative transitive meant the loss of the ergative argument, which entails the development of an unaccusative construction, as per Burzio's (1986) generalisation. Therefore, the lack of both ergative and accusative case in the construction could only mean that the patient had to obtain structural case in the form of the nominative. As time went on, nominative and accusative case presumably spread by analogy, which resulted in the assignment of nominative case to the subject of both transitive and unergative intransitive clauses.

The only remnant of the ergative in Malay now is the agent of the patient voice. As we have seen in Section 1.5.2, the trigger of the patient voice in Modern Malay corresponds to the grammatical subject that resides in SpecTP, whilst the agent occupies Spec*v*P and carries inherent ergative case. The reanalysis of the trigger (i.e. the absolutive argument in languages like Tagalog) as the grammatical subject in the non-active voices is hinted by the use of *yaṃ* (*yang* in Modern Malay) with subjects in Old Malay, which is a cognate of Tagalog absolutive marker *ang*, as shown below:

(423) a. *Çānti yaṃ uraṃ ni-galar=ku maṃ-rakṣa di kāmu.* (Nonverbal)
blessed YAṂ people PASS-order=1 ACT-protect PREP 2
'Blessed are the people whom I ordered to watch over you.'
(Telaga Batu inscription – 700 AD)

b. *Tatkālā=ña yaṃ maṃmaṃ sumpaḥ ini ni-pāhat…* (Passive Verb)
during=3 YAṂ imprecation curse PROX PASS-engrave
'When the imprecation of this curse was engraved…'
(Kota Kapur inscription – 686 AD)

c. *Pulaṃ ka iya muaḥ yaṃ doṣā=ña vuat=ña ǰāhat inan.* (Unaccusative Verb)
return to 3 OPT YAṂ sin=3 do=3 evil DIST
'May the sins of those evil deeds return to them.'
(Kota Kapur inscription – 686 AD)

d. *Yaṃ mitrā=ña tuvi ǰāṅan ya kapaṭa.* (Unergative Verb)
YAṂ companion=3 verily PROH 3 deceive
'As for his companion, verily may he not deceive.'
(Talang Tuwo inscription – 684 AD)

Yaṃ was already phasing out and was also used as a sort of definite marker on DPs in general at the time (Mahdi, 2005), which resembles the definitising effect of the absolutive marker in Tagalog. The case-marking use of this element has since been lost, but survives in Modern Malay as a topic marker, as in (324), repeated below:

(424) *Yang dia, sedikit pun dia tak kesah.*
TOP 3.SG little.bit also 3.SG NEG care
'As for him, he didn't even care the slightest bit.'
(Yap, 2011, p. 4)

Ultimately, the development of the antipassive into the transitive and the emergence of the true passive from the passive-patient split allowed the subject of a transitive clause to align with the subject of an intransitive clause, hence the reanalysis of the syntactic alignment of the language from ergative to accusative. Given that structural case had become available to objects in transitive clauses and the unaccusative passive construction had emerged, a new nominative-accusative alignment had started to develop in the language.

To summarise, as the most conservative language, Tagalog retains the antipassive and ergative transitive. On the other hand, the antipassive in Malay and Malagasy has evolved into the transitive. As for the ergative transitive construction, it is only in Malay that the construction has evolved into the passive construction. Although Malay has achieved a majorly accusative alignment, ergative syntax survives in the patient voice, in which the patient corresponds to the syntactic subject of the clause, but the agent retains argumenthood. Because of the retention of the patient voice in Modern Malay, the change in the alignment of Malay is argued by Aldridge (2011a) to be ongoing.

7.1.2 Change in Word Order and Reanalysis of the Trigger as the Subject

As we have seen in Chapter 6, the development of the copulas in Malay was made possible by the change in word order in Classical Malay, which was set in motion by information-structural factors. Since the correlation between the change in word order and the emergence of overt copulas in Malay has been established to be information-structural in nature, it could very well be the case that this occurrence could hold of the other SVO languages in Table 34 that have overt copulas too. Indeed, Aldridge (2011a) suggests that the change in alignment in Austronesian languages is related to information structure: “the reanalysis of the absolutive as subject is mediated by topicalization” (p. 20).

However, it has been made clear in Section 6.1 that it was not until the end of the Classical Malay period that topicalisation and the subsequent loss of topic marker *pun* cemented the SVO order of Malay through the reanalysis of the topic as the canonical subject. As for the reanalysis of the trigger as subject, it is shown below that it had already occurred even prior to the reanalysis of topic as subject. In fact, the trigger had already achieved a stable footing in the specifier of TP, even in Old Malay. In the following examples, the subject, whether it be the agent or the patient, is preverbal across all voice alternations in Old Malay:

(425) a. *Ājñā=ṇḍa ku=juṃjuṃ nitya.* (Patient Voice)
instruction=3 1=uphold always
'I always uphold <u>his instructions</u>.'
(Mañjuśrīgṛha inscription – 793 AD)

b. *Iya maṅ-astuti guṇa=ṇḍa Dang Karayān Partapān.* (Active Voice)
3.SG ACT-praise use=3 D.
'He praises the virtues of Dang Karayan Partapan.'
(Gandasuli inscription – 832 AD)

c. *Dayang Aṅkatan… di-bari waradāna…* (Passive Voice)
D. PASS-give favour
'Dayang Angkatan... was given a favour…
(Laguna Copper-Plate inscription – 900 AD)

The movement of the trigger cum subject to SpecTP is evident in example (420a), repeated below, whose representation is provided below. As argued in Section 6.1, the verb-initial word order in Old Malay was obtained via V-T-C movement. Therefore, movement of the prohibitive mood marker *jāṅan* to C^0 in the example blocks movement of the verb to C^0, subsequently revealing the position of the patient *ya* between VoiceP and CP. The patient precedes the agent *savañakña yaṃ upasargga* and the passive verb *niknāi* but does not occupy the clause periphery. This constituent order entails that the patient occupies SpecTP.

(426) *Jāṅan ya ni-knāi sa-vañak=ña yaṃ upasargga.*
PROH 3.SG NI-afflict one-many=3 LIG misfortune
'May <u>all misfortunes</u> not afflict him.'
(Talang Tuwo inscription – 684 AD)

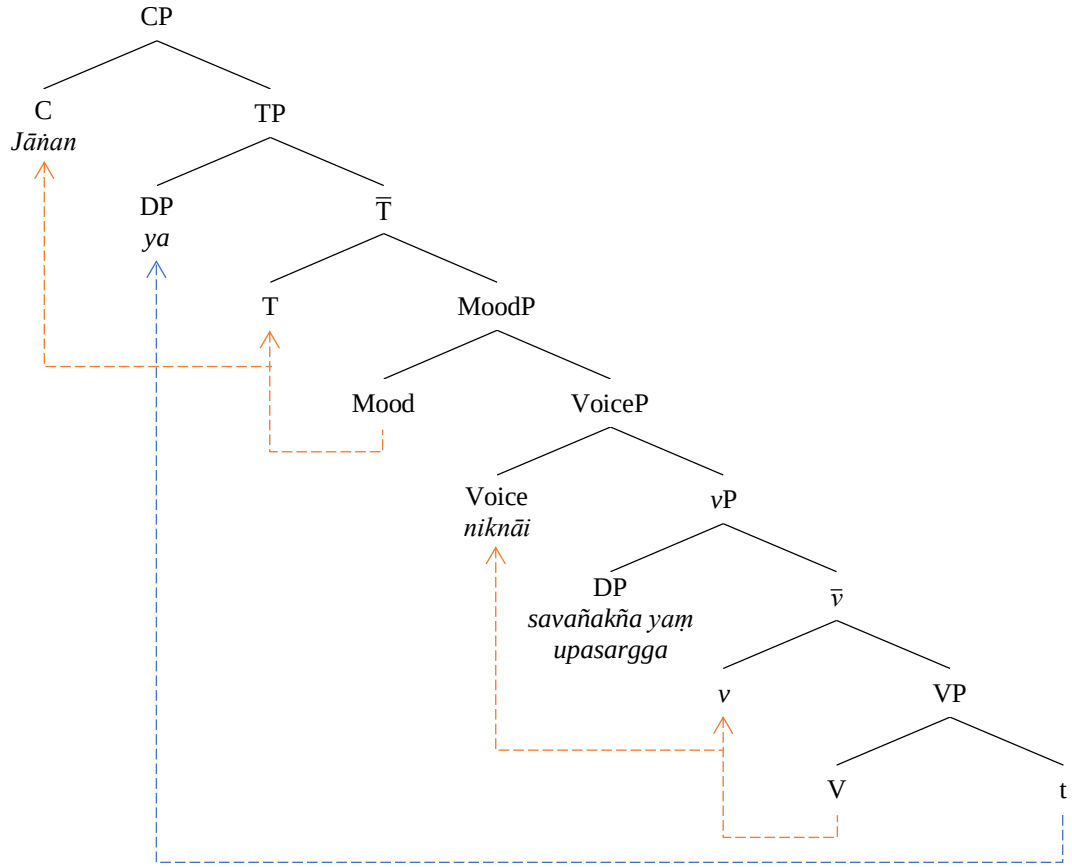


Figure 45: The Position of the Patient in SpecTP in the Ni- Construction in Old Malay

The finding that the trigger consistently occupied SpecTP across all voice alternations in Old Malay makes it possible to conclude that the trigger in Old Malay had already been reanalysed as the canonical subject, many centuries before the change in word order in Classical Malay from VSO to SVO via reanalysis of the topic as the subject. As such, Old Malay can be said to have technically obtained SVO word order, making it a stage-three language. However, it is only via the additional rule of V-T-C movement of the verb that VSO word order is maintained. Otherwise, the subject would have taken clause-initial position, as is apparent in examples (425) and in environments in which V-T-C movement is bled, such as conditional clauses and other subordinate clauses. Judging from the position of the trigger in SpecTP in both the patient voice and the *ni-* construction, the reanalysis of the trigger as the grammatical subject occurred prior to the passive-patient split. Otherwise, there would have been no need for the patient to move to SpecTP in either construction.

To be specific, there is an EPP feature on T^0. The trigger cum grammatical subject agrees with it during the process of obtaining nominative case and it is triggered to move to SpecTP by the EPP feature. Apparently, the movement of the patient in the patient voice and in the *ni*-construction entails that it gets assigned nominative case by T^0. Clearly, accusative case is not available in the passive voice, and this is also seen in the patient voice, as the patient moves to SpecTP in agreement with T^0. This further provides evidence that, even though the patient voice remains a vestige of ergative syntax in the language – as demonstrated by the position of the agent in Spec*v*P having inherent ergative case and its invisibility to the T^0 probe – absolutive case was no longer available in Old Malay. T^0 was consistently the nominative case assigner to the agent of the active voice, the patient of the patient voice, the patient of the passive voice, and the subject of intransitive clauses.

In contrast to Old Malay, in the Philippine-type languages such as Tagalog, the notion of subject is tied to the notion of trigger, which variably occurs in different positions in the clause, hence no single dedicated structural position for the so-called subject. Thus, there being no canonical subject and its privileged structural position precludes the reanalysis of topic as subject. For example, in Tagalog, the absolutive argument in the actor voice occupies Spec*v*P, whereas the absolutive argument in the patient voice occupies the complement position of the VP. They do not occupy one specific position within the clause or its periphery but remain in the positions in which they are base-generated. To illustrate, the absolutive in example (427) occurs within the VP layer, sandwiched between the oblique constituents and the ergative DP. Given that topics are presupposed, the indefinite interpretation of the oblique DPs indicates that they are not topics in a clause-peripheral topic position, and the same goes for the absolutive.

(427) *Ib‹in›igay ng lalaki ang libro sa babae sa aklatan.*
‹TR.PRF›give ERG man ABS book OBL woman OBL library
'The man gave <u>the book</u> to a woman at a library.'

Unlike Old Malay, the VSO order of Tagalog does not involve movement of the verb to C^0. According to Aldridge (2004), the verb moves only as far as AspP, which entails that SpecTP is empty. The following examples provide some word order facts concerning negation and polar interrogatives to support this claim:

(428) a. *K‹um›akanta ba ang lalaki?*
‹INTR.PRF›sing Q ABS man
'Does the man sing?'

b. *Hindi ba k‹um›akanta ang lalaki?*
NEG Q ‹INTR.PRF›sing ABS man
'Does the man not sing?'

c. **Hindi ba ang lalaki k‹um›akanta?*
NEG Q ABS man ‹INTR.PRF›sing
(Does the man not sing?)

In the formation of a polar interrogative, the interrogative marker *ba*, which should be conceived to head CP, attracts the verb to move C^0, as in (428a). When a negative polar interrogative is formed, the interrogative marker attracts the negator to $C^{0,}$ as in (428b). The attraction of the negator by the interrogative marker prevents the verb from moving farther from AspP. Should it be the case that the absolutive occupies SpecTP (in whatever voice, be it transitive or otherwise), we would expect it to intervene between the interrogative marker and the verb. However, this prediction is not borne out, as the absolutive occurs in post-verbal position, as shown in (428b). The ungrammaticality of the expected constituent order, as in (428c), thus confirms that the absolutive remains in *v*P and that SpecTP is not filled by any material. The syntactic representation of example (428b) is provided below:

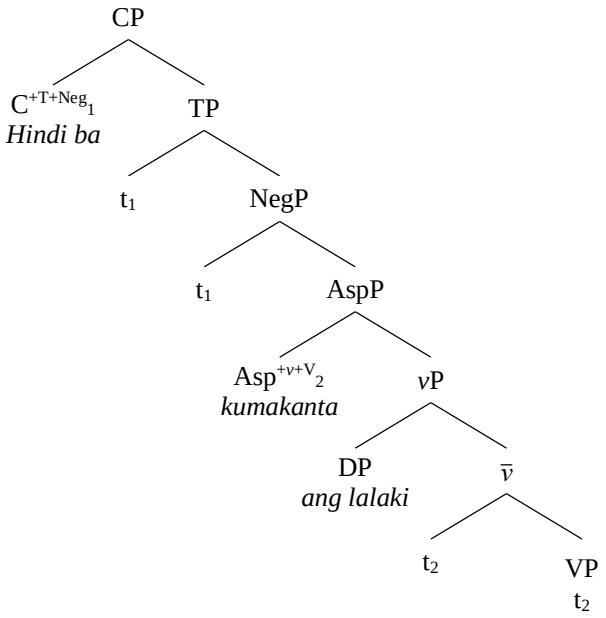


Figure 46: The Structure of an Intransitive Negative Polar Interrogative in Tagalog

Moreover, from the word order facts presented above, we find evidence that there is no structural position in the language that is dedicated to hosting the subject, owing to there being no notion of grammatical subject. This privileged subject position is important in the emergence of copulas in a language as, without it, topicalisation would not be able to trigger reanalysis of the topic as subject (and subsequent reanalysis of the resumptive pronoun as a copula). TP and its specifier are pivotal in the emergence of copulas in the Austronesian languages. As agreed by scholars such as Stassen (1997), Van Gelderen (2011), Givón (2015), and many others, pronominal copulas commonly grammaticalise from a resumptive pronoun in a topical construction, as undergone by Malay *ialah*. However, before spec-to-head reanalysis of the resumptive pronoun could occur, topic-to-subject reanalysis must precede it, and even before that, the notion of subject must first exist in the language.[100]

Assembling the findings in this chapter so far and the findings about grammaticalisation of pro-copulas, we can flesh out Figure 43 to yield the cline below. The cline is implicational such that one event may not occur without the preceding one having occurred.

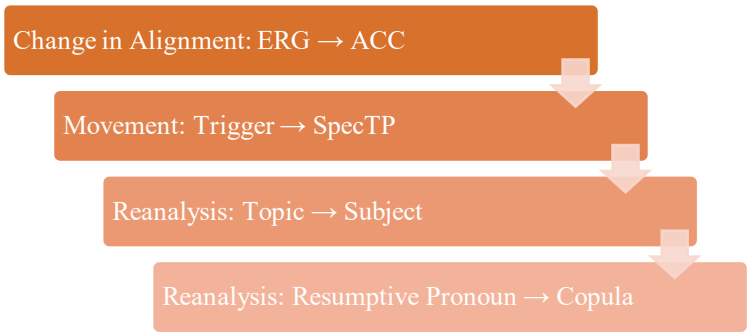


Figure 47: The Detailed Cline of the Emergence of Pronominal Copulas in Austronesian

[100] Certain scholars argue that the Philippine-type languages do not even project a TP, based on the assumption that there is no need for T^0 to inherit the features of C^0, as per the framework by Chomsky (2008) on C-T inheritance. According to Aldridge (2017), C^0 is directly responsible for assigning nominative case to the subject of a clause in Philippine-type Austronesian languages, which obviates the need for C-T inheritance. Moreover, C^0 does not possess an EPP feature unless for the sake of *wh*-movement. Therefore, there is no A-movement of the subject to the front of the clause, be it to SpecCP or SpecTP. A direct consequence of no A-movement is the predominantly verb-initial word order of the Philippine-type languages.

To connect this cline with the four stages in Table 39, we start off with an unchanged ergative language, which is in stage one. Stage two would involve the change in alignment, which allows the reanalysis of the trigger as the subject, owing to its movement to SpecTP by the EPP feature on T^0 with which the subject agrees during the process of being assigned nominative case. In stage three, the creation of this dedicated subject position makes way for change in word order from VSO to SVO, unless there is an additional rule in the syntax that forces the verb to move to the left periphery, as in the V-T-C rule in Old Malay. Lastly, stage four must involve information structure such that topicalisation structures spell out resumptive pronouns in nonverbal clauses and the topic gets reanalysed as the canonical subject, subsequently compressing the resumptive pronoun into a head that gets reanalysed as a copula.

In conclusion, the absence of copulas in the ergative languages in Austronesian is attributed to the syntactic alignment of the languages being ergative. The absence of the notion of subject and its dedicated position in ergative languages precludes any sort of reanalysis of dislocated topics into canonical subjects, which is a common precursor of the grammaticalisation of copulas from resumptive pronouns.

7.2 Pronominal Copulas

We have seen how left dislocation can pave the way for the reanalysis of a resumptive pronoun into a copula, following the reanalysis of the antecedent topic into the canonical subject, as in the case of Malay *ialah*. A similar trend can be said to have occurred in a number of Austronesian languages that have pronominal copulas. For instance, Soriente (2018) states that the copula *nah* in Punan Tubu can be derived from the demonstrative *inah*. In Cham, the pro-copula may co-occur with a demonstrative of the same form contiguously, meaning that the second instance of the lexeme does not play the role of a regular demonstrative. Blood (1977) claims that "the two phrases or clauses (of a topic-comment sentence) may be juxtaposed or may be linked by *năn* 'to be'. In the latter case, *năn* functions as the main verb of the simpler or complex sentence" (p. 63). Other than that, there are many other languages with demonstratives functioning as copulas, e.g. Javanese, Iban (Sea Dayak), etc.

Cham (Austronesian – Chamic)

(429) *Ông năn năn urang tôl.*
mister DIST COP person guest
'That gentleman is a guest.'
(Blood, 1977, p. 63)

Javanese (Austronesian)

(430) *Séng paling banter mlayu-né yo iku pe-menang-é.*
REL most fast run.AV-NMZ also DIST AG-win-DEF
'He who is the fastest in the run is the winner.'
(Kraußе, 2017, p. 54)

Iban (Austronesian – Malayic)

(431) *Besaki nya haram.*
adultery DIST prohibited
'Adultery is prohibited.'

The table below lists down several Austronesian languages that employ a demonstrative in copular clauses.

Language	Demonstrative
Cham	*năn*
Iban	*nya*
Javanese	*iku*
Malay	*itu*
Minangkabau	*itu*
Punan Tubu	*nah*

Table 40: Austronesian Languages with Demonstrative Copulas

Of course, there remain many accusatively aligned languages that do not have pronominal copulas. Although the correlation between SVO word order and the emergence of pronominal copulas has been established, it is not the case that all accusative Austronesian languages with the SVO word order must have pronominal copulas. This condition is a corollary of the finding in the previous section that the SVO order is not necessarily achieved via information-structural means. As in Old Malay, the SVO order is achieved simply by the projection of TP and the requirement by the EPP that the subject move to SpecTP. It is only by an additional V-T-C movement operation that the verb-initial order is achieved in Old Malay, as evidenced in clauses in which V-T-C movement is banned, e.g. conditional clauses.

In light of this, not all the accusative SVO Austronesian languages achieved the SVO order via topicalisation. In languages that did so without topicalisation, there should have been no spec-to-head reanalysis of some resumptive pronoun into a copula, considering that there was no dislocation and resumption to begin with. Consequently, such languages lack pronominal copulas. In fact, the copulas of most of the languages examined are verbal, as tabulated in Table 41 in Section 7.3. However, the development of pronominal copulas is more complex than that as there are many more factors that could have affected pronominal copularisation, even if the SVO order is achieved through topicalisation, e.g. the lack of resumption following dislocation or topicalisation, the lack of specialisation of the resumptive pronoun in copular clauses, etc. Without a detailed examination of the diachrony of such languages, e.g. Matbat, Teop, Nakanai, etc., it cannot be confirmed whether the languages achieved SVO order via information-structural means or otherwise.[101]

7.2.1 No Subject, No Pronominal Copula in Tagalog

In contrast to the languages in Table 40, there is no pronominal copula in Tagalog. On account of there being no notion of grammatical subject in the language, it is impossible for a topic to reanalyse as the canonical subject, which then precludes reanalysis of a resumptive pronoun as a copula. However, for the sake of confirming the hypothesis that the TP is central in pronominal copularisation, let us attempt to apply the processes undergone by Malay copula *ialah* to copular clauses in Tagalog.

---

[101] Matbat does have what Remijsen (2010) calls a *copular clitic*. However, as illustrated below, the *i-* element appears to be more generally a 3rd person agreement marker, which is also used with verbal predicates.

(xxviii) a. *Ak-nɔ^{3}ŋ* *i-ma^{3}ntri^{3}.* (Nonverbal)
1.SG-brother 3.SG-health.visitor
'My brother is a health visitor.'

b. *I^{21}* *ga^{21}* *i-bɔ^{3}t.* (Verbal)
3.SG just.now 3.SG-arrive
'He has arrived just now.'
(Remijsen, 2010)

Interestingly, the grammaticalisation of agreement markers is also known to involve dislocation and resumption, as do pronominal copulas, as argued by Givón (2015). Therefore, it could be the case that, instead of specialising in copular clauses and grammaticalising into copulas, the *i-* element in Matbat originated as a resumptive pronoun and underwent spec-to-head reanalysis across the board and became agreement markers.

First, note that copular clauses in Tagalog are ordered Predicate-Subject:

(432) *Doktor ang lalaki.*
doctor ABS man
'The man is a doctor.'

Applying dislocation to a copular clause in Tagalog does not produce a result that feeds reanalysis of the resumptive pronoun into a copula. As illustrated in example (433), the resumptive pronoun does not occur between the topic and the predicate, i.e. in the *compression zone* in which it could be pressured to be reanalysed as a head following reanalysis of the topic as the canonical subject, if that even takes place. SVO word order is important in the copularisation of resumptive pronouns as the reanalysis of the topic as subject feeds spec-to-head reanalysis of the pronoun by compressing, as it were, the phrasal constituent into a head. Therefore, dislocation and resumption in Tagalog copular clauses do not allow spec-to-head reanalysis of the resumptive pronoun, hence no copularisation.

(433) *Ang lalaki*$_1$ *doktor siya*$_1$.
ABS man doctor 3.SG.ABS
'That man, he is a doctor.'

In addition to that, there is a topical construction that involves an operation called *ay*-inversion, which makes use of a sort of topic marker, *ay*. It is characterised by the movement of the subject to the front of the clause and the spell-out of *ay*, producing apparent inversion of the subject and the predicate. In the case of a verbal clause, it is the trigger that is preposed. To illustrate, the following verbal and nonverbal examples exhibit *ay*-inversion.

(434) a. *Ang lalaki*$_1$ *ay doktor* t$_1$. (Nonverbal)
ABS man TOP doctor
'The man is a doctor.'

b. *Ang lalaki*$_1$ *ay gwapo* t$_1$.
ABS man TOP handsome
'The man is handsome.'

c. *Ang lalaki*$_1$ *ay s‹um›ayaw* t$_1$. (Verbal)
ABS man TOP ‹INTR.PRF›dance
'The man danced.'

d. *Ang lalaki*$_1$ *ay s‹in›ampal ng babae* t$_1$.
ABS man TOP ‹TR.PRF›slap ERG woman
'A/the woman slapped the man.'

Seeing that the *ay* element resembles a copula that intervenes between a subject and a nonverbal predicate, one might be inclined to hypothesise that the *ay* element in the examples above could be a copula, especially considering its medial position in the Top-*ay*-Pred order, comparable to SVO. As a matter of fact, it is very commonly referred to as a copula in pedagogical grammars (e.g. Totanes, 1745); however, as Grant (2009) states, "such authors favoured this admittedly stylistically marked construction because they felt that sentences containing the construction corresponded point by point to a large extent to sentences with similar semantic content in Latin and Spanish which contained obligatory copulas" (p. 234).

On the contrary, it is characterised as a topic marker since it also occurs in verbal clauses.[102] By definition alone, a copula links a subject to a nonverbal predicate, and a topic marker roughly corresponds to something that links a topic to a comment, regardless of whether the clause is verbal or nonverbal. Since copulas are not generally used in verbal clauses, such topic markers cannot be construed as copulas.

[102] In other languages as well, a topic marker is often referred to as a copula. Paul (2008) argues that *dia* in Malagasy is a topic marker. The fact that *dia* also surfaces when constituents in non-copular clauses are topicalised is one among a few pieces of evidence.

(xxix) *Ny mpianatra dia mamaky teny.*
DET student TOP AT.read word
'The students, they are reading.'

In Roviana, there is the element *si*. Although Blust (2013) glosses the element *si* as a copula, it is arguably a topic marker as it also occurs in verbal clauses, as shown in the example below by Schuelke (2020). Waterhouse (1928) describes *si* as an assertive element and that "there is no verb to be, but in some constructions… *si* [has] almost the force of the copula" (p.248).

(xxx) *Sa vivinei si ele toz=ia rau.*
ART story SI PRF tell=3.SG.OBJ 1.SG
'I already told the story.'

In Dusun, the element *nopo-nga* (loosely translated as *thing that is*) surfaces in topical constructions in both verbal and nonverbal clauses and is called a copula by some (see Price, 2007).

(xxxi) *Iti nopo-nga pingungaranan do Bundu Dusun.*
PROX TOP named INDEF B.
'This language is named Bundu Dusun.'

7.3 Verbal Copulas

So far, the discussion has revolved around pronominal copulas, which emerge as a result of the reanalysis of a resumptive pronoun as the head of TP. Common as it may be, this strategy involving dislocation and spec-to-head reanalysis is by no means the only way for a copula to emerge in a language. Alongside pro-copulas, there are verbal copulas of the likes of English *be*, which descended from Proto-Indo-European *b^huH- "to be, exist, grow" (Ringe, 2017). Indeed, in many of the Austronesian languages examined, there are verbal copulas. The following table shows the types of copulas present in the Austronesian languages examined:

Language	Verbal	Pronominal	Other	Origin
Malay	*adalah*	*ialah; itu*		EXIST-COM; 3.SG-COM; DIST
Iban		*nya*		DIST
Minangkabau	*adolah*	*iyolah; itu*		EXIST-COM; 3.SG-COM; DIST
Urak Lawoi'	*jadi*			'Become'
Cham		*năn*	*ciə*	DIST; ?
Jarai	*jing*			'Become'
Tsat			*si*	Mandarin copula *shi* (DEM)
Javanese		*iku*		DIST
Moklen			*pɪn*	Thai copula *pen* 'be alive'
Rejang	*adeba*			EXIST
Batak		*ém-kap*		PROX-EMPH
Ambel	*be*			'Become'
Biak	*is*			?
Irarutu			*rau*	?
Lelepa	*pi*			Make/do
Daakaka	*i*			Make/do
Bierebo	*ve/pe*			Make/do
Erromangan	*ete*			'Live/stay/dwell'
Puyuma	*amau*			?

Table 41: The Origins of the Copulas in Austronesian Languages

At first glance, not many of the copulas in the table share a similar form. Although there are copulas shared by very closely related languages such as Malay and Minangkabau, as well as Lelepa and Bierebo, there is no similarity between groups of languages. Even in some closely related languages, there remains variation, as in the case of Iban and Urak Lawoi', close relatives of Malay and Minangkabau, all of which comprise Malayic. Interestingly, Rejang presents a curious case as it is as distant to Malay as Malay is to Javanese, but it shares *ada* as the root of the copula.

The lack of similarity disallows reconstruction of a proto-form and suggests that the origins of the copulas in different Austronesian languages are not uniform. There not being a single proto-form that can tie together all the different copulas makes for a high degree of variation as many languages developed their own copulas individually. Certainly, Starosta et al. (1982) state that there were no copulas in Proto-Austronesian, which is why most languages across Austronesian lack overt copulas altogether. Because a copula cannot be reconstructed for Proto-Austronesian, it can be said that the copulas of the Austronesian languages developed independently and relatively late. In the case of the Malayic languages, *adalah* and *ialah* developed during the Classical Malay era (circa 14th century – 19th century).

As for the Southern Oceanic languages in the table (Lelepa, Daakaka, Bierebo), the copulas seem to be cognates, which suggests that there is a proto-form. The copula in Erromangan seems to be different as it is slightly more distantly related than the three other languages. The three other languages are classified as Central Vanuatu languages, whereas Erromangan is a Southern Vanuatu language. Clark (1985) suggests that the copulas in the languages of Central Vanuatu can be traced to Proto-Central-Vanuatu **vei*, which itself originates from Proto-Oceanic **pai* (make/do). Other than these two branches, there do not seem to be cognates shared by slightly more distantly related Austronesian languages, which again suggests that the copulas of the Austronesian languages developed relatively late.

One major question that arises is why the Philippine-type languages do not have verbal copulas. It is the strict observance of the Austronesian voice system that disallows verbal copularisation, as explained in the following section.

7.3.1 Posture Verbs and Light Verbs

Copulas are known to frequently develop from posture verbs such as *sit* and *stay*, whilst light verbs such as *make* or *do* often turn into auxiliaries, which in turn may develop into copulas (Anderson, 2006). In fact, the Southern Oceanic languages in Table 34 are found to have develop their copulas from such verbs. Arguably, the same development could have occurred in an ergative Austronesian language like Tagalog; however, copularisation of a posture or light verb is not attested in any of the ergative languages in the table. It can be hypothesised that this instance of copularisation is not

possible in Austronesian languages that strictly observe the Austronesian voice system due to two important properties: the form of the verb changes according to the trigger; the argument structure of the verb changes according to the trigger.

To elaborate, suppose that the word *upo* (sit) grammaticalised into a copula and replaced the null copula in Tagalog. Depending on the thematic role of DPs present in the clause, the verb is expected to undergo morphological changes. Specifically, *upo* may inflect for the agent, locative, and benefactive voices, as shown in (435); however, it may not inflect for the patient voice because it is an unergative verb. On the other hand, copular clauses are unaccusative.

Tagalog (Austronesian – Philippine)

(435) ‹*Um*›*upo* *ang lalaki* *sa* *sofa.* (Actor Voice)
‹INTR.PRF›sit ABS man OBL sofa
‘The man sat on a/the sofa.’

P‹in›aupo *ng babae* *ang lalaki* *sa* *sofa.* (Benefactive Voice)
‹TR.PRF›sit ERG woman ABS man OBL sofa
‘A woman sat for the man on a/the sofa.’

‹*In*›*upu-an* *ng lalaki* *ang sofa.* (Locative Voice)
‹TR.PRF›sit-APPL ERG man ABS sofa
‘A man sat on the sofa.’

None of the voices are compatible with the thematic role of the subject of a copular clause. The difference in thematic roles between the unergative posture verb and the unaccusative copula poses a problem to a language like Tagalog, considering that the Philippine-type languages assign high importance to thematic role information and the associated forms of the verb. Conversely, the accusatively aligned Indonesian-type languages are very much morphologically impoverished, which obviates the importance of thematic role information. As reported by Foley (2005), “Austronesian Oceanic languages typically do not have an overt morphological contrast between unergative and unaccusative intransitive verbs” (p. 398). The result is the possibility of verbs with agent thematic roles to develop into copulas, as seen in Erromangan, Lelepa, Daakaka, Bierebo, etc. Therefore, there is no possibility for any posture verbs or even unergative verbs in general to develop into copulas in Tagalog and other Philippine-type languages.

The copularisation of posture verbs typically involves the reinterpretation of locative adjuncts as the main predicate. In time, the verb undergoes bleaching until it reaches the extent that it loses its locative connotation and becomes compatible with nonverbal predicates of other syntactic categories (Hengeveld, 1992). Reinterpretation of the adjunct is made possible by the intransitive nature of posture verbs. However, the benefactive and locative forms of *upo*, *pinaupo* and *inupuan* respectively, are transitive, taking the ergative argument and the *ang*-marked benefactor or location as applied arguments. The transitivity of the verb poses a problem for the copularisation of the verb, as copular clauses are strictly intransitive. Therefore, copularisation is impossible, on account of the transitive nature of the benefactive and locative voices.

Based on the failure of the development of the postulated *upo* into a copula, the different voices of the ergative languages and their associated thematic roles are therefore a major reason why there are no verbal copulas in the ergative languages. Perhaps if the verb lost its voice affixes and became morphologically invariable, it could develop into a copula.

7.3.2 Verbs of Becoming

Notice that a few of the verbal copulas in Table 41 developed from verbs of becoming. For example, the copula in Urak Lawoi' evolved from inchoative *jadi*. It can be said to have undergone semantic bleaching to a certain degree, allowing its use in the non-inchoative sense.

Presumably, a language like Tagalog might be able to resort to dispensing with the inchoative aspect on the verb *maging* (become). However, it is well-known that the marking of aspect on the verb in the Philippine-type languages is very robust. In fact, almost all of the ergatively-aligned languages in Table 34 morphologically encode aspect on the verb. It seems that semantic bleaching of the inchoative verb does not occur in languages like Tagalog owing to the richness of morphological marking on the verb in the Philippine-type languages, especially with respect to aspect, as opposed to the impoverished morphology of the Indonesian-type languages. Seeing that the ergative Austronesian languages may not develop copulas from aspect-related verbs of becoming, tense and aspect are clearly notions that are not suitable for the development of copulas in such languages.

Besides, developing a full-fledged copula from an inchoative copular verb can be said to require more effort on the part of the listener, as the same overt form would then carry two interpretations – episodic vs. atemporal. Indeed, the copulas in Urak Lawoi', Jarai, and Ambel carry both episodic and atemporal interpretations. To illustrate, the following examples contrast a temporal reading and an episodic reading of *jadi*, *jing*, and *be*:

Urak Lawoi' (Austronesian – Malayic)

(436) a. *Nya jadi kaka ku.* (Atemporal)
3.SG COP brother 1.SG
'He is my brother.'

b. *Nya jadi bercac.* (Episodic)
3.SG become bedbug
'He became a bedbug.'
(Hogan & Pattemore, 1988)

Jarai (Austronesian – Chamic)

(437) a. *Ñu anŭn jing đah-kơmơi tuai.* (Atemporal)
3.SG MED COP woman visitor
'She is a foreigner.'

b. *Ñu amra jing hĭ bơnai kâo.* (Episodic)
3.SG FUT become PRT wife 1.SG
'She will certainly become my wife.'
(Jensen, 2013)

Ambel (Austronesian – South Halmahera-West New Guinea)

(438) a. *Ia ŋ=be mákay bábo. rín.* (Atemporal)
3.SG.AN 3.SG.AN=COP child young CONT
'He is still a youngster.'

b. *Bey ne aŋ=be i-pil pórin.* (Episodic)
sago ART 3.SG.INAN=become 3.INAN-price NEG.CONT
'Sago had not yet become expensive.'
(Arnold, 2018)

To dispel the ambiguity, extra effort needs to be made, which goes against overarching principles of economy. For example, the use of the verb in the atemporal sense is restricted to copular clauses whose subject or predicate is of a certain type. In Urak Lawoi, for *jadi* to be atemporal, the post-copular DP must be referential (Hogan

& Pattemore, 1988). Ambel relies on a different strategy involving the animacy of the subject, whereby *be* is used in the atemporal sense only with animate subjects (Arnold, 2018). The differences in animacy or referentiality are likely to prevent a change-of-state interpretation. Furthermore, the atemporal sense is restricted to nominal predicates, which are less likely to carry a change-of-state interpretation, i.e. less dynamic. This restriction is exemplified in copular clauses in Jarai, whereby the copula is not used with adjectival predicates, lest it be interpreted as an inchoative verb, as stated by Jensen (2013).

7.4 Summary

To summarise, Austronesian alignment is, in one way or another, responsible for the lack of overt copulas in a great number of the Austronesian languages, particularly those that are categorised as the Philippine-type languages. The notion of subject and the privileged subject position is absent in these languages, which is a result of the way in which the Philippine-type languages encode different constituents as the trigger, in accordance with Austronesian alignment. There being no grammatical subject obviates the reanalysis of the topic as the subject and interferes with pronominal copularisation, the development of which is reliant on the notion of subject. On the other hand, the word order of most of the Indonesian-type languages examined is SVO, which is obtained following change in alignment from ergative to accusative. A consequence of this change in alignment is the emergence of the grammatical subject and the privileged subject position. The subject position gets filled by a resumptive pronoun in topical left-dislocation constructions and the resumptive pronoun undergoes spec-to-head reanalysis following reanalysis of the topic as the canonical subject, resulting in the pronoun grammaticalising into a copula.

As for the effects of the Austronesian voice system, given that thematic structure is of great importance in the Philippine-type languages, posture verbs and light verbs are unable to develop into copulas due to a clash between the unergativity of posture/light verbs and the unaccusativity of copulas. The agentivity indicated by the actor-voice marking on the posture/light verb is incompatible with the non-agent argument in copular clauses. There is no patient voice for posture verbs, which could be compatible with the thematic structure of copular clauses, seeing that posture verbs

are unergative. Furthermore, the transitive nature of the locative and benefactor voices conflicts with the intransitive nature of copular clauses.

Lastly, the morphologically rich nature and the robust aspectual marking of the verb of the Philippine-type languages disallows verbs of becoming from losing their inchoative aspect to develop into copulas.

Chapter 8: Conclusion

As we have seen, the study of nonverbal predication in Malay has not ventured very far. Noticeably, research on copular clauses in Malay is wanting on all levels of linguistic representation. My intention has been to bridge this gap by examining copular clauses in Malay to achieve a profound understanding of the construction and engender a grammar of Malay copular clauses of a satisfactory breadth and depth.

In examining copular clauses from synchronic, diachronic, and typological perspectives, this thesis has been able to uncover various important findings that not only concern the copulas and their immediate linguistic environments, but also the Austronesian language family at large. Even though I have taken nonverbal predication to be the focus of this thesis, the findings pertaining to other facets of the grammar may be relevant to linguists in general. For example, my examination of extraction from copular clauses in Malay has revealed that, similar to the voice markers, the *yang* head also carries the feature [*u*Trigger] that identifies a constituent as the trigger of the clause. Considering this, nothing may be clefted other than the trigger. Conversely, the null C^0 head allows free extraction, as supported by the possibility of extracting either arguments or adjuncts. More broadly, the typological survey has revealed that the constellation of several syntactic properties of the ergatively aligned Austronesian languages such as their word order, voice-marking properties, and argument structure, each of which influences the other, conspire to prevent the emergence of copulas. Therefore, the findings presented herein may be informative to Austronesianists working on both verbal and nonverbal predication in other Austronesian languages.

The lack of research on copular clauses in Malay has made new discoveries possible, and with new findings, previous claims about the copulas and phenomena or constructions related to them can be revisited, reanalysed, and reaffirmed or refuted. Since this thesis covers quite a broad range of topics related to copular clauses, from morphology to syntax, and synchrony to diachrony, it is hoped that linguists may draw from the observations, results, and generalisations from this thesis to supplement their own research and propel the study of nonverbal predication and Malay grammar in general to newer and unexplored horizons.

8.1 Revisiting the Research Questions

After having discussed at great length copular clauses in Malay with respect to their synchrony, diachrony, and typology, we can finally revisit the research questions posed in Section 1.2.

To what extent are the copulas predictable from their component parts and the way they are combined?

The copulas do not transparently reflect the morphemes that they are believed to have evolved from. Specifically, *ialah* is no longer a focused pronoun, whilst *adalah* is no longer a focused existential verb. The inseparability of *lah* from *ia* and *ada* indicates that the two copulas are monomorphemic in their current form. The demorphologisation of *lah* has caused the former morpheme to fuse together with *ia* and *ada*. Therefore, the two copulas are no longer predictable from their component parts as their behaviours are completely different from 3rd person pronoun *ia*, existential verb *ada*, and their combinations with focus marker *lah*.

What is the syntactic structure of a copular clause in Malay?

Based on facts pertaining to word order, the copulas are T^0 heads that are base-generated in TP. They are obviously not verbs as *adalah* may co-occur with linking verbs. Although *ialah* never co-occurs with verbs due to its non-predicational nature, the two copulas are analysed to be one and the same head but with different properties. Furthermore, both copulas occur in complementary distribution with epistemic modals, which precede negator *tidak* and contrastive marker *bukan*, confirming their status as auxiliaries and the heads of TP.

As for the structure below the verb, the postulation of a PredP is supported by the possibility of the $Pred^0$ head to surface in certain environments and the possibility of a pronoun in its specifier to encliticise onto the verb, which entails that the verb immediately selects the PredP, given that cliticisation must be local. Therefore, the structure of a copular clause in Malay is as such:

$$[_{TP}\ \text{COPULA}\ [_{VoiceP}\ [_{vP}\ [_{VP}\ [_{PredP}\ [\text{SUBJECT}]\ [_{Pred'}\ Pred^0\ [\text{PREDICATE}]]]]]]]$$

What governs overt encoding of the copulas in Malay?

Aspect governs overt encoding of the copula. It is only in copular clauses that have an atemporal interpretation, as opposed to a temporally bound interpretation, that overt encoding of the copula is possible. The temporally bound interpretation is analysed to be introduced by an inner aspectual projection located between *v*P and VP, which clashes with the atemporal property of the copula. Given that aspect also interacts with the notion of eventuality and the distinction between stage-level and individual-level predicates, the copula is also affected by them such that the use of overt copulas is permitted with Kimian states and individual-level predicates. Therefore, the inner AspP is also projected in clauses that contain the opposing Davidsonian states and stage-level predicates, which do not allow overt encoding of the copula.

How are cleft constructions in Malay derived?

Given the discovery of a matrix copular clause above the visible part of the cleft, the cleft clause introduced by *yang* is shown to raise to the subject position of the matrix copular clause to form the clausal subject of a pseudocleft. Otherwise, the subject could be occupied by a null expletive. The null expletive prevents raising of the clause, which derives a normal cleft.

When did the copulas emerge in Malay and how did they grammaticalise?

Whilst no visible copula was used in the Old Malay period (7^{th} ~ 14^{th} century), what have now become copulas *ialah* and *adalah* had only emerged in the beginning of the Classical Malay period. Although both *ia-lah* and *ada-lah* were used in copular clauses throughout the Classical Malay period (14^{th} ~ 19^{th} century), they functioned as an an argument and auxiliary marking the topic-comment boundary respectively. It was only towards the end of the Classical Malay period (circa the 19^{th} century) that the copular use of *ia-lah* and *ada-lah*, as opposed to their use to mark new information and the comment portion of the clause, became apparent.

As opposed to the assumption that the focus marker *lah* was involved in the grammaticalisation of the copulas *ialah* and *adalah*, I have shown that it was the variant of *lah* as a comment marker that combined with 3rd person *ia* and dummy auxiliary *ada* to develop the respective copulas. The decline in movement of the verb to clause-initial position effected a decline in the use of *lah*, and with it, *ada-lah*. However, prior to that, the *lah* morpheme had fused onto the *ada* root in nonverbal clauses and *ada-lah* had specialised in nonverbal clauses, which allowed its use as a copula to remain unaffected by the change in word order towards the end of the Classical Malay period. At the same time, the rise in topical constructions and the subsequent loss of topic marker *pun* allowed the topic to be reanalysed as the canonical subject, catalysing the VSO to SVO change. This change forced the resumptive pronoun *ia*, which was spelt out in topical nonverbal clauses to host comment marker *lah*, to undergo spec-to-head reanalysis and grammaticalise into a copula heading TP.

What factors condition the distribution of copulas across Austronesian?

SVO word order is very conducive to the emergence of copulas across Austronesian. This is because pronominal copulas often develop from topical constructions in which the left-dislocated topic gets reinterpreted as the canonical subject, and the resumptive pronoun that was previously in the subject position gets compressed into a head via spec-to-head reanalysis and undergoes grammaticalisation into a copula. In the ergatively-aligned Austronesian languages, the predominantly VSO word order bleeds spec-to-head reanalysis of the resumptive pronoun as it does not occur between the topic and the predicate, hence no grammaticalisation.

Also important with respect to copularisation is the voice system of the ergatively-aligned languages, which morphologically reflects the argument structure of the verb. Posture verbs and light verbs cannot develop into copulas due to this voice marking. Given that copular clauses are unaccusative, the encoding of the agent on the unergative posture and light verbs in the actor voice clearly presents a problem. Additionally, the other voices such as the locative voice and the benefactive voice cannot be used as the absolutive is encoded in the structure as an applied argument, which derives a transitive clause. This clashes with the strictly intransitive nature of copular clauses.

References

Main References

Abdullah, I. H. (1993). *The Semantics of the Modal Auxiliaries of Malay*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Abu Bakar, H. (2009). Code-switching in Kuala Lumpur Malay: The "Rojak" phenomenon. *Explorations*(9), 99-107.

Adelaar, K. A. (1992). *Proto-Malayic: The reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: The Australian National University.

Adelaar, K. A. (1993). The Internal Classification of the Malayic Subgroup. *Bulletin of the School of Oriental and African Studies, University of London, 56*(3), 566-581. Retrieved from http://www.jstor.org/stable/620695

Adelaar, K. A. (2004). Where does Malay come from? Twenty years of discussions about homeland, migrations and classifications. *Journal of the Humanities and Social Sciences of Southeast Asia, 160*(1), 1-30. doi:https://doi.org/10.1163/22134379-90003733

Adelaar, K. A. (2005). Malayo-Sumbawan. *Oceanic Linguistics, 44*, 357-388. doi:10.1353/ol.2005.0027

Ajamiseba, D. C. (1983). *A classical Malay text grammar : insights into a non-western text tradition*. Canberra, Australia: Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

Akmajian, A. (1970a). *Aspects of the grammar of focus in English.* (Ph.D. Dissertation). Massachusetts Institute of Technology,

Akmajian, A. (1970b). On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry, 1*(2), 149-168.

Aldridge, E. (2004). *Ergativity and word order in Austronesian languages.* (Ph.D. Dissertation). Cornell University, Ann Arbor. ProQuest Dissertations & Theses Global; ProQuest One Literature database. (3140799)

Aldridge, E. (2007). Phase-based account of extraction in Indonesian. *Lingua, 118*(10), 1440-1469.

Aldridge, E. (2011a). Antipassive in Austronesian alignment change. In D. Jonas, J. Whitman, & A. Garrett (Eds.), *Grammatical Change: Origins, Nature, Outcomes*: Oxford University Press.

Aldridge, E. (2011b). Neg-to-Q: The historical origin and development of question particles in Chinese. *The Linguistic Review, 28*(4), 411. doi:https://doi.org/10.1515/tlir.2011.012

Aldridge, E. (2017). *ϕ-feature competition: A unified approach to the Austronesian extraction restriction.* Paper presented at the Proceedings of the 52nd meeting of the Chicago Linguistic Society (CLS).

Alharbi, B. (2017). *The Syntax of Copular Clauses in Arabic.* (Ph.D. Dissertation). University of Wisconsin-Milwaukee,

Alwi, H., Dardjowidjojo, S., Lapoliwa, H., & Moeliono, A. M. (2003). Tata bahasa baku bahasa Indonesia.

Anderson, G. (2006). *Auxiliary Verb Constructions*: Oxford University Press.

Arbak Othman. (1987). Masalah Kenahuan dan Kesalahan Bahasa [problems in grammaticality and errors in language]. In Abdullah Hassan (Ed.), *Kesalahan Bahasa dalam Bahasa Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Arche, M., Fábregas, A., & Marín, R. (2017). Towards a unified treatment of Spanish copulas. In S. Perpiñán, D. Heap, I. Moreno-Villamar, & A. Soto-Corominas (Eds.), *Romance Languages and Linguistic Theory 11: Selected papers from the 44th Linguistic Symposium on Romance Languages (LSRL)* (pp. 33-52). London, Ontario.

Arche, M., Fábregas, A., & Marín, R. (2019). Main questions in the study of copulas. In M. J. Arche, A. Fábregas, & R. Marín (Eds.), *The Grammar of Copulas Across Languages* (Vol. 73, pp. 1-30). Oxford: Oxford University Press.

Ariel, M. (2000). The development of person agreement markers: From pronouns to higher accessibility markers. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 197-260). Stanford, CA: Center for the Study of Language and Information.

Arka, I. W. (2013). *Nonverbal predicates in Austronesian and Papuan languages: an LFG perspective*. Paper presented at the Proceedings of the 6th International Seminar on Austronesian and non-Austronesian Languages and Literatures.

Arka, I. W., & Manning, C. (1998). *Voice and grammatical relations in Indonesian: A new perspective*.

Arnold, L. M. (2018). Grammar of Ambel, an Austronesian language of Raja Ampat, west New Guinea.

Asraf. (1988). *Sebutan baku bahasa melayu berdasarkan prinsip fonemik* (Vol. 7). Petaling Jaya: Sasbadi.

Awang Sariyan. (1984). *Isu-Isu Bahasa Malaysia [issues in Bahasa Malaysia]*. Petaling Jaya: Fajar Bakti.

Bailey, L. R. (2013). *The syntax of question particles.* (Ph.D. Dissertation). Newcastle University, Retrieved from https://kar.kent.ac.uk/80123/

Baker, M. (1985). The Mirror Principle and Morphosyntactic Explanation. *Linguistic Inquiry, 16*(3), 373-415. Retrieved from http://www.jstor.org/stable/4178442

Baskoro, S. (2016). Stone Heritage of Indonesia. In H. Kato, A. Reedman, Y. Shimazaki, T. Uchida, N. T. M. Ngoc, & A. Surinkum (Eds.), *Stone Heritage of East and Southeast Asia*. Ibaraki, Japan: Geological Survey of Japan (GSJ).

Batllori, M. (2011). Grammaticalization of ser and estar in Romance. In D. Jonas, J. Whitman, & A. Garrett (Eds.), *Grammatical Change: Origins, Nature, Outcomes*: Oxford University Press.

Bellwood, P. (2007). *Prehistory of the Indo-Malaysian Archipelago*: ANU Press.

Benjamin, G. (2009). Affixes, Austronesian and iconicity in Malay. *Bijdragen tot de Taal-, Land- en Volkenkunde, 165*(2/3), 291-323. Retrieved from http://www.jstor.org/stable/43817830

Bickerton, D., & Givón, T. (1976). Pidginization and syntactic change: From SXV and VSX to SVX. *Papers from the parasession on diachronic syntax*, 9-39.

Biggs, B. (1969). *Let's Learn Maori*. Wellington: Reed.

Blood, D. L. (1977). A three-dimensional analysis of Cham sentences. *Papers in South East Asian linguistics no. 4: Chamic studies*, 53-76.

Blust, R. (2001). Notes on the history of 'focus' in Austronesian languages. In F. Wouk & M. Ross (Eds.), *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics.

Blust, R. (2013). *The Austronesian Languages*. Canberra, Australia: Asia-Pacific Linguistics, Research School of Pacific and Asian Studies, the Australian National University.
Boechari. (1966). Preliminary report on the Discovery of an Old-Malay Inscription at Sodjomerto. *Madjalah Ilmu-Ilmu Sastra Indonesia, 3*(2&3), 241-251.
Bowers, J. (1993). The syntax of predication. *Linguistic Inquiry, 24*(4), 591-656.
Bowrey, T. (1701). *A Dictionary: English and Malayo, Malayo d English*. London: Sam. Bridge.
Bresnan, J., & Grimshaw, J. (1978). The Syntax of Free Relatives in English. *Linguistic Inquiry, 9*(3), 331-391. Retrieved from http://www.jstor.org/stable/4178069
Bril, I. (2005). Semantic and functional diversification of reciprocal and middle prefixes in New Caledonian and other Austronesian languages. *9*(1), 25-76. doi:doi:10.1515/lity.2005.9.1.25
Budd, P. S. (2009). *Topics in the grammar of Bierebo, Central Vanuatu, with a focus on the Realis/Irrealis categories.* (Ph.D. Dissertation). School of Oriental and African Studies (University of London),
Burzio, L. (1986). *Italian syntax*. Netherlands: Springer.
Carlson, G. N. (1977a). Reference to kinds in English.
Carlson, G. N. (1977b). A unified analysis of the English bare plural. *Linguistics and philosophy, 1*(3), 413-457.
Chantanakomes, V. (1980). *A description of Moken: A Malayo-Polynesian language.* (Master's Dissertation). Mahidol University,
Cheng, L. L. S. (1991). *On the Typology of Wh-Questions.* (Ph.D. Dissertation). Massachusetts Institute of Technology, Cambridge, MA.
Chomsky, N. (1981). *Lectures on Government and Binding* (1st ed.). Dodrecht: Foris Publications.
Chomsky, N. (2001). Derivation by Phase. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1-52). Cambridge, MA: MIT Press.
Chomsky, N. (2008). On phases. In R. Freidin, C. P. Otero, & M. L. Zubizarreta (Eds.), *Current Studies in Linguistics Series* (Vol. 45, pp. 133). Cambridge, Massachusetts: MIT Press.
Chung, S. (1976). On the subject of two passives in Indonesian. In C. N. Li (Ed.), *Subject and topic* (pp. 57-98). London ; New York: Academic Press.
Chung, S. (1982). Unbounded Dependencies in Chamorro Grammar. *Linguistic Inquiry, 13*(1), 39-77. Retrieved from http://www.jstor.org/stable/4178259
Clark, R. (1985). *Languages of North and Central Vanuatu: Groups, Chains, Clusters and Waves.* Paper presented at the Austronesian Linguistics at the 15th Pacific Science Congress, Canberra.
Cœdès, G. (1930). Les Inscriptions Malaises de Çrīvijaya. *Bulletin de l'École française d'Extrême-Orient, 30*(1/2), 29-80. Retrieved from http://www.jstor.org/stable/43731057
Cole, P., & Hermon, G. (2000). Partial Wh-Movernent: Evidence from Malay. In U. Lutz, G. Müller, & A. v. Stechow (Eds.), *Wh-scope marking* (Vol. 37, pp. 101-130). Amsterdam: John Benjamins Publishing Company.
Cole, P., & Hermon, G. (2008). VP Raising in a VOS Language. *Syntax, 11*(2), 144-197. doi:10.1111/j.1467-9612.2008.00106.x

Cole, P., Hermon, G., & Yanti. (2008). Voice in Malay/Indonesian. *Lingua, 118*(10), 1500-1553.
Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems* (Vol. 2): Cambridge university press.
Crowley, T. (1998). An erromangan (Sye) grammar. *Oceanic Linguistics Special Publications*(27), i-294.
Cumming, S. A. (1988). *Syntactic function and constituent order change in Malay.* (Ph.D. Dissertation). University of California, Los Angeles, Los Angeles.
Cumming, S. A. (1995). Agent position in the Sejarah Melayu. In P. Downing & M. Noonan (Eds.), *Word order in discourse* (Vol. 30, pp. 51-83). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Den Dikken, M. (2006). *Relators and linkers: The syntax of predication, predicate inversion, and copulas* (Vol. 47). Cambridge, MA: MIT press.
Deshpande, M. M. (1979). Genesis of Ṛgvedic Retroflexion. A Historical and Sociolinguistic Investigation. In M. M. Deshpande & P. E. Hook (Eds.), *Aryan and Non-Aryan in India* (pp. 235-316): University of Michigan Press.
Dewan Bahasa dan Pustaka. (2013). Korpus DBP. Retrieved from http//:www.sbmb.dbp.gov.my/korpusdbp/. from Dewan Bahasa dan Pustaka http//:www.sbmb.dbp.gov.my/korpusdbp/
Diedrich, D. (2018). *A grammar of Paku: a language of Central Kalimantan.* (Ph.D. Dissertation). The University of Melbourne, Victoria.
Dik, S. C. (1997). *The theory of functional grammar: the structure of the clause* (2 ed.): Mouton de Gruyter.
Dixon, R. M. W. (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
Donohue, M. (2005). Word Order in New Guinea: Dispelling a Myth. *Oceanic Linguistics, 44*, 527 - 536.
Donohue, M. (2007). Word order in Austronesian from north to south and west to east. *11*(2), 349-391. doi:doi:10.1515/LINGTY.2007.026
Dowty, D. R. (1979). *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel: Springer.
Elbourne, P. (2001). E-Type Anaphora As NP-Deletion. *Natural Language Semantics, 9*(3), 241-288. Retrieved from http://www.jstor.org/stable/23748811
Erlewine, M. Y., Levin, T., & Urk, C. v. (2017). Ergativity and Austronesian-Type Voice Systems. In J. Coon, D. Massam, & L. d. Travis (Eds.), *The Oxford handbook of ergativity* (pp. 373–396). Oxford: Oxford University Press.
Fischer, S. (2010). *Word-order Change as a Source of Grammaticalisation* (Vol. 157): John Benjamins Publishing.
Foley, W. (2005). Semantic parameters and the unaccusative split in the Austronesian language family. *Studies in Language, 29*, 385-430. doi:10.1075/sl.29.2.05fol
Fortin, C. R. (2007). *Indonesian sluicing and verb phrase ellipsis: Description and explanation in a minimalist framework.* (Ph.D. Dissertation). University of Michigan, Ann Arbor. ProQuest Dissertations & Theses Global database. (3287506)
Frege, G. (1892). Über begriff und gegenstand. *Vierteljahrsschrift für wissenschaftliche Philosophie, 16*(2), 192-205.

Gallego, Á. J., & Uriagereka, J. (2016). Estar = Ser + P. *Borealis: An International Journal of Hispanic Linguistics, 5*. doi:10.7557/1.5.1.3634
Givón, T. (1984). *Syntax: A Functional-Typological Introduction*. Amsterdam: Benjamins.
Givón, T. (2015). Topic, Pronoun and Grammatical Agreement. In T. Givón (Ed.), *The Diachrony of Grammar* (pp. 163-196). Amsterdam: John Benjamins,.
Grant, A. P. (2009). Contact, complexification and change in Mindanao Chabacano structure. In E. O. Aboh & N. Smith (Eds.), *Complex processes in new languages* (pp. 223-241).
Griffiths, A. (2018). The Corpus of Inscriptions in the Old Malay Language. In: École française d'Extrême-Orient.
Griffiths, A. (2020). The Old Malay Mañjuśrīgrffha inscription from Candi Sewu (Java, Indonesia). In V. Tournier, V. Eltschinger, & M. Sernesi (Eds.), *Archaeologies of the written: Indian, Tibetan, and Buddhist studies in honour of Cristina Scherrer-Schaub* (pp. 225-262).
Gudai, D. H. (1985). *A grammar of Maanyan: a language of Central Kalimantan.* (Ph.D. Dissertation). Australian National University, Canberra. Available from http://worldcat.org /z-wcorg/ database.
Gundel, J. K., & Fretheim, T. (2004). Topic and focus. *The handbook of pragmatics, 175*, 196.
Halliday, M. A. (1967). Notes on transitivity and theme in English: Part 2. *Journal of linguistics, 3*(2), 199-244.
Harahap, D. (1991). *Persoalan Bahasa [the question of language]*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Hassan, A., & Mohd., A. (1994). *Tatabahasa Dinamika*. Kuala Lumpur: Utusan Publications & Distributors Sdn. Bhd.
Hedberg, N. (2000). The referential status of clefts. *Language*, 891-920.
Heggie, L. A. (1989). *The syntax of copular structures.* (Ph.D. Dissertation). University of Southern California, Los Angeles.
Heine, B., & Kuteva, T. (2005). *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.
Hengeveld, K. (1992). *Non-verbal predication: theory, typology, diachrony*: Mouton de Gruyter.
Heycock, C., & Kroch, A. (1997). Inversion and equation in copular sentences. *ZAS Papers in Linguistics, 10*, 71-87.
Heycock, C., & Kroch, A. (2002). *Topic, focus, and syntactic representations.* Paper presented at the Proceedings of WCCFL.
Higgins, F. R. (1979). *The pseudo-cleft construction in English*. New York: Garland Pub.
Hikayat Amir Hamzah. (1987). Kuala Lumpur: Dewan Bahasa dan Pustaka.
Hogan, D. W., & Pattemore, S. W. (1988). *Urak Lawoi': Basic Structures and a Dictionary*: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
Holm, J. A. (1988). *Pidgins and creoles*. Cambridge, New York: Cambridge University Press.
Hopper, P. (1979). Aspect and Foregrounding in Discourse. In T. Givón (Ed.), *Syntax and Semantics* (Vol. 12, pp. 213-241): Academic Press.

Hopper, P. (1991). On some Principles of Grammaticalization. In E. C. Traugott & B. Heine (Eds.), *Approaches to Grammaticalization* (Vol. 1, pp. 17-35). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hopper, P., & Traugott, E. C. (2003). *Grammaticalization* (2 ed.). Cambridge: Cambridge University Press.

Jackendoff, R. S. (1972). *Semantic Interpretation in Generative Grammar*: MIT Press.

Jackson, J. A. J. (2014). *A grammar of Irarutu, a language of West Papua, Indonesia, with historical analysis.* (Ph.D. Dissertation). University of Hawai'i at Manoa,

Jensen, J. M. (2013). *The structure of Jarai clauses and noun phrases.* (Ph.D. Dissertation). University of Texas at Arlington,

Kader, M. B. (1981). *The syntax of Malay interrogatives*: Dewan Bahasa dan Pustaka, Kementerian Pelajaran, Malaysia.

Kader, M. B. (1986). Aspek-Aspek Sintaksis di dalam Bahasa Melayu Tinggi [aspects of syntax in high Malay]. In N. S. Karim (Ed.), *Bahasa Melayu Tinggi: Teori dan Penerapan*. Kuala Lumpur: Persatuan Linguistik Malaysia.

Karim, N. S., Onn, F. M., Musa, H. H., & Mahmood, A. H. (2014). *Tatabahasa Dewan [the Dewan grammar]* (3rd ed.). Kuala Lumpur: Dewan Bahasa dan Pustaka.

Khairul Taufiq, A. B., & Nor Hashimah, J. (2017). Kata Pemeri Ialah dan Adalah: Analisis Pragmatik [copulas ialah and adalah: a pragmatic analysis]. *Jurnal Linguistik, 21*(1), 14-28.

Kiss, K. É. (1998). Identificational focus versus information focus. *Language*, 245-273.

Koizumi, M. (1999). *Phrase Structure in Minimalist Syntax*. Tokyo: Hituzi Syobo.

Koster, J. (1978). Why subject sentences don't exist. In S. J. Keyser (Ed.), *Recent transformational studies in European languages* (pp. 53–64). Cambridge, MA: MIT Press.

Kozok, U. (2015). *A 14th Century Malay Code of Laws: The Nitisarasamuccaya* (Vol. 16): Institute of Southeast Asian Studies.

Kratzer, A. (1995). Stage-Level and Individual-Level Predicates. In G. N. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 125-175). Chicago, London: The University of Chicago Press.

Kraußе, D. (2017). *A description of Surabayan Javanese with special reference to its linguistic etiquette.* (Master's Dissertation). Goethe-Universität Frankfurt am Main,

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica, 55*(3-4), 243-276.

Kroeger, P. (2014). External negation in Malay/Indonesian. *Language, 90*(1), 137-184.

Lacrampe, S. (2014). *Lelepa: Topics in the grammar of a Vanuatu language.* (Ph.D. Dissertation). The Australian National Uiversity,

Lakoff, G. (1965). *On the nature of syntactic irregularity.* (Ph.D. Dissertation). Indiana University, Cambridge, Massachusetts.

Lakoff, G. (1976). Toward Generative Semantics. In J. D. McCawley (Ed.), *Notes from the Linguistic Underground* (pp. 43-61): Brill.

Larish, M. D. (2005). Moken and Moklen. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 513-533): Routledge.

Legate, J. A. (2012). Subjects in Acehnese and the nature of the passive. *Language, 88*, 495-525.

Levin, B., & Malka, R. (1986). The Formation of Adjectival Passives. *Linguistic Inquiry, 17*(4), 623-661. Retrieved from http://www.jstor.org/stable/4178511
Lewis, M. B. (1947). *Teach yourself Malay*. London: English Universities Press.
Lipták, A. (2011). The structure of the topic field in Hungarian. In B. P. & M. N. (Eds.), *Mapping the left periphery: The cartography of syntactic structures* (Vol. 5): Oxford University Press.
Lohndal, T. (2009). The copula cycle. In E. Van Gelderen (Ed.), *Cyclical change* (pp. 209-242).
Mahdi, W. (2005). Old Malay. *The Austronesian languages of Asia and Madagascar*, 182-201.
Maienborn, C. (2005). On the limits of the Davidsonian approach: The case of copula sentences. *Theoretical Linguistics, 31*, 275-316. doi:10.1515/thli.2005.31.3.275
Maienborn, C. (2008). On Davidsonian and Kimian States. In I. Comorovski & K. von Heusinger (Eds.), *Existence: Semantics and Syntax* (Vol. 84, pp. 107-130). Dordrecht: Springer.
McCawley, J. D. (1968). *Lexical insertion in a transformational grammar without deep structure*. Chicago: University of Chicago.
McCawley, J. D. (1996). The focus and scope of only. In B. Partee & P. Sgall (Eds.), *Discourse and meaning: Papers in honor of Eva Hajicová* (pp. 171-193).
McCloskey, J. (1996). Subjects and subject positions in Irish. In R. D. Borsley & I. G. Roberts (Eds.), *The syntax of the Celtic languages: a comparative perspective* (pp. 241-283): Cambridge University Press.
McGinn, R. (1998). Anti-ECP effects in the Rejang language of Sumatra. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *43*(3-4), 359-376.
Meira, S., & Gildea, S. (2009). Property concepts in the Cariban family: Adjectives, adverbs, and/or nouns? *LOT Occasional Series, 13*, 95-133.
Merchant, J. (2001). *The syntax of silence : sluicing, islands, and the theory of ellipsis / Jason Merchant*. Oxford ;: Oxford University Press.
Mikkelsen, L. (2005a). *Copular clauses: Specification, predication and equation* (Vol. 85): John Benjamins Publishing Company.
Mikkelsen, L. (2005b). Subject choice in copular clauses. *Unpublished MS, UC Berkeley*.
Moro, A. (1997). *The raising of predicates: Predicative noun phrases and the theory of clause structure* (Vol. 80): Cambridge University Press.
Moro, A. (2006). Copular Sentences. In *The Blackwell companion to syntax* (pp. 1-23).
Müller-Gotama, F. (1995). *Complementation in Classical Malay*. Paper presented at the the third annual meeting of the Southeast Asian Linguistics Society 1993.
Müller-Gotama, F. (1996). Clause structure in Classical Malay. *General Linguistics, 36*(1), 89.
Mustaffa, A. R. (2015). *Focus and Topic: Triggers of Copula Deletion in Interrogatives and Relative Clauses in Malay.* (Master's Dissertation). University College London, London, UK.
Mustaffa, A. R. (2018). Silent Syntactic Structures: Copular Clauses. *Journal of Modern Languages, 28*.
Mustaffa, A. R. (2020). Null Expletives in Malay. *GEMA Online® Journal of Language Studies, 20*(2), 224-243. Retrieved from http://ejournal.ukm.my/gema/article/view/37434

Nathesan, S. (2015). *Etimologi bahasa Melayu*: Dewan Bahasa dan Pustaka.
Neeleman, A., & Vermeulen, R. (2012). The Syntax of Topic, Focus, and Contrast. In N. Ad & V. Reiko (Eds.), *Chapter 1. The Syntactic Expression of Information Structure* (pp. 1-38): De Gruyter Mouton.
Noguchi, T. (1997). Two types of pronouns and variable binding. *Language*, 770-797.
Omar, A. H. (1968). Word Classes in Malay. *Anthropological Linguistics, 10*(5), 12-22. Retrieved from http://www.jstor.org/stable/30029361
Omar, A. H. (2008). *Nahu kemas kini : panduan bahasa yang baik dan betul*. Kuala Lumpur: PTS Professional Publishing.
Omar, A. H. (2014). *Nahu Melayu Mutakhir [a grammar of contemporary Malay]* (5th ed.). Kuala Lumpur: Dewan Bahasa dan Pustaka, Kementerian Pelajaran, Malaysia.
Omar, A. H., & Rama, S. (1968). *An Introduction to Malay Grammar*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Overall, S. E., Vallejos, R., & Gildea, S. (2018). *Nonverbal predication in Amazonian languages* (Vol. 122): John Benjamins Publishing Company.
Partee, B. (1977). John is easy to please. In A. Zampolli (Ed.), *Linguistic Structures Processing* (pp. 281-312). Amsterdam: North-Holland Publishing Co.
Partee, B. (1998). *Copula inversion puzzles in English and Russian.* Paper presented at the Formal Approaches to Slavic Linguistics: The Seattle Meeting.
Paul, I. (2008). On the topic of pseudoclefts. *Syntax, 11*(1), 91-124.
Percus, O. (1997). *Prying open the cleft.* Paper presented at the Proceedings of NELS.
Peyraube, A., & Wiebusch, T. (1994). Problems relating to the history of different copulas in Ancient Chinese. In M. Y. Chen & O. J. L. Tseng (Eds.), *In Honor of William SY. Wang: Interdisciplinary studies on language and language change* (pp. 383-404). Taipei: Pyramid Press.
Polinsky, M., & Potsdam, E. (Forthcoming). Austronesian Syntax. In B. Palmer (Ed.), *Oceania*. Berlin: Mouton.
Postma, A. (1992). The Laguna Copper-Plate Inscription: Text and Commentary. *Philippine Studies, 40*(2), 183-203. Retrieved from http://www.jstor.org/stable/42633308
Price, D. C. (2007). Bundu Dusun sketch grammar. *Unpublished Manuscript*.
Proudfoot, I. (1991). Concordances and classical Malay. *Bijdragen tot de taal-, land- en volkenkunde/Journal of the Humanities and Social Sciences of Southeast Asia, 147*(1), 74-95.
Pustet, R. (2003). *Copulas: Universals in the Categorization of the Lexicon*: OUP Oxford.
Reinhart, T. (1982). Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica, 27*, 53-94.
Remijsen, B. (2010). Nouns and verbs in Magey Matbat. In *East Nusantara:* (Vol. 618, pp. 281-311). Canberra: Pacific Linguistics, Canberra :.
Ringe, D. (2017). *From Proto-Indo-European to Proto-Germanic* (2 ed.). Oxford: Oxford University Press.
Rizzi, L. (1990). *Relativized minimality*: The MIT Press.
Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of Grammar: Handbook of Generative Syntax* (pp. 281-337). Dodrecht: Kluwer.

Rizzi, L. (2010). On Some Properties of Criterial Freezing. In P. Panagiotidis (Ed.), *The Complementizer Phase: Subjects and Operators*: Oxford University Press.
Rooth, M. (1985). Association with focus.
Ross, J. (1967). *Constraints on variables in syntax.* (Ph.D. Dissertation). Massachusetts Institute of Technology, Cambridge, Massachusetts.
Saddy, D. (1992). A versus A-bar movement and wh-fronting in Bahasa Indonesia. *Ms., University of Queensland, Australia*.
Salleh, R. (1993). Verb Movement in Malay. *Jurnal Persatuan Linguistik, 6*.
Salleh, R. (1995). *Sintaksis bahasa Melayu: penerapan teori kuasaan dan tambatan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Schuelke, P. (2020). *Grammatical Relations and Syntactic Ergativity in Roviana: A little-described language of Solomon Islands.* University of Hawai'i at Manoa,
Simin, A. M. (1988). *Discourse-Syntax of 'Yang' in Malay (Bahasa Malaysia)*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Smith, G. (2008). Tok Pisin: morphology and syntax. In K. Bernd, W. S. Edgar, & B. Kate (Eds.), *3 The Pacific and Australasia* (pp. 488-513). Berlin, New York: De Gruyter Mouton.
Sneddon, J. N. (1996). *Indonesian: A comprehensive grammar*. London: Routledge.
Soh, H. L., & Nomoto, H. (2011). The Malay verbal prefix meN- and the unergative/unaccusative distinction. *Journal of East Asian Linguistics, 20*(1), 77-106.
Soriente, A. (2018). Deixis in Borneo: Kenyah and Punan.
Starosta, S., Pawley, A., & Reid, L. A. (1982). *The Evolution of Focus in Austronesian*.
Stassen, L. (1997). *Intransitive predication*. Oxford, England: Oxford University Press.
Stassen, L. (2013). Zero Copula for Predicate Nominals. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Steinhauer, H. (2005). Colonial history and language policy in insular Southeast Asia and Madagascar. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 65-86): Routledge.
Stowell, T. (1983). Subjects Across Categories. *The Linguistic Review, 2*(3), 285–312.
Suhadi, M. (1983). *Seven Old-Malay inscriptions found in Java.* Paper presented at the SPAFA Consultative Workshop on Archaeological and Environmental Studies on Srivijaya.
Svenonius, P. (1998). Clefts in Scandinavian: an investigation. *ZAS Papers in Linguistics, 10*, 163-190.
Teng, S. F.-C. (2007). *A reference grammar of Puyuma, an Austronesian language of Taiwan.* (Ph.D. Dissertation). Australian National University,
Thomason, S. G., & Kaufman, T. (1992). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
Thurgood, G. (1999). From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change: With an Appendix of Chamic Reconstructions and Loanwords. *Oceanic Linguistics Special Publications*(28), i-407. Retrieved from http://www.jstor.org/stable/20006770
Thurgood, G., & Li, F. (2002). *Contact induced variation and syntactic change in the Tsat of Hainan.* Paper presented at the Annual Meeting of the Berkeley Linguistics Society.

Tjia, J. (2015). Grammatical relations and grammatical categories in Malay. The Indonesian prefix meN - revisited. *Wacana: Journal of the Humanities of Indonesia, 16*, 105-132.
Totanes, S. d. (1745). *Arte de la Lengua Tagala y Manual Tagalog*.
Totsuka, M. (2015). *On Phasehood Of Functional Categories In The Left Periphery.* (Ph.D. Dissertation). Tohoku University,
Travis, L. d. (1984). *Parameters and effects of word order variation.* Massachusetts Institute of Technology,
Travis, L. d. (2010). *Inner Aspect: The Articulation of VP*.
van den Heuvel, W. (2006). *Biak Description of an Austronesian Language of Papua*. Utrecht: LOT.
Van Gelderen, E. (2011). The Pronominal Copula Cycle. In E. Van Gelderen (Ed.), *The Linguistic Cycle: Language Change and the Language Faculty*: Oxford University Press.
Van Gelderen, E. (2015). The copula cycle. *Lingue e linguaggio, 14*(2), 287-301.
van Minde, D. (2008). The pragmatic function of Malay yang. *Journal of Pragmatics, 40*(11), 1982-2001. doi:https://doi.org/10.1016/j.pragma.2008.03.006
Verhaar, J. W. M. (1982). On the Syntax of Yang in Indonesian. In A. Halim, L. Carrington, & S. A. Wurm (Eds.), *Papers from the Third International Conference on Austronesian Linguistics* (pp. 43-70). Dept. of Linguistics, Research School of Pacific Studies, Australian National University.
Verhaar, J. W. M. (1991). The Function of I in Tok Pisin. *Journal of Pidgin and Creole Languages, 6*(2), 231-266. doi:https://doi.org/10.1075/jpcl.6.2.04ver
von Prince, K. (2012). *A grammar of Daakaka.* (Ph.D. Dissertation). University of Humboldt, Berlin.
Waterhouse, J. L. (1928). *A Roviana and English dictionary: with English-Roviana index and list of natural history objects*: Melanesian Mission Press.
Wilkinson, R. J. (1908). *An abridged Malay-English dictionary (romanised)*.
Williams, E. (1980). Predication. *Linguistic Inquiry, 11*(1), 203-238. Retrieved from http://www.jstor.org/stable/4178153
Willson-Sturman, H. (2014). Pseudocleft constructions in Marshallese wh- questions. *Lingua, 145*, 1-35. doi:https://doi.org/10.1016/j.lingua.2014.01.014
Winstedt, R. O. (1913). *Malay Grammar*: Oxford University Press.
Wok Awang, H. (1981). *Some Aspects of Malay Relativization: A Transformational Account by Deletion.* (Ph.D. Dissertation). University of London,
Woolford, E. (1979). Variation and Change in the i 'Predicate Marker' of New Guinea Tok Pisin. *Papers in Pidgin and Creole Linguistics No. 2*.
Woollams, G. D. (2005). Karo batak. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 534-561): Routledge.
Yap, F. H. (2007). *Nominalizers (and copulas) in Malay.* Paper presented at the International Workshop on Nominalizers and Copulas in East Asian and Neighboring Languages, Chinese University of Hong Kong.
Yap, F. H. (2011). Referential and non-referential uses of nominalization constructions in Malay. In F. H. Yap, K. Grunow-Hårsta, & J. Wrona (Eds.), *Nominalization in Asian Languages: Diachronic and Typological Perspectives* (pp. 627-658). Amsterdam: John Benjamins Publishing Company.
Zheng, Y. (1997). *Huihui Yu Yanjiu [A Study of Cham]*: Shanghai Yuandong Chuban She [Shanghai Far East Publishers].

Zhiming, B., & Aye, K. K. (2010). Bazaar Malay topics. *Journal of Pidgin and Creole Languages, 25*(1), 155-171. doi:https://doi.org/10.1075/jpcl.25.1.06bao

Newspaper Articles

Abd Jamil, N. S. (2021, July 23). Perbalahan Ebit Lew, Caprice dan Nadir mengundang fitnah... 2 mufti harap individu berbalah kembali kepada ajaran al-Quran dan sunah Rasulullah. *MStar*. Retrieved from https://www.mstar.com.my/lokal/semasa/2021/07/23/perbalahan-ebit-lew-caprice-dan-nadir-mengundang-fitnah-2-mufti-harap-individu-berbalah-kembali-kepada-ajaran-al-quran-dan-sunah-rasulullah

Abd Mutalib, Z., & Halid, S. (2020, May 3). Rakyat perlu buat kawalan kendiri cegah penularan COVID-19. *Berita Harian*. Retrieved from https://www.bharian.com.my/berita/nasional/2020/05/684671/rakyat-perlu-buat-kawalan-kendiri-cegah-penularan-covid-19

Abd Mutalib, Z., & Shahrul Annuar, S. (2019, December 17). 10,000 rumah mampu milik dibina di Bandar Malaysia. *Berita Harian*. Retrieved from https://www.bharian.com.my/berita/nasional/2019/12/639527/10000-rumah-mampu-milik-dibina-di-bandar-malaysia

Abdul Karim, M. A. F. (2021, August 26). 'Tuhan beri keizinan rehatkan diri' – Shahrol Shiro. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/nasional/2021/08/tuhan-beri-keizinan-rehatkan-diri-shahrol-shiro/

Abdul Rahim, A. (2016, September 13). Apakah itu pengaruh mikro? *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/bisnes/korporat/apakah-itu-pengaruh-mikro-1.382134

Abdul Rahman, H. (2018, October 28). Isteri mati dipercayai ditetak suami. *Berita Harian*. Retrieved from https://www.bharian.com.my/berita/kes/2018/10/491441/isteri-mati-dipercayai-ditetak-suami

Abllah, N. (2019, 7 May). China cadang bina 100,000 rumah mampu milik. *Berita Harian*. Retrieved from https://www.bharian.com.my/bisnes/lain-lain/2019/05/561137/china-cadang-bina-100000-rumah-mampu-milik

Abu Hassan, Z. (2021, February 10). 'Ini jiwa saya yang sebenar' – Amira Othman. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/gaya/hiburan/2021/02/ini-jiwa-saya-yang-sebenar-amira-othman/

Adnan, A. F. (2021, September 24). Dos lengkap vaksin berkesan lebih 80 peratus. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/terkini/2021/09/dos-lengkap-vaksin-berkesan-lebih-80-peratus/

AFP. (2019, January 25). Kvitova ada strategi benam Osaka. *Berita Harian*. Retrieved from https://www.bharian.com.my/sukan/raket/2019/01/524014/kvitova-ada-strategi-benam-osaka

Agensi. (2021, June 26). Mancini buru trofi, Foda tak gentar. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/terkini/2021/06/mancini-buru-trofi-foda-tak-gentar/

Amran, S. N. M. E. (2018, November 27). 'Kita tengok mereka terima ke tidak...'. *Berita Harian*. Retrieved from

https://www.bharian.com.my/berita/politik/2018/11/502727/kita-tengok-mereka-terima-ke-tidak
BERNAMA. (2015, March 3). SPM 2014 : Keputusan masih ikut trend - Muhyiddin. *Berita Harian*. Retrieved from https://www.bharian.com.my/node/38259
BERNAMA. (2017, March 5). Pembunuhan Jong-nam segarkan kembali kejadian 39 tahun lalu. *Berita Harian*. Retrieved from https://www.bharian.com.my/node/256280
BERNAMA. (2019, December 18). MH17: MMU bertafakur pada konvokesyen hujung minggu ini. *Berita Harian*. Retrieved from https://www.bharian.com.my/node/1961
Che Noh, N. (2018, January 2). Macam-macam ragam. *Berita Harian*. Retrieved from https://www.bharian.com.my/berita/pendidikan/2018/01/369828/macam-macam-ragam
Dahali, R. (2021, September 29). JKNS nafi mesej palsu. *Harian Metro*. Retrieved from https://www.hmetro.com.my/mutakhir/2021/09/760620/jkns-nafi-mesej-palsu
Fadzlyana, R. (2017, August 19). Kenal pasti punca sembelit. *Harian Metro*. Retrieved from https://www.hmetro.com.my/sihat/2017/08/255481/kenal-pasti-punca-sembelit
Fauzi, F. (2019, March 26). 'Jangan dengar cakap orang'. *Harian Metro*. Retrieved from https://www.hmetro.com.my/arena/2019/03/438016/jangan-dengar-cakap-orang
Hussein, N. (2022, February 13). Manifesto politik dari ayah. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/gaya/sastera/2022/02/manifesto-politik-dari-ayah/
Idris, R. (2018, May 24). Shahidan 'tak mahu' Azlan. *Berita Harian*. Retrieved from https://www.bharian.com.my/berita/politik/2018/05/429854/shahidan-tak-mahu-azlan
Idris, S. R. (2017, August 5). Etnik Siam turut sumbang dalam sektor pelancongan. *Berita Harian*. Retrieved from https://www.bharian.com.my/berita/wilayah/2017/08/308941/etnik-siam-turut-sumbang-dalam-sektor-pelancongan
Isa, I. S. (2022, February 18). PRN Johor: AMK P.Pinang ingatkan Syed Saddiq jangan ‘melampaui batas’ *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/berita/2022/02/prn-johor-amk-p-pinang-ingatkan-syed-saddiq-jangan-melampaui-batas/
Linch, C. (2022, April 17). Pemimpin Sarawak, Sabah ada Peluang jadi TPM. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/nasional/2022/04/pemimpin-sarawak-sabah-ada-peluang-jadi-tpm/
Mahmud, N. H. (2021, October 28). Cara elak demam kuning. *Harian Metro*. Retrieved from https://www.hmetro.com.my/amp/WM/2021/10/771148/cara-elak-demam-kuning
Mohamad Yatim, M. N. (2021, February 6). Jadikan STEM lebih kreatif, seronok. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/rencana/2021/02/jadikan-stem-lebih-kreatif-seronok/

Mohamad, Z. (2015, March 30). Jamal tak cukup tidur gara-gara persembahan 'maraton'. *Berita Harian*. Retrieved from https://www.bharian.com.my/taxonomy/term/8/2015/03/44229/jamal-tak-cukup-tidur-gara-gara-persembahan-maraton

Mohamad, Z. (2016, July 18). Gambar Adi, isteri tersenyum timbulkan tanda-tanya. *Berita Harian*. Retrieved from https://www.bharian.com.my/node/174700

Mohamada, K. A. (2020, August 11). Sejarah khat dalam al-Quran. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/gaya/2020/08/sejarah-khat-dalam-al-quran/

Mohd Yusoff, A. N. (2016, January 26). Abai hak syariah punca rumah tangga pincang. *Berita Harian*. Retrieved from https://www.bharian.com.my/node/118089

Muhammad Malik. (2019, October 27). Subri idam piala bersama TFC. *Harian Metro*. Retrieved from https://www.hmetro.com.my/arena/2019/10/511154/subri-idam-piala-bersama-tfc

Najib, jangan lupa saman Tun M. (2015, July 4). *Astro Awani*. Retrieved from http://www.astroawani.com/berita-malaysia/najib-jangan-lupa-saman-tun-m-blogger-64784

Nasir, A., & Abu Bakar, J. M. (2021, March 28). 'Sumpah, saya tak terima habuan RM20,000'. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/berita/2021/03/sumpah-saya-tak-terima-habuan-rm20000/

Ngah, N. (2018, July 6). Ayah sanggup menderita jika itu membahagiakan rakyat - Nazifuddin. *Berita Harian*. Retrieved from https://www.bharian.com.my/berita/nasional/2018/07/445718/ayah-sanggup-menderita-jika-itu-membahagiakan-rakyat-nazifuddin

Riduan, J. (2020, February 1). Sabah tetap dengan pendirian. *Harian Metro*. Retrieved from https://www.hmetro.com.my/mutakhir/2020/02/540494/sabah-tetap-dengan-pendirian

Shahimi, I. M. A. (2018, July 30). Jadi diri sendiri sepuas-puasnya. *Harian Metro*. Retrieved from https://www.hmetro.com.my/hati/2018/07/363159/jadi-diri-sendiri-sepuas-puasnya

Shahimi, I. M. A. (2019, November 14). Rekod penting perlu ditanda ibu [METROTV]. *Harian Metro*. Retrieved from https://www.hmetro.com.my/WM/2019/11/516976/rekod-penting-perlu-ditanda-ibu-metrotv

Sulaiman, M. (2016, January 4). Tip hadapi musibah dari perspektif Islam. *Berita Harian*. Retrieved from https://www.bharian.com.my/node/110845

Sulaiman, N. A. (2021, October 24). 5,666 jangkitan baharu hari ini. *Berita Harian*. Retrieved from https://www.bharian.com.my/berita/nasional/2021/10/879580/5666-jangkitan-baharu-hari-ini

Yusoff Bakri, M. H. (2021, February 14). Pengasingan kuasa libatkan pindaan Perlembagaan Persekutuan. *Utusan Malaysia*. Retrieved from https://www.utusan.com.my/berita/2021/02/pengasingan-kuasa-libatkan-pindaan-perlembagaan-persekutuan/